Empirical Studies in the Economics of Education

Tommaso Sartori

A thesis submitted for the degree of

Doctor of Philosophy in Economics

Institute for Social and Economic Research

University of Essex

Declaration

No part of this thesis has been submitted for another degree.

Chapter 1 is co-authored with Professor Adeline Delavande, Professor Emilia Del Bono, and Dr. Angus Holford. Chapter 2 is co-authored with Professor Emilia Del Bono and Dr. Angus Holford. Chapter 3 is my sole, independent work.

All chapters have been prepared in accordance with the University of Essex Principal Regulations for Research Degrees.

Chapter 2 uses data from the Aberdeen Children of the 1950s (ACONF), available through the Grampian Data Safe Haven (DaSH), a secure data analysis and storage facility established by the University of Aberdeen and NHS Grampian. The use of this data in this work does not imply endorsement by DaSH, the University of Aberdeen, or NHS Grampian with respect to the interpretation or analysis presented.

Chapter 3 uses data from the Ministry of Education of Italy (*Ministero dell'Università e della Ricerca*), provided under a research agreement between me and the Directorate for the Statistical Services. The use of this data in this work does not imply endorsement by the Ministry of Education with respect to the interpretation or analysis presented.

I acknowledge the use of OpenAI's ChatGPT and Grammarly Inc.'s Grammarly for tasks including brainstorming, drafting, and language refinement. I take full responsibility for the content, interpretation, and conclusions of this work.

Acknowledgments

I want to thank my examiners, Emma Duchini and Ingo Isphording, for accepting the role and for their thoughtful and constructive feedback.

I am grateful to my supervisors, Emilia, for supporting me in seizing the opportunities I had throughout my doctorate and pushing me to strive for continuous improvement, and Angus, for always having his door open and offering insightful suggestions at every step of my work. I also want to thank Massimiliano for his mentorship throughout all the stages of my journey, and Rigissa for the support and the advice she shared along the way.

I have been incredibly fortunate to be surrounded by many wonderful people during the time I spent in Milan, Essex, and Melbourne, far too many to thank individually. But in particular, to Anna and Omar, with whom I shared this journey (and many dinners). To Pietro, who I am sure will become a brilliant economist. To my friends who have become like family, Nicola and Ludo. To those I met along the way, especially Flavio and Lorenz. And above all, to Andrea, who has always been there.

To my family — my aunts Dina, Flavia, Barbara, and Lorena, my uncles and cousins, whose warmth and support have always been there. To Camilla, who tried her hardest to make me a cool person, without ever succeeding, and to Virginia, who is still my baby after almost 22 years. To Mum and Dad, who always did their best and allowed me the freedom to find my own path, for which I feel beyond blessed.

To Brooke, for her love and support, and for always trying to make each step of our journey together a little lighter. I look forward to what lies ahead.

Finally, to those who were a big part of my life and are no longer with me. My uncle Ferruccio, my grandfather Corrado, my aunt Margherita, and especially my grandmother, Rosetta. I wish I could talk to you again.

Contents

List of Figures

List of Tables

Abstract

This thesis consists of three papers exploring different topics in the economics of education. Each chapter addresses a distinct question related to student behavior, academic outcomes, and educational interventions.

Chapter 1 studies whether goal-setting can improve student attendance. Using a randomized controlled trial, we find that students who set attendance goals attend one additional lecture over the term, on average. While this does not translate into improved academic performance, treated students report greater interest in their field of study. The effect is strongest for students with poor planning ability.

Chapter 2 looks at the effect of academic rank in primary school using data on all children enrolled in primary school in Aberdeen, Scotland, in 1962. Higher rank within a school-cohort group improves performance on the high-stakes 11-plus exam, raises long-term educational attainment, and strengthens internalizing skills — especially for girls. However, only boys experience lasting income gains, likely reflecting historical constraints on women's access to higher education and skilled employment.

Chapter 3 analyzes the impact of a program introducing additional training in mathematics and physics in Italian high schools. The overall effect on university STEM enrollment is null, but significant gender heterogeneity emerges. Boys attending the program are more likely to choose STEM majors, while girls do not respond on average, effectively widening the gender gap. However, this pattern varies by cultural context: girls from areas with more traditional gender norms are significantly more likely to pursue STEM if they attended a program school, suggesting that curricular interventions can counteract restrictive social norms and reduce gender disparities in STEM participation.

Introduction

Education is a critical driver of economic growth and social mobility, shaping the skills and opportunities available to individuals over their lifetime. High educational attainment is linked to higher earnings (Card, 1999), better health and longer life expectancy (Grossman, 2006), and is widely viewed as a key determinant of national productivity and innovation (Waldinger, 2016; Moser et al., 2014). However, the factors that shape educational outcomes are diverse, ranging from students' behavioral traits to peer dynamics to curricular content. Understanding these factors is essential for creating learning environments that effectively promote skill development, with benefits for both individuals and national economies.

This thesis examines the determinants of educational outcomes at three critical stages of the academic lifecycle, moving from higher education to primary school and then to secondary education. It focuses on how student behavior, peer interactions, and curricular exposure influence short- and long-term educational trajectories. While each chapter addresses a distinct aspect of this broader question, they share a common interest in how individual characteristics and contextual factors interact to shape educational success.

The first chapter examines the impact of goal-setting interventions on student attendance in higher education. Attendance has a dualistic nature, as it is both a critical determinant of students' academic achievement (Delavande et al., 2023; Dobkin et al., 2010) and (often) a sign of their overall commitment to the academic journey. While the returns to education are well-documented, students do not always behave in ways that align with these long-term benefits. Limited information has been identified as a potential explanation for this misalignment (Hoxby and Turner, 2013; Jensen, 2010), but many students also face less tangible but powerful behavioral challenges, leading them to prioritize immediate gratification over long-term benefits. These present-biased behaviors result in myopic educational decisions (Cadena and Keys, 2015; O'Donoghue and Rabin, 1999) that undermine students' ability to follow through on academic commitments, including attending lectures and engaging in coursework (Lavecchia et al., 2016; Oreopoulos and Petronijevic, 2013; Oreopoulos and Salvanes, 2011). In fact, many students continue to miss classes, potentially limiting their long-term educational and career prospects.

Behavioral economics provides a promising framework for addressing this challenge. Goal-setting strategies, in particular, have emerged as effective soft commitment devices, helping students overcome present bias by creating a clear reference point for the completion of a task and reducing the gap between intention and action (Patterson, 2018; Kaur et al., 2015). In this study, we conducted a randomized controlled trial at a UK public university. At the beginning of the term, treated students were asked to set an attendance goal, received weekly feedback, and were given the option to enter a commitment contract that withheld part of their compensation unless the goal was met. We find that the intervention led to a statistically significant increase in attendance, roughly equivalent to one additional lecture over the term. This increase did not translate into improved performance in final exams, though there was a positive effect on students' interest in their field of study. We also examine those entering the commitment contract, who show a larger increase in attendance and interest, but still do not see a performance improvement.

The findings highlight both the potential and the limitations of goal-setting interventions. On one hand, these strategies are easily scalable and, once set up, relatively inexpensive to implement, making them a promising approach for improving student engagement. Moreover, their effectiveness seems to be concentrated among students with weaker self-regulation and planning abilities, who may benefit the most from external structure and accountability. However, the lack of effects on academic performance aligns with a broader literature showing that first-order behavioral changes, like increased attendance, do not always translate into second-order outcomes, such as improved grades or graduation rates (Clark et al., 2020; Brade et al., 2018). This disconnect suggests that the relationship between attendance and academic performance may be less straightforward than commonly assumed.

The second chapter shifts the focus to the study of rank effects, a particular form of peer influence determined by a student's relative academic standing within their peer group. This relative position can shape a wide range of outcomes, including future academic performance, educational attainment, career choice (Megalokonomou and Zhang, 2024; Carneiro et al., 2023; Elsner et al., 2021; Murphy and Weinhardt, 2020; Elsner and Isphording, 2018, 2017), major choice (Goulas et al., 2023; Delaney and Devereux, 2021), and future earnings (Denning et al., 2023). We contribute to this growing literature by extending the range of outcomes considered, the measurement approaches employed, and the mechanisms explored. Specifically, we examine the effect on non-cognitive skills and parental investment, as well as on long-term measures of educational attainment and income. Tracing the impact of rank across this broad set of outcomes allows us to provide a comprehensive view of how early relative academic position can shape life trajectories.

To address these questions, we draw on a unique population-level survey covering all children born in Aberdeen (Scotland) between 1950 and 1955 and enrolled in primary school. Our data combine detailed administrative records with survey data on family background, early academic achievement, and long-term outcomes measured through a follow-up survey conducted approximately 40 years later. Crucially, we demonstrate that peer group composition within these schools is conditionally random, with no evidence of systematic sorting based on observed characteristics. We find that higher rank within the school-cohort group significantly improves both cognitive and non-cognitive outcomes, including performance on subsequent test scores and measures of internalizing skills, which capture traits related to self-concept and emotional stability. Rank also positively affects long-term educational attainment and annual earnings 40 years later, but with important gender differences. Girls benefit more in terms of test scores and educational attainment, while boys experience larger gains in future earnings.

An important contribution of this chapter is the use of rich, teacher-reported measures of non-cognitive skills, which allow us to derive reliable measures of externalizing and internalizing skills (Attanasio et al., 2020a; Boyle and Jones, 1985). We find that the impact of rank is concentrated on internalizing skills — traits like self-esteem, emotional regulation, and social confidence. In contrast, externalizing skills, reflecting more outward-directed behaviors like hyperactivity or aggression, show weaker and less consistent effects. We also explore parental involvement as a potential mechanism, but find limited evidence that parents adjust their behavior based on their child's rank. Finally, while girls benefit more from higher rank in terms of educational attainment, these gains do not consistently translate into higher earnings, likely reflecting the historical constraints on women's access to higher education and skilled employment for these cohorts.

The third chapter examines the impact of increasing the scientific content of secondary school curricula on students' educational choices, focusing on the gender gap in STEM participation. The demand for STEM workers has grown substantially in recent decades, driven by rapid technological advancements and the critical role of innovation in sustaining economic growth (Directorate General for Employment, 2023; BusinessEurope, 2023). STEM graduates are essential for productivity and technological progress (Waldinger, 2016; Moser et al., 2014), but a persistent gender gap in these fields limits the available pool of talent. This gap is a significant factor contributing to the broader gender wage gap (Wiswall and Zafar, 2017; Zafar, 2013), alongside other structural barriers such as differences in career/family trade-offs (Bertrand, 2020), occupational segregation (Blau and Kahn, 2017), and school schedules (Duchini and Van Effenterre, 2022).

In this chapter, I study the impact of a program available to Italian high school students opting for a scientific education, which introduced additional weekly hours of mathematics and physics. The Italian educational system provides a quasi-experimental setting for evaluating this program, as the decision to adopt the curriculum was made at the school level and depended on a range of factors unrelated to student characteristics, creating quasi-random variation in program availability. To account for students' selection, I instrument the attendance to a school offering the program with the availability of those schools within a student's commuting area. I show that the instrument is conditionally unrelated to municipality characteristics, mitigating concerns that local economic or social factors might confound the estimates. Importantly, the instrument is also uncorrelated with the initial decision to enroll in a scientific track, supporting the validity of the identification strategy. The results indicate modest average effects on STEM enrollment and no significant impact on dropout or first-year performance. However, I find substantial gender heterogeneity. While boys exposed to the program are significantly more likely to pursue STEM degrees, the average effect for girls is close to zero. Notably, the impact on girls varies significantly with local gender norms: in more conservative areas, girls exposed to the program are significantly more likely to choose STEM majors.

These findings contribute to the growing literature on the impact of additional mathematics and science instruction, which has consistently shown that early exposure to quantitative subjects can have long-term effects on educational attainment and labor market outcomes (De Philippis, 2023; Goodman, 2019; Cortes et al., 2015). They also add to the literature on the interaction between cultural norms and educational choices, highlighting the potential for curricular interventions to counteract restrictive gender norms and reduce disparities in STEM participation. While previous studies have documented substantial gender gaps in the response to STEM-promoting policies (De Philippis, 2023; Joensen and Nielsen, 2016), this study shows that the effectiveness of such interventions can be highly context-dependent, with potentially large gains for girls in more conservative environments.

Chapter 1

Boosting Attendance through Goal Setting: Evidence from a Randomized Experiment

1.1 Introduction

A highly educated workforce is a crucial driver of economic growth and innovation in advanced economies[1]. Recognizing this, policymakers have invested significant resources in expanding access to education and improving student outcomes, aiming to increase the pool of skilled workers necessary for maintaining economic competitiveness. Given the overwhelming evidence demonstrating how individuals themselves also benefit from completing more years of education (Card (1999) provides an excellent summary of the literature), we might expect students to invest in their education, and to be naturally committed to it. However, research shows that students do not always behave in ways that align with the strong individual returns to education.[2]

[1]Barro (2001) highlights the relationship between human capital accumulation and economic growth; Moretti (2004) finds that education has positive spillover effects on productivity; Hanushek and Woessmann (2008) emphasize cognitive skills as key drivers of economic development; Acemoglu and Autor (2011) discuss skill-biased technological change and the consequent demand for education; Goldin and Katz (2008) highlight the historical role of education in fostering economic growth.

[2]There is substantial evidence that students underinvest in higher education due to informational constraints or misperceptions about the value or demands of post-secondary education. Stinebrickner and Stinebrickner (2008) show that many students drop out of college because they misjudge the academic

Even when financial barriers and informational gaps are addressed, students do not always engage with education in ways that maximize their long-term outcomes. Present bias can play a crucial role in this, leading individuals to prioritize immediate gratification over long-term benefits, resulting in myopic educational decisions (Cadena and Keys, 2015; O'Donoghue and Rabin, 1999). This tendency can make it challenging for students to follow through on their academic commitments, such as attending lectures and engaging in coursework, even when they recognize the long-term benefits of doing so (Lavecchia et al., 2016; Oreopoulos and Petronijevic, 2013; Oreopoulos and Salvanes, 2011). Understanding how to mitigate the effects of present bias is therefore crucial to fostering consistent academic engagement and improving student outcomes.

One tool that has been extensively tested to improve student effort and outcomes is the use of performance-based financial incentives. These programs have yielded limited success, often failing to produce sustained improvements in academic performance. Moreover, their high implementation costs make them a challenging and inefficient policy tool for large-scale educational reforms (Cohodes and Goodman, 2014; Rudd et al., 2013; De Paola et al., 2012; Leuven et al., 2010; Angrist et al., 2009; Cornwell et al., 2005; Henry et al., 2004). In contrast, behavioral interventions have emerged as a promising alternative. These approaches leverage psychology and behavioral economics insights to encourage students towards better academic habits. They are often cost-effective, simple to implement, and easily scalable, making them an attractive option for policymakers and educators. For instance, commitment devices, including self-imposed deadlines, have proven effective in reducing procrastination and improving educational outcomes (Lavecchia et al., 2016; Ariely and Wertenbroch, 2002). Nudging strategies, such as personalized reminders, have successfully increased student retention and academic persistence (Damgaard and Nielsen, 2018; Castleman and Page, 2015). Among these, goal-setting interventions have shown potential in improving self-regulation and encouraging behaviors aligned with academic success, such as engagement with study and task completion

difficulty, rather than due to financial constraints. Hoxby and Turner (2013) find that even high-achieving, low-income students often do not apply to selective colleges they could access, due to limited information and outreach. Similarly, Jensen (2010) and Dinkelman and Martínez (2014) provide experimental evidence that improving information about the returns to education can significantly influence schooling decisions.

(Clark et al., 2020; Himmler et al., 2019).

The effectiveness of goal setting as an intervention is grounded in insights from behavioral economics, which suggest that individuals with self-control deficiencies driven by present bias can improve their behavior through commitment devices (Patterson, 2018; Kaur et al., 2015; DellaVigna and Malmendier, 2006; Ashraf et al., 2006; Thaler and Benartzi, 2004; Ariely and Wertenbroch, 2002; Wertenbroch, 1998). Goals function as internal commitment devices by providing a clear reference point that helps individuals stay on track and reinforces their intentions. In this way, goals can shape how individuals perceive their future outcomes, reducing the gap between short-term impulses and long-term objectives. This mechanism is consistent with the framework proposed by Koch and Nafziger (2011), which suggests that students who are present biased and loss averse are more likely to adjust their behavior when goals frame success and failure in a way that makes immediate costs and benefits more salient, aligning with prospect theory (Kahneman and Tversky, 1979).

We study how effective goal-setting is at improving students' attendance, as well as other academic outcomes. Using data from the BOOST2018 study (Delavande et al., 2022), linked with administrative records on student demographics and attendance, we track a cohort of undergraduate students in a public UK university from enrollment to graduation. During the spring term of their second academic year, we conducted an experiment in which students were randomized into treatment and control groups using a stratified design. Students in the treatment group watched a short video emphasizing the benefits of goal-setting and were then prompted to set an attendance goal for their lectures and classes. They received weekly feedback and reminders via SMS or email. Additionally, they had the option to enter a commitment contract, locking a portion of their participation reward in a special account that would be released only if they met their attendance goal. In contrast, the control group completed a non-verbal reasoning test. Roughly 1,000 students participated in the experiment.

Our findings show that the intervention had a positive and statistically significant effect on lecture attendance, with treated students attending one additional lecture over the term, on average. However, this increase in attendance did not translate into improvements in academic performance, as we find no significant effect on GPA. While greater attendance may increase exposure to course content, it may not be sufficient to improve outcomes — possibly due to variation in study habits, the effectiveness of self-directed learning, or simply because the attendance increase was too small to generate meaningful academic gains. We find no evidence of changes in students' daily time allocation across study, work, sleep, or exercise. That said, the treatment led to a notable increase in students' interest in their field of study.

The effect seems, on average, considerably stronger for students who chose to enter the commitment contract. However, since only students in the treatment group were given this option, we cannot interpret these differences causally. To address this, we use a question from a follow-up wave of the survey to construct a suitable control group. Two months after the intervention, students assigned to the control group during the goal-setting experiment were asked whether they would have opted into the same commitment contract had they been given the opportunity. Based on their responses, we identify a control group of comparable size — approximately 130 students in each group — with nearly identical average characteristics. We find that students who opted into the commitment contract experienced a substantially larger increase in lecture attendance, equivalent to roughly three additional lectures over the term. However, consistent with the broader pattern of results, we see no effect on academic performance.

To better understand who benefits from goal-setting interventions and why, we draw on the theoretical model developed by Koch and Nafziger (2011), which we adopt as the primary framework for interpreting our findings. The model posits that individuals who are present biased — placing disproportionate weight on immediate costs — can use self-imposed goals as internal commitment devices, particularly when they are also loss averse, since failing to meet a goal generates a psychologically salient loss.

Our results are partially consistent with the model's predictions. As expected, we find that students with low self-regulation skills respond strongly to the intervention. However, contrary to the model, we do not observe stronger treatment effects among loss averse students. One possible explanation lies in the importance of using domain-specific measures, as emphasized by (Augenblick et al., 2015). We proxy present bias through planning efficacy, an index derived from survey questions capturing students' ability to plan ahead and manage their workload — traits directly relevant to academic effort. In contrast, our measure of loss aversion is based on a monetary task, which may not adequately reflect students' sensitivity to failure in educational contexts.

Our study makes several contributions to the growing literature on goal-setting interventions. First, we provide new evidence on the effectiveness of task-based goals (as setting attendance targets is). Consistent with previous findings (Clark et al., 2020; Brade et al., 2018), we show that while these goals can effectively drive first-order effects — such as increasing attendance — they struggle to affect second-order outcomes. That is, their impact tends to be confined to the specific behavior being targeted, without necessarily translating into wider academic improvements.

Second, we reinforce the idea that goal-setting benefits students with weaker self-regulation skills. Like Brade et al. (2018), we find that students who struggle with planning and self-discipline experience the largest positive effects from task-based goal interventions. This supports the notion that goal-setting works best as a behavioral tool for those most in need of external structure and commitment.

Our findings also contribute to the broader education literature by questioning the commonly assumed link between lecture attendance and academic performance. While previous research suggests that greater attendance should lead to improved outcomes (Delavande et al., 2023; Dobkin et al., 2010), we find no evidence of an effect on GPA. This result may reflect the non-linear and context-dependent nature of the attendance-performance relationship, as highlighted in recent studies showing that increases in attendance do not always translate into better academic results (Goulas et al., 2023; Kapoor et al., 2021). These findings may suggest that the student production function may be more complex than often assumed, and that engagement alone is not always sufficient to improve achievement.

1.2 Theoretical Foundation and Current Evidence

Koch and Nafziger (2011) develop a formal model to explain when and how goal setting can serve as an effective tool for improving individual performance, particularly in the presence of behavioral frictions. Their framework draws on core insights from behavioral economics, incorporating two well-documented features of decision-making: present bias and loss aversion.

The model assumes individuals are sophisticated but time-inconsistent, meaning they are aware that their future selves may fail to follow through on plans made today. This is captured using the quasi-hyperbolic discounting framework (O'Donoghue and Rabin, 1999; Laibson, 1997), in which individuals place disproportionately high value on immediate costs and benefits compared to future ones. As a result, they may underinvest in effort today, even when they recognize the long-term benefits of doing so.

To address this self-control problem, the model allows individuals to set self-imposed goals that act as internal commitment devices. These goals create reference points in the utility function, meaning that performance is evaluated relative to the stated goal, rather than in absolute terms. The psychological power of this mechanism comes from loss aversion—a core principle of prospect theory (Kahneman and Tversky, 1979) — which holds that individuals experience the pain of losses more acutely than the pleasure of equivalent gains. In this framework, failing to meet a goal is experienced as a loss, while exceeding it is seen as a gain. Because individuals are loss averse, they exert additional effort to avoid the psychological discomfort of underperforming relative to their goal. This mechanism makes goal-setting particularly effective for individuals who are both present biased and loss averse, as the pain of falling short helps offset their tendency to procrastinate or discount the future.

However, the model also highlights important limits to goal-setting. If goals are perceived as too ambitious or unattainable, they may be ignored altogether or lead to demotivation. Thus, the effectiveness of goal-setting hinges on careful calibration: goals must be challenging enough to trigger the motivational force of loss aversion, but still seen as within reach.

A growing body of empirical research has tested these predictions in educational settings, though the findings are mixed. Dobronyi et al. (2017) report no effect of GPA-based goals among university students, even when paired with reminders. In contrast, Brade et al. (2018) study a non-binding goal agreement paired with reminders and find reduced dropout and faster degree completion, though no effect on GPA. Notably, effects are stronger for students who apply closer to the admission deadline, who are more likely to exhibit present bias — a finding confirmed by Himmler et al. (2019) using the same setting.

Several studies highlight the importance of the type of goal. Clark et al. (2020) find that task-based goals (e.g., completing problem sets) are more effective than performance-based goals (e.g., aiming for a certain grade), though the effects differ by gender. Similarly, van Lent (2019) and Kaiser et al. (2021) stress the dangers of setting goals that are too rigid or unmodifiable, especially early in a course when effort requirements are uncertain. Islam et al. (2020) observe no improvement in test scores from performance-based goals in Tanzanian high schools, though they find positive effects on effort, time use, and self-discipline — especially among students from disadvantaged backgrounds. In contrast, van Lent and Souverijn (2020) show that performance goals can be effective when students are supported by mentors, suggesting that interpersonal reinforcement may be key.

Finally, consistent with the theoretical insights of Koch and Nafziger (2011), Liu (2019) finds that narrow, bracketed goals work better than broad, long-term objectives. Importantly, many of these findings come from settings outside standard academic tasks, such as lab-based real-effort experiments, which may limit their generalizability to university coursework.

Taken together, the literature supports several core insights from the theory. Goal-setting can act as a powerful internal commitment device, especially for individuals with poor self-regulation. The structure and framing of the goal matter: task-based goals appear more effective than performance-based ones, and flexibility can enhance motivation. Our intervention draws on these insights, targeting a specific behavior—lecture attendance—that students often struggle to sustain, and embedding the goal-setting process in a supportive environment that includes feedback, reminders, and an optional commitment device.

1.3 Data and Experimental Design

1.3.1 The BOOST2018 Study

We use data from the BOOST2018 longitudinal study, which followed the 2015 cohort of bachelor students in a UK public university until graduation. The study aimed to improve our understanding of how students spend their time during their university careers. It focused on their attitudes, expectations, goals, and attainments.

The survey was widely promoted across the university campus to maximize enrollment. Of the 2,619 eligible students, 1,978 agreed to participate, representing both Home (UK residents) and International (EU and Overseas) backgrounds. As shown in Figure 1.1, which also provides a monthly breakdown of the academic year, they were interviewed 12 times in total - one in the Autumn term (November), two in the Spring term (January and March), and one in the Summer term (May). Most of these interviews were performed online, lasting either 60 minutes (the November and March sessions) or between 10 and 20 (the May sessions). The three January sessions involved attending the Social Science Experimental Laboratory of the university to participate in an experiment. While the online questionnaires aimed to gather information on students' study hours, attendance, and study habits, the lab surveys enabled the collection of experimentally elicited cognitive and non-cognitive traits and the delivery of randomized interventions.[3] The survey data were linked to administrative records held by the university, containing information on the demographics of the students, their socioeconomic status, and their degree courses. Moreover, we also utilize the weekly records of attendance to lectures and classes, which the university collects through a swipe-card electronic system.

[3]To encourage students to sign up for the survey, we offered a £5 incentive upon registration. In addition, participation in the various sessions was rewarded with monetary compensation—ranging from £8 to £20 for online surveys and averaging around £30 for laboratory sessions.

1.3.2 The Goal-Setting Experiment

In January of their second year at the university (2017), coinciding with the start of the Spring term, students participating in the study were invited to the Social Science Experimental Laboratory on the university campus. The sample of potential participants consisted of 1,871 individuals, of whom 1,063 attended the lab.[4] They were assigned to either a control or treatment group, using stratified random sampling (based on observable characteristics from the administrative dataset and their assignment and participation to treatment and control groups in an earlier RCT conducted in January 2016, during the student's first year). The day and time of the sessions were randomized between the treatment and control groups and there were always 2 control and 2 treatment sessions per day.[5] Students who came to the lab sat in individual partitioned booths with their own computer screens and noise-canceling headphones. The sessions were rolled out starting from the 23rd of January 2017, until the 15th of February 2017, covering the 2nd, 3rd, 4th, and one day of the 5th academic week in the spring term. Over these four weeks, 49% of the students attended in week one, 31% in week 2, 17% in week 3, and 3% in an originally unscheduled week 4, added to increase participation.

Most of the session was common to treated and control participants. Initially, both groups were asked about their enjoyment of lectures and classes.[6] At the start and end of the session, all students were asked to complete a set of incentivized tasks designed to elicit several cognitive and non-cognitive traits.[7] The rest of the session was different from the treatment and control groups.

[4]The change from the initial 1,978 accounts for those who dropped out from their studies or the survey participants pool, and those who joined the survey in their second year at the university.

[5]The email invitations offered each group a different menu of sessions, such that each session could only be booked by students from the same treatment group. Students asking to take part in a session they were not offered (but available to their friends) were told that this was because the session included a competitiveness task where they were asked to compete against other participants and we wanted to minimize the chance of people who knew each other well being paired together. This explanation was always accepted.

[6]We measured enjoyment using a "willingness to accept" question on how much they would want to be paid to trade one hour of leisure for one hour of lecture or class, depending on schedule and whether an assignment was due the following day (assuming that would not have any effect on their final grade).

[7]This took about 35 minutes in total. These tasks were identical for the treated and control groups and the average payoff was very similar.

The control group - Students in the control group were asked about their expected attendance rates at lectures and classes. Then, they had to answer 10 multiple-choice questions based on a non-verbal reasoning test (where the goal was to find the missing element of a pattern). They could choose between an "easy" and a "difficult" set of questions. Both sets would get progressively harder. They would receive compensation for each correct answer, corresponding to £2 if they selected the "easy" set and to £4 if they picked the "difficult" one. The amount of time available to complete this task was 8 minutes.

The treatment group - We first informed students in the treatment group of the benefits of attendance for their performance in course assignments and final exams and told them that setting a goal could help motivate them not to skip any academic event. To reinforce the argument, we showed them a 2-minute video named "One-step-at-a-time - goal achieving cartoon doodle video". The clip explained very intuitively the importance of goal-setting in improving the chances of achieving what you aimed for.

Each student was then shown their own attendance rate at lectures and classes, during the preceding autumn term. They were then asked to set a goal for their attendance, first for lectures, then for classes, in the remaining weeks of the spring term. They were assured that the attendance goals they set would be just for themselves and would not be seen by lecturers or teaching staff. These goals were reported in terms of the percentage of lectures and classes they will attend each week for the remaining weeks of this term.

Next, we had them write an essay to reflect briefly on the reasons that led them to miss some of their academic events, whether that had negative consequences, and how to avoid repeating that same mistake. As a reward for their writing, they would receive £1.50 for every 200 characters of coherent text they wrote (up to an amount of £15). They were required to spend at least 3 minutes planning, and 10 minutes writing this essay.

Students were next told they would receive feedback on how they are doing with their goals and reminders to help them hit their target over the next weeks. They were told that this would be measured using the count-me-in records (based on the swipe cards) and could choose to receive these by text message or by email.

Finally, we offered them the possibility of entering a commitment contract. Accepting the offer meant that the students would give up a portion of what they earned that day at the lab if they did not stick to the attendance goal that was previously set.[8] Finally, we asked the students how committed they were to the goals they set on a scale from 0 to 100. As for the control group, we also asked them about their expected attendance rates at lectures and classes.

However, the treatment continued as, at the beginning of each week, each treated student received a message reminding them of the goal they had set, revealing their attendance during the previous week, and stating whether they managed to achieve their attendance goal. The messages were different depending on whether they reached their goal, barely failed it (actual attendance was within 10% point of their goal), or failed it by a significant margin.[9]

1.4 Empirical Strategy

1.4.1 The Outcome Variables

The experiment aims to verify if goal-setting improves students' attendance as well as other outcomes that might be directly affected, such as academic performance, sense of belonging, and interest in the field of study. By increasing attendance, we may displace other academic inputs, such as study hours. We change the structure of our data depending on the outcome studied. In the case of attendance, since it is measured in each of the 10 weeks that are part of a term, we have 10 observation per student. Other variables are

[8] The amount of money put into the contract was selected by the student without knowing the exact amount earned using a list of payments (e.g., "If I have got £[Y] for my essay, I want to allocate a £[X] amount to my commitment contract" where Y varies from 1 to 10 and X is any amount chosen by the student).

[9] The message content depended on whether the attendance goal was achieved for both classes and lectures, whether it was achieved for only one of the two, or whether it was not achieved for both. If the goal was reached it tried to motivate the student to keep going ("You reached your goal! Now do the same this week!"). If it was not achieved by less than 10 percentage points, or if it was not achieved by a larger margin, it tried to push the student to do better during the following week (the final part of the message stated "You almost reached your goal! You can do it this week!" in the first case and "You have not reached your goal! You can do it this week!" in the second one).

measured on a year-to-year or term-to-term basis, so that only one observation per student is needed. In particular, students take their final exams each academic year after the end of the Summer Term (between May and June). Variables referring to students' well-being and time allocation are collected in the sessions of the survey taking place around the Autumn term of 2016 (end of October) and the end of the Spring term of 2017 (end of March, hence roughly one and a half month after the experiment), respectively.

Attendance: We obtained administrative records of students' timetables as well as weekly records of their attendance from a swipe-card electronic system, which allows us to derive measures of attendance that are not affected by self-reporting.[10] Attendance is hence the percentage of course-related events attended by a student concerning all the events they are expected to attend. We use both the aggregate category "events" and distinguish between "lectures" and all other event types, which we call "classes". Lectures involve the more theoretical and passive part of the teaching, where typically all the students registered on a particular module are taught together. The other event types focus on the practical and interactive side of teaching, and students are typically taught in smaller groups.[11]

Academic Performance: Performance variables are those that involve marks and the academic progression of a student. Students' marks are scaled from 0 to 100. A "year mark" is a weighted average of marks scored in between 4 and 8 "modules" (also referred to as "courses") that together make up an overall degree program. We measure the intervention effect on the student's "year mark" (or GPA), the percentage of modules that are passed, the percentage of grades that are above 70/100 points (which is the threshold for a "first class" grade in the UK system), on their degree mark and on whether they did graduate in time. Excluding the last two, all the other performance variables refer to the outcomes measured at the end of the second academic year and are computed

[10] The swipe-card system was put in place to record non-EU students' compliance with their visa and immigration requirements. The measures of attendance obtained using administrative records are not error-free (e.g. students may forget their swiping card), but they are such that it is much more unlikely that the measurement errors are correlated to individual characteristics.

[11] We included in the latter category "classes", "support classes", "laboratory sessions", "seminars", "tutorials" and "workshops". We also look at the effect on the attendance to "core events" and "core lectures": "core" indicates modules that a student has necessarily to pass to advance to the following year.

using only marks assigned in the exams undertaken after the intervention.[12]

Measures of Well-Being: We exploit each iteration of the online questionnaire to ask students how satisfied they are with their relationship with the university, their fellow students, the staff, and the course of choice. Satisfaction is captured by simply asking the participants, on a scale from 1 to 100, "All things considered, how satisfied are you with your life at University as a whole nowadays?". Similarly, and on the same scale, we ask them "How interested are you in your field of study?". Both the questions are asked in the Autumn term of 2016 - survey of November 2016 - and in the Spring term of 2017 - survey of March 2017.

Time Allocation: Hours of study are self-reported from the question "Not counting hours spent in class and lectures, how many hours in a typical week during term time do you usually study?". Hours of sleep/work/exercise are derived from the time diary students are asked to complete during the online surveys; specifically, they had to answer the following question "We now would like you to fill a time diary which is a list of activities during the week. If today is Saturday, Sunday, or Monday, please report your activity starting at 6am last Thursday. If today is Tuesday, Wednesday, Thursday or Friday, please report your activity starting at 6 am yesterday.". [13] Both the questions are asked in the Autumn term of 2016 - survey of November 2016 - and in the Spring term of 2017 - survey of March 2017.

1.4.2 Empirical Specification

We use a simple OLS regression for the analysis and run two different specifications. We include analytical weights that account for the number of events each individual was expected to attend during a certain week when the dependent variable is related to attendance. This is to ensure that weeks in which fewer events are scheduled are given

[12] The share of passed modules and the degree mark may, in small part, reflect marks for coursework and mid-term tests taken in the autumn term of Year 2. First-year marks do not count for the final degree marks and no modules extend across both the first and second years.

[13] They could choose to fill each of the 24 hours in a day with one of the following activities: "Class Attendance", "Lecture Attendance", "Club, Society or Association Attendance", "Commuting/Transport", "Eating", "Exercising", "Partying", "Personal (Hygiene, Laundry, etc)", "Recreation (alone)", "Recreation (with Friends)", "Shopping", "Sleeping", "Studying", "Working", or "Other".

lower weights.[14] We cluster standard errors at the individual level.

We begin by regressing the treatment dummy T on the dependent variables of interest, which will be our equation (1). Then, in equation (2), we add a control for the previous level of the dependent variable; for attendance variables, we include week fixed effects W. To improve the precision of our estimates we include a vector that controls for the variables that were used in the stratification process X in equation (3) (McKenzie, 2012; Bruhn and McKenzie, 2009). We report specification (3) below, first for attendance Equation 1.1 and then for performance Equation 1.2. As we anticipated, the former contains subscripts for individual student, term and week, the latter only for individual and term.

$$Y_{iwt} = \beta_0 + \beta_1 T_i + \beta_2 Y_{iwt-1} + \beta_3 \boldsymbol{W}_{iwt} + \beta_4 \boldsymbol{X}_i + \epsilon_{iwt} \tag{1.1}$$

$$K_{it} = \beta_0 + \beta_1 T_i + \beta_2 K_{it-1} + \beta_3 \boldsymbol{X}_i + \beta_4 C_i + \epsilon_{it} \tag{1.2}$$

1.4.3 Restrictions Based on Treatment Timing

A key feature of this study is that the treatment is provided over 3 weeks, following the first week of the term. The staggered treatment timing is very likely related to some unobserved student characteristics (we will discuss heterogeneous treatment effects based on the week of attendance later). Comparing the average attendance of treated and untreated students over all the weeks of the term would include also weeks in which all or some of the students did not participate in the experiment.

For each specification involving attendance, we run Equation 1.1 using two different restrictions. The first restriction, named “Post Week 16”, includes all the weeks after the 1st of the term, when the experiment began. The second restriction, called “Post Week 19”, includes all the attendance records taken after the end of 4th week of the term when the experiment was over.

[14] Naturally, the weight changes with the dependent variable. We use the number of “events” in a given week when we estimate the treatment effect on attendance to “events”, the number of “lectures” in a given week when we estimate the treatment effect on attendance to “lectures”, and so on.

1.5 Sample Characteristics

1.5.1 The 2015/2016 Cohort

Participation in the BOOST2018 study was open to all the BSc students of the 2015/2016 entry cohort, which counted up to 2619 individuals. The available demographic data allows us to paint a first picture of our first-year students. The group is fairly balanced in terms of gender, as 49.75% of the individuals are men and 50.25% women. Overall, slightly less than 10% of the students are "mature students", i.e. older than 21 years old. Looking at the type of tuition fee paid allows inferring their area of origin: home, EU and overseas students account for 67%, 17%, and 16% of the first years, respectively.

We establish students' socioeconomic status based on the occupation of the parents, which is available for 81% of the 1790 Home students. We categorize it 'high' for 51% of them (higher or lower "managerial and professional", and "intermediate" occupations, or categories 1 to 3 of the 8-point Office for National Statistics Socio-Economic Classification) and low for the remaining 30% (all other occupations). 59% of the students are White; those of Asian and Black ethnicity include 16% of the population each and in the remaining 9% we aggregate all the other ethnic minorities.

For the sake of simplicity, in this descriptive section, we will group departments into 3 macro categories, corresponding to the faculties in place at the University: Humanities, Science and Health, and Social Sciences.[15] They account for 28%, 32% and 40% of the students, respectively. More detailed data on the distribution by the department are available in Table 1.1.

We also have data on the distribution of the tariff score of each student. This indicates students' performance in pre-university qualifications, which we standardized among the eligible population at the study university.[16]

[15] Humanities includes degrees in Art History, History, Interdisciplinary Studies, Law, Literature and Philosophy. Sciences and Health groups programs in Biological Sciences, Computer Science, Health and Human Sciences, Maths, and Psychology. Social Sciences includes courses in Business, Economics, Government, Language and Linguistics, and Sociology.

[16] The tariff points are available through the linkage with the university administrative data and come from the Universities and Colleges Admissions Service (UCAS). The UCAS Tariff points are a way of

1.5.2 BOOST2018 Participants' Characteristics

Since participation in the survey and its various sessions was not compulsory, we will now present the differences, in terms of demographic characteristics, between the population and the participants in different parts of our experiment. All the figures are available in Table 1.2. 75% of our freshmen (1978 individuals) initially decided to participate, but 109 students dropped out from the survey by the end of the year. Since our analysis is focused on an experiment that was run during the second academic year (to be precise, in spring 2017), we excluded the dropouts from the comparison. Conversely, we include those who signed up for the survey in their second year. Overall, we are left with 1876 individuals.

The comparison between these 3 groups highlights several differences. Our population is almost perfectly balanced in terms of gender. However, female students tend to participate slightly more in the survey, especially in the January 2017 session, as they account for almost 58% of the sample. Other relevant differences we can highlight involve EU students, who were also more likely to come to the lab for the goal-setting experiment. The opposite was true for overseas students. Also, students of black ethnicity were slightly more represented in those who participated in the January 2017 session compared to their share in the population (18% vs 16%), as those who attended a public high school (36% vs 38%). Similarly, the share of those belonging to the bottom tariff quintile was 2 percentage points higher in the session participants than in the whole cohort.

We opted for the exclusion of those students who attended the lab during the 4th, unplanned week of the experiment. We show in column (4) the mean characteristics after excluding that particular subgroup. The comparison with column (3) highlights that the exclusion does not materially alter the observable characteristics of our estimation sample.

comparing the value of all post-16 qualifications in the UK, as students can access university by gaining academic qualifications, vocational qualifications or a mixture of the two. The total score is obtained by assigning a numerical value to each grade and qualification and summing these up. The higher the grade the student achieves per each qualification, the higher the number of points awarded. The standardized tariff score and quintiles shown are derived with respect to the population of enrollees at the study university who have a non-missing tariff score. This includes non-British students who took UCAS-recognized qualifications.

1.6 Balance Checks and Placebo Test

We will now present the mean comparison between the demographic and baseline behavioral characteristics of the treatment and control groups. While the former came either from the university database or from the participation questionnaire the students filled in when enrolling in the survey, we gathered the latter over the first academic year and the first term of the second academic year. We randomized students into treatment and control groups using stratified random sampling. The stratifying variables are gender, age, department, socioeconomic status, tariff quintile, the fee type, and their assignment and participation to treatment and control groups in an earlier RCT conducted in January 2016, during the student's first year.

As we can see from Table 1.3, the two groups are quite balanced. The only noticeable differences (at the 10% significance level), are in the share of students that attended a private high school (5.3% in the control group, 8.3% in the treatment group) and in the share of students from the faculty of Humanities (30% in the control group, 25% in the treatment group), which are approximately compensated by those from the faculty of Science and Health (30.5% in the control group, 34.8% in the treatment group), although this difference is not statistically significant.

The randomization on administrative characteristics does not result in statistically significant differences in any other dimension: Table 1.4 shows that the two groups are equal in terms of the share of students that participated in the January 2016 lab experiment and in the share that were assigned to the treatment or the control group. First-year marks, and attendance to events and lectures during the previous terms, are also not statistically different.

In Table 1.4 we also see that we achieve balance in terms of several behavioral characteristics measured in previous waves of the survey. We elicited present bias, discount rate, overconfidence, and grit during the survey of wave 2.[17] Except for overconfidence,

[17]They were calculated based on the outcomes of a real choice problem that students encountered in the wave 2 survey. We detail the method through which the main non-cognitive traits we use in our heterogeneity analysis, in the appendix.

which was measured for all the participants in the goal-setting experiment, the other indicators are not available for 23% of our sample. Data on openness, conscientiousness, extroversion, agreeableness, and neuroticism were collected in wave 3, and they are available for more than 97% of the students in our wave.[18] We also measured risk aversion, loss aversion, and willingness to pay for lectures and classes during the wave of January 2017. The former is based on a series of tasks that asked to commit to certain bets based on the payoffs that were offered, the other two report how much students would require to be paid to attend an extra hour of lecture or class. We have observations for all the students.

To further assess the balance between the treatment and control groups, we conduct a placebo check, estimating the effect of the treatment on pre-intervention outcomes. Since no treatment has occurred yet, we expect to find no significant differences if the groups are truly comparable. We do that for all the outcomes we study, finding no evidence of pre-treatment differences between the groups. We show these estimates in Appendix A.1. This supports the validity of our random assignment and strengthens the case for a causal interpretation of our results.

1.7 Context and Descriptive Evidence

1.7.1 Motivating Evidence: The Link Between Academic Performance and Attendance

Improving student attendance is particularly important given its strong association with academic performance and its uneven distribution across ability groups. As shown in the top panel of Figure 1.2, students with higher high school achievement tend to maintain higher attendance rates throughout their university studies, with those in the top tariff quintile consistently attending more than 5 percentage points more than those in the bottom tariff quintile. This gap suggests that students with weaker academic backgrounds

[18] We calculate these indicators, also known as the "Big 5" Personality Traits, using 15 questions from the questionnaire developed by Goldberg (1992).

— who might benefit the most from consistent classroom engagement — are the ones most likely to miss out on its advantages. The bottom panel further illustrates the connection between attendance and academic success, showing that students who attend a greater share of lectures in their first year of study achieve better marks in their end-of-year exams. Together, these patterns underscore the dual nature of attendance, which is both an input of students' production function and a sign of their commitment to their academic career. Informed by preceding research providing evidence of the former (Delavande et al., 2023; Dobkin et al., 2010; Romer, 1993), it highlights how students at greater academic risk could benefit from participating more in academic activities.

1.7.2 Goal-Setting Patterns and Achievement

We examine the characteristics of the attendance goals set by students in the treatment group. The top panel of Figure 1.3 shows a clear pattern: students who had the lowest attendance rates in the preceding term tended to set the most ambitious goals, aiming for improvements averaging around 40 percentage points. In contrast, students who had already achieved high attendance in the autumn term set more modest targets, often only slightly above their previous performance. This is indeed a consequence of the attendance rate being naturally bounded, but also speaks to the feasibility of the goals, a key factor emphasized in theoretical models such as Koch and Nafziger (2011), which warn against the demotivating effects of setting goals perceived as unattainable.

This intuition is confirmed by the patterns of goal achievement. Students with initially high attendance were significantly more likely to meet their weekly goals, suggesting that goal attainability plays a central role in sustaining motivation. The bottom panel of Figure 1.3 highlights how our interpretation of success matters greatly: if we define success as never failing to meet one's goal in any week, less than 5% of students achieve this until the final week of the term. However, looking at weekly goal achievement, we observe that between 40% and 50% of students meet their goal in any given week up to week 8, after which achievement rates drop below 40%, and then below 30% in the final weeks of the term.

Taken together, these patterns offer a key insight: students set goals that appear responsive to their previous behavior, with those who had low attendance in the prior term setting the most ambitious targets. While this reflects a desire to improve, the descriptive evidence also suggests that such ambitious goals may be less likely to be achieved consistently, particularly as the term progresses. This is consistent with theoretical concerns regarding goal feasibility, as highlighted by Koch and Nafziger (2011), and with prior empirical findings (Islam et al., 2020; van Lent, 2019; Dobronyi et al., 2017) showing that overly demanding or rigid goals can be counterproductive when they exceed what students perceive as achievable.

1.7.3 Measuring Present Bias and Loss Aversion

To interpret the effects of our intervention through the lens of the theoretical model presented in Koch and Nafziger, 2011, it is essential to capture the behavioral traits the model identifies as central — namely, present bias and loss aversion. These traits shape how individuals respond to goals, particularly in contexts involving effort and self-control.

Planning Efficacy

According to the theoretical framework, goal-setting is particularly effective for individuals who are present biased, as goals can help mitigate their tendency to procrastinate and underinvest in effort. We need a reliable measure of present bias to evaluate this mechanism empirically. While present bias is often elicited through monetary choices, applying such measures in our context raises concerns. Our intervention is centered on an effort-based behavior —lecture attendance — rather than financial decision-making. Relying on time preferences in the monetary domain may thus fail to capture the relevant behavioral tendencies that affect academic effort.

This concern is supported by prior work showing that behavioral traits are domain-specific. For example, Augenblick et al. (2015) demonstrate that individuals may exhibit different time preferences when decisions involve money versus effort. Consistent with this insight, we adopt an effort-based proxy for present bias: a planning efficacy index derived from survey responses. The idea is that students who plan ahead, manage their workload, and avoid last-minute behaviors are less likely to suffer from present bias in the academic context.[19]

We validate this index in Appendix A.2.1, where we present two pieces of evidence supporting its interpretation as a domain-relevant measure of (non-) present biasedness. First, planning efficacy - measured in Autumn 2016, the term before the intervention - is positively correlated with attendance and GPA - the former is measured in the same term, the latter in the preceding academic year. Second, we show that students with higher planning efficacy are significantly more likely to book their lab session earlier, consistent with forward-looking behavior and reduced procrastination. Taken together, these patterns support the use of planning efficacy as a behaviorally meaningful, effort-based proxy for present bias in the context of our study.

Loss Aversion

Loss aversion — defined as the tendency to weigh losses more heavily than gains — is a central feature in models of goal-setting behavior. According to the theory, individuals who are loss averse should be especially responsive to goal-setting interventions, since failing to meet a self-imposed goal creates a psychologically salient loss (Koch and Nafziger, 2011).

To measure loss aversion in a riskless setting, we implement a variation of the classic mug experiment introduced by Kahneman et al. (1990). Upon entering the lab, all students find a mug on their desk. We randomly assign participants to one of two roles:

1. **Buyers**, who do not own the mug and are asked to report their *willingness to pay (WTP)* for it.

[19]We construct the planning efficacy index by averaging responses to four Likert-scale items (1 = "Strongly Disagree", 7 = "Strongly Agree") collected across multiple survey waves. We measure planning efficacy during the November and March waves of all three years. The items are: "I usually do my work assignment the day before it is due"; "I usually keep track of my work assignment on a schedule or planner"; "I do not need to plan ahead to get good marks"; "I often underestimate the time that will be required to finish a project". Responses are recoded where necessary so that higher values reflect greater planning ability.

2. **Sellers**, who are told they own the mug and asked to report their *willingness to accept (WTA)* to give it up.

At the end of the session, roles are reversed. We define the endowment effect as the difference between WTA and WTP, and use it as a proxy for loss aversion, following Gächter et al. (2022); Jefferson and Taplin (2011); Kahneman et al. (1990). Students whose WTA is equal to or below their WTP are classified as loss neutral, while those with a strictly positive difference are considered loss averse. 85% of the students fall into the latter category in our sample.

As discussed in Appendix A.2.2, our results closely align with the existing literature in terms of magnitude and distribution of the endowment effect. However, we find no systematic correlation between loss aversion and pre-treatment behaviors, reinforcing the importance of domain-specificity and suggesting that its effects may emerge only in interaction with the intervention.

1.8 Results

This section presents the main findings of our analysis. We begin by estimating the overall effect of the goal-setting intervention on key outcomes, including lecture and class attendance, academic performance, measures of subjective well-being, and time use during a typical weekday. We then explore heterogeneous treatment effects, focusing on the two behavioral traits central to our theoretical framework: planning efficacy, as a proxy for present bias, and loss aversion, as measured through the endowment effect. Finally, we examine treatment effects for the subgroup of students who opted into the commitment contract, providing insight into the added value of combining goal-setting with self-imposed incentives.

1.8.1 The Impact of the Goal-Setting Treatment: Attendance

We begin by estimating the effect of the goal-setting intervention on overall attendance at academic events, before distinguishing between lectures and classes. Results are reported in Table 1.5, where we estimate Equation 1.1 separately for two periods: the weeks following the start of the intervention ("Post W16"), and the weeks after its completion ("Post W19").

We find a positive and marginally significant effect on overall attendance, significant at the 10% level only. Although the effect size is larger when considering the "Post W16" period (2.6% vs. 3.3% in "Post W19"), this difference is mechanical: students had, on average, 64 academic events in the nine weeks following the start of the experiment, compared to 49 events in the seven weeks following its end. As a result, the treatment effect corresponds to approximately 1.7 additional academic events attended in the longer window and 1.6 in the shorter one.

When disaggregating events by type, we observe a clear distinction in treatment effects between lectures and classes. Lectures — which typically involve larger cohorts, are more theory-focused, and last about 120 minutes — exhibit a positive and statistically significant treatment effect, ranging from 3.3% to 3.7%, depending on the post-treatment window considered. In contrast, we find no measurable effect on attendance to classes - generally smaller, more interactive, and lasting about 60 minutes. Given that students were enrolled in an average of 36 lectures ("Post W16") and 28 ("Post W19"), these percentages imply that treated students attended roughly one additional lecture over the term — equivalent to about two more hours of teaching exposure.

1.8.2 The Impact of the Goal-Setting Treatment: Academic Performance, Well-Being, and Time Allocation

We next examine whether the increase in attendance brought about by the treatment translates into improved academic performance. Results are presented in Table 1.6. We find no significant effect of the intervention on students' GPA, the probability of passing more modules, or the share of exams passed with a score above 70/100 — the threshold for earning a first-class degree. Similarly, there is no effect on longer-term outcomes, such as the final degree mark or the probability of graduating on time.

To interpret these null results, we draw on prior estimates of the attendance-performance relationship in the literature. Dobkin et al. (2010) suggest that a 10% increase in attendance should lead to a 0.17 standard deviation increase in test scores, assuming a linear relationship. Likewise, Delavande et al. (2023) estimate that 10 additional hours of attendance in a term corresponding to a 0.15 standard deviation increase in GPA. Based on our observed effect — equivalent to one additional lecture, or roughly two extra hours of teaching — we would expect a GPA increase in the range of 0.035 to 0.08 standard deviations (i.e., 3.5% to 8% of one standard deviation). However, the estimated treatment effect on GPA is smaller and statistically indistinguishable from zero. These results may reflect the non-linearity of the relationship between attendance and performance, as well as the limited statistical power to detect small changes in academic outcomes. However, our null findings also align with studies suggesting that the relationship between attendance and achievement is complex and context-dependent. For instance, Kapoor et al. (2021) find that mandatory attendance policies for low-performing students raised attendance but had no positive effect on test scores. Similarly, Goulas et al. (2023) show that high-ability students benefit from greater flexibility in managing their study time. It is also possible that the increase in attendance we cause is too small to have an impact on the following exams.

Finally, we explore whether the intervention had any impact on students' well-being and time use. As shown in Table 1.7, there is no significant effect on students' overall satisfaction with university life. However, we find a small increase in students' interest in their field of study — about 2.6 percentage points, significant at the 10% level. In terms of time allocation, we find no evidence that attending more academic events crowds out other activities. The treatment had no measurable impact on hours spent studying, sleeping, working, or exercising during a typical weekday.

1.8.3 Heterogeneity Analysis: Present Bias and Loss Aversion

We now turn to heterogeneity analysis to explore whether the treatment effect varies depending on individual behavioral characteristics. This analysis is motivated by the theoretical framework discussed in section 1.2, which highlights two key mechanisms underlying the effectiveness of goal-setting: present bias and loss aversion. These traits shape how individuals respond to internal commitment devices, such as goals, by influencing their sensitivity to immediate costs and perceived failure. Beyond theory, studying heterogeneity by these traits provides insight into who is most likely to benefit from the intervention — an important consideration for the design and targeting of behavioral policies. In what follows, we examine how treatment effects differ across students with varying levels of planning efficacy, our proxy for present bias in the effort domain, and loss aversion, as measured through the endowment effect.

To examine whether the treatment effect varies with present-biasedness, we exploit our measure of planning efficacy, collected in the online survey conducted in November 2016, prior to the goal-setting intervention. While not all students who participated in the January 2017 experiment responded to the earlier survey, we observe a high rate of overlap, with approximately 90% of the experimental sample included. We divide students into terciles based on their planning efficacy scores — 41% fall in the bottom tercile, 27% in the middle, and 32% in the top — and estimate treatment effects by interacting the treatment indicator with tercile dummies. Results are presented in Table 1.8.

Descriptively, we observe stark differences across planning efficacy groups. Students in the top tercile attend lectures at rates nearly 7 percentage points higher than those in the bottom group, and they also perform slightly better on long-run outcomes, reporting higher degree marks, greater satisfaction with university life, and stronger interest in their field of study. However, when it comes to treatment effects, the picture reverses: it is students in the bottom tercile — those with the weakest planning ability — who respond most positively to the intervention. For this group, the goal-setting treatment leads to a statistically significant increase in lecture attendance, equivalent to an 8% improvement over the baseline, as well as a stark increase in interest in their field of study, of roughly 6.6%.

These findings are consistent with our theoretical framework, which suggests that goal-setting is most effective for individuals who struggle with self-regulation. The treatment appears to support those students who have difficulty planning ahead — precisely the group the intervention was designed to help. This suggests that goal-setting interventions can be effective in reducing behavioral gaps in academic engagement, particularly when targeted toward students exhibiting signs of present-biasedness in the effort domain.

We also explore heterogeneity by loss aversion, measured through the endowment effect elicited during the lab session via a variant of the mug experiment. While this approach follows standard methods to capture aversion to losses in a riskless setting, it primarily reflects preferences in the monetary domain, which may not perfectly map onto the effort-based context of academic goal pursuit. In our sample, the vast majority of students — roughly 85% — exhibit a strictly positive endowment effect and are thus classified as loss averse; about 15% are loss neutral.

The descriptive patterns in Table 1.9 suggest that loss averse students attend more lectures on average and perform significantly better academically — scoring around 5% higher in their second-year exams — though they paradoxically report lower interest in their field of study (by about 6%). In terms of treatment effects, the overall patterns observed in the full sample are largely driven by the loss averse group, whose responses closely mirror the average estimates. However, it is the small group of loss neutral students who appear to benefit more meaningfully from the treatment. For this group, we estimate a 7% increase in lecture attendance in the "Post W19" period, just shy of statistical significance at the 10% level, and a significant 6.3% improvement in GPA, along with a positive and statistically significant effect on final degree marks. By contrast, gains in interest in the field of study are observed only among loss averse students.

These findings are somewhat at odds with the theoretical model, which predicts that loss averse individuals should be more responsive to goal-setting, given their heightened sensitivity to falling short of a target. One potential explanation lies in the domain mismatch between how we measure loss aversion — through a monetary framing — and the effort-based nature of our intervention. It is possible that the endowment effect does not fully capture how students perceive and internalize "academic losses" such as missing a goal. Moreover, as shown in Table 1.10, excluding the small subset of non-loss averse students (approximately 15% of the sample) leads to only a slight decrease in the magnitude of the estimated effects. This suggests that the core treatment effect is attributable to the loss averse majority, consistent with the model's focus on this group. However, these results may reflect complex interactions between motivation, baseline performance, and intrinsic interest, which we are not fully able to unpack with the available data. Nonetheless, these patterns highlight the importance of carefully aligning behavioral measures with the domain of the intervention and suggest that loss aversion, while conceptually relevant, may not always operate in predictable ways across contexts.

1.8.4 Commitment Contract

We next look at the treatment effect among students who self-selected to take a commitment contract in the treatment group. As described above, students could put aside some reward money during the lab session and be entitled to collect it only if they reach their attendance goals. 26% of the students in the treatment group took up the commitment contract. They showcase higher attendance rates during the previous term, Autumn 2016 - 70.42% vs. 67.76%. The difference is similar when we compare them in terms of attendance to classes, 65.41% against 62.53%. On the other hand, we do not observe any significant gap in the mean of the first-year marks, 62.64/100 vs 62.49/100.

To create a credible control group for students who took up the commitment contract, we leverage a hypothetical question asked to respondents at the online wave of the survey held in March 2017, the one following the goal-setting intervention. To those previously assigned to the control group, we asked whether they would sign up for a commitment contract to have the incentive to maintain a certain level of attendance.[20] 31% of the students formerly in the control group were interested in the commitment contract, similarly to what we observed among treated students. As we can see in Table 1.11, the two groups are closely matched in terms of past attendance, GPA, and general demographic characteristics.

We then estimate the combined effect of the goal-setting treatment and the take-up of

[20]The overlap between the January 2017 and March 2017 survey sessions is 90%.

the commitment contract, by comparing those who took it, to those in the control group who would have done that if assigned they were assigned to the treatment group. As we can see in Table 1.12, treated students exhibit a higher attendance to lectures in both periods of interest. In fact, they attend 8.5% more lectures in the "Post Week 16" period and 10% more in the "Post Week 19". This corresponds to 3 more lectures attended over the term (as the total remaining lectures were 36 and 28, depending on which of the two periods we consider). Still, we do not find any statistically significant treatment effect on GPA or the well-being of the students.

1.9 Robustness Checks

In this section, we present a series of robustness checks aimed at testing the validity and interpretation of our main results. First, we address the possibility that the treatment may have influenced swiping behavior rather than actual attendance, given that our primary attendance measure is based on students swiping their ID cards upon entering academic events. Second, we assess whether our findings are driven by students' pre-existing preferences for attending lectures, by examining treatment heterogeneity based on the amount they would expect to attend an extra hour of lecture. This allows us to rule out the possibility that only students who already enjoy attending are responding to the treatment. Finally, we evaluate how our results change when we include students who participated in the fourth week of the experiment, who were excluded from the main analysis due to the shorter exposure to the intervention.

1.9.1 Impact on Swiping Behavior

A key concern when interpreting our main findings is that the treatment may have influenced swiping behavior rather than attendance itself, given that our attendance measure relies on students swiping their ID cards upon entering academic events. To probe this, we exploit the idea that the effort required to swipe may vary by lecture size: in larger lectures, students may have to queue to access the reader, which could discourage swiping even if they attend, while in smaller lectures this cost is likely minimal. If the treatment merely changes how students record attendance, we might expect stronger effects in lectures where swiping is less costly.

To test this, we re-estimate the treatment effect by breaking the data down at the lecture level and interacting treatment status with terciles of lecture size, defined by the average number of students attending each lecture. On average, lectures in the first tercile include 65 students, in the second 150, and in the third 234. We adapt our main specification to a difference-in-differences setup that compares attendance changes from the autumn to the spring term between treated and untreated students. Specifically, we estimate:

$$Y_{liwt} = \beta_0 + \beta_1 \mathrm{T}_{iw} \times \mathrm{S2}_{lwt} + \beta_2 \mathrm{T}_{iw} \times \mathrm{S3}_{lwt} + \beta_3 \boldsymbol{W}_{wt} + \gamma_i + \epsilon_{liwt} \tag{1.3}$$

where Y_{liwt} is a binary indicator Indicator for student i's attendance at lecture l, in week w, and term t. T_{iw} is a treatment indicator expressed at the week level, $\mathrm{S2}_{lwt}$ and $\mathrm{S3}_{lwt}$ denote the second and third terciles of lecture size (with the first tercile as the omitted category), $\boldsymbol{W}_{wt}$ includes week fixed effects, and γ_i denotes individual fixed effects. The regression is weighted by the expected number of lectures each student could attend.

The results in Table 1.13 suggest that lecture size does moderate the treatment effect: the attendance improvement does not originate from students attending more medium-sized lectures, but rather from an increased participation to small- and medium-sized lectures. Interestingly, the effect is largest for students attending larger lectures, which contradicts the idea that swipe costs alone drive the observed increase in attendance. While the treatment effects for the first and third terciles are not statistically distinguishable from one another, these findings reinforce the interpretation that the treatment affects actual attendance behavior, not just the likelihood of students recording it.

1.9.2 Is Enjoyment of Lectures the Key?

Second, we assess whether the treatment effect is driven by students' pre-existing preferences for lecture attendance, by examining treatment heterogeneity based on their reservation price for attending an extra hour of lecture. This measure, elicited at the beginning of the goal-setting experiment for all the participants, captures the amount of compensation each student would require to attend an additional lecture hour. We divide students into terciles based on their responses: students in the bottom tercile report an average of £3.92, those in the middle £9.34, and those in the top nearly £32.

As shown in Table 1.14, we find no clear pattern of heterogeneity: treatment effects are similar in magnitude across the three groups and not statistically significant in any of them. While effects appear slightly lower for students in the middle tercile, this difference is neither large nor robust. These findings suggest that the effectiveness of the goal-setting intervention does not depend on students' initial enjoyment of lectures, reinforcing the idea that the treatment affects behavior regardless of baseline preferences for attending.

1.9.3 Adding Week 4 Participants

Finally, we assess whether the exclusion of students who participated in the experiment during Week 4 materially affects our results. As discussed earlier, these students were excluded from the main analysis due to their shorter exposure to the treatment and their small number, which limits the reliability of any subgroup analysis. In Table 1.15, we re-estimate our main specifications including Week 4 participants and find that the results are virtually unchanged. This supports the robustness of our findings and provides additional justification for our initial sample restriction.

1.10 Conclusion

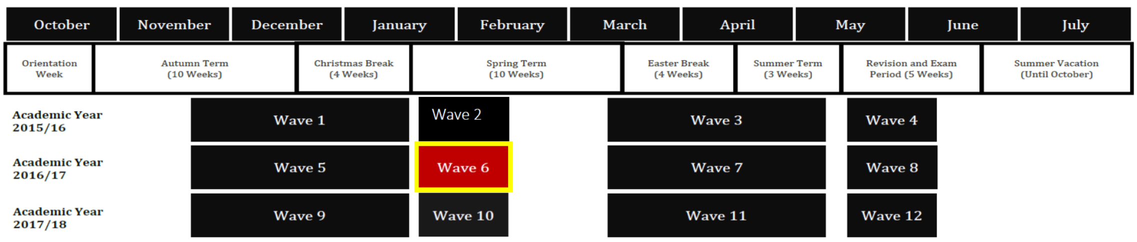
This paper evaluates a goal-setting intervention designed to improve attendance among university students. We find that the treatment led to a modest but statistically significant increase in lecture attendance, equivalent to roughly one additional lecture over the term. However, this change did not translate into improved academic performance, as we detect no significant effects on GPA, exam success rates, or longer-term academic outcomes. In line with recent studies such as Clark et al. (2020) and Brade et al. (2018), our findings suggest that task-based goals can successfully influence specific behaviors, but that inducing downstream gains in academic achievement remains more elusive.

One of the most informative results comes from the heterogeneity analysis. Students with low planning efficacy — our proxy for present bias in the effort domain — are the ones who respond most positively to the intervention, increasing their attendance by 8% relative to baseline. This finding is consistent with the predictions of Koch and Nafziger (2011), who emphasize that goal-setting acts as a form of internal commitment device, particularly effective for individuals who struggle with self-regulation. In contrast, we find no meaningful pattern of heterogeneity based on our measure of loss aversion, which was elicited through a monetary endowment effect task. This divergence highlights the importance of domain-relevant behavioral measurement, echoing the concerns raised in Augenblick et al. (2015) about the context-dependence of time and loss preferences.

Although the treatment did not affect performance, we find marginal evidence of increased interest in students' field of study, particularly among those who were least likely to attend lectures in the first place. Importantly, our robustness checks confirm that the treatment effect on attendance is not driven by swiping behavior, nor by pre-existing enjoyment of lectures.

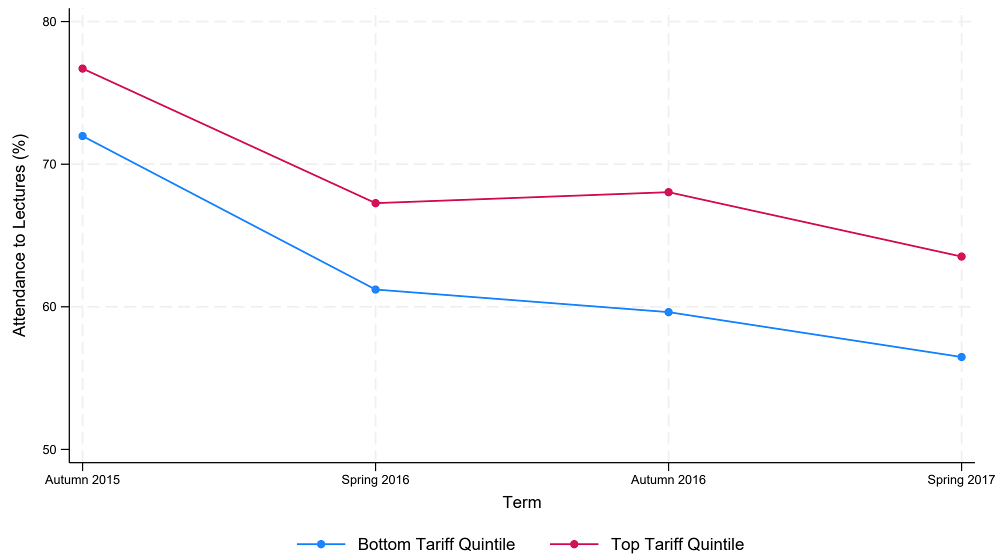
Taken together, our findings underscore the potential of behavioral interventions like goal-setting to influence concrete, observable outcomes in higher education — particularly when targeted toward students with lower planning ability. At the same time, they also highlight the limits of what these interventions can achieve in isolation. Future work should explore how to reinforce goal-setting with complementary supports, and how to design goal structures that are not only feasible, but more clearly aligned with broader academic progress.

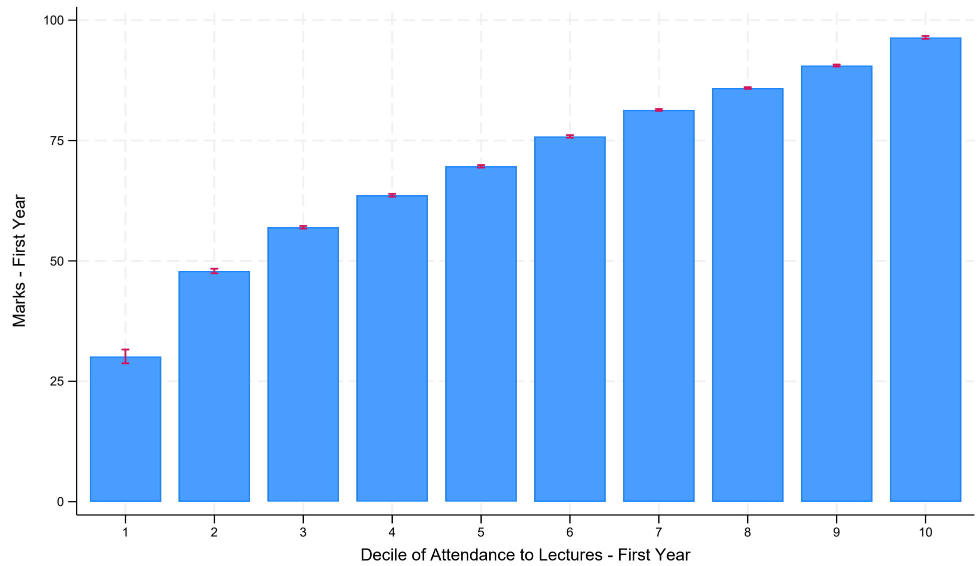
Figure 1.1: Structure of the BOOST2018 Study



Notes: The image summarizes the different waves of the BOOST2018 study, showing how they were distributed across the three years in which the students attended their bachelor's degree, from October 2015 until July 2018.

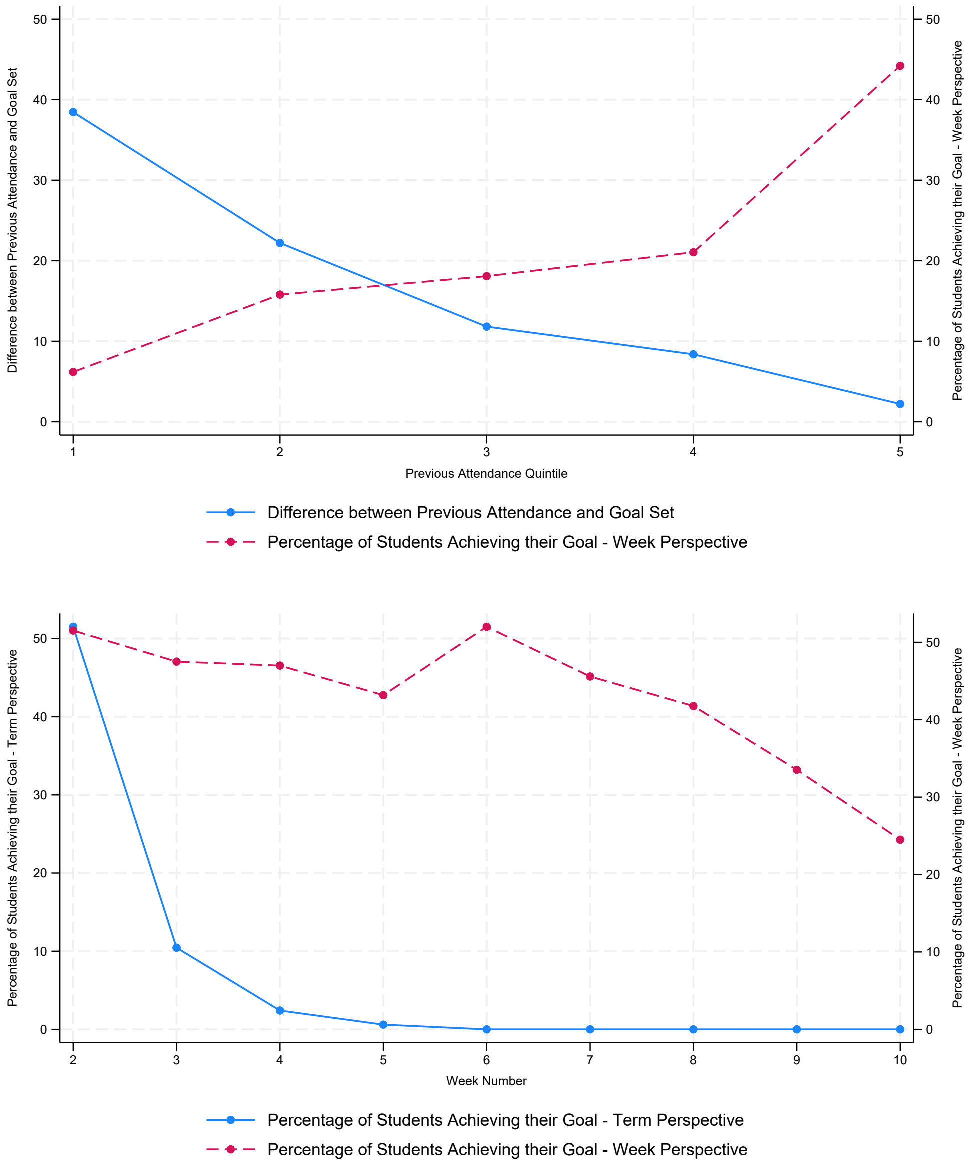
Figure 1.2: The Association between Academic Attainments and Attendance





Notes: The graph in the top panel showcases the evolution of the percentage of lectures attended by the students, based on whether they belong to the first or fifth tariff quintile. Attendance is measured as the weighted average of the percentage of the academic events classified as 'lectures' students attend over the 10 weeks of a term, where the weight is the number of these academic events during the week. The tariff quintile is calculated based on the UCAS Tariff points, which are a scoring system that allows comparing the value of all post-16 qualifications in the UK, and are derived with respect to the population of enrollees at the study university who have a non-missing tariff score. We report attendance as it is measured in the Autumn and Spring terms of the academic years 2015/2016 and 2016/2017, which are the first two that students in the sample spend at the university. The first tariff quintile includes 283 students, while the fifth includes 287. The graph in the bottom panel showcases the relationship between the marks obtained in the first year and the decile of attendance. Attendance is measured as the weighted average of the percentage of the academic events classified as 'lectures' students attend over the 10 weeks of the Autumn and Spring Term of their first year in higher education, where the weight is the number of these academic events during the week. We measure attendance for all students signing up to participate in the survey, for a total of 1,865 individuals.

Figure 1.3: Lecture Goal Description - Previous Attendance to Lectures



Notes: The graph in the top panel showcases, on the left-hand side (solid line), the path of the difference between the percentage of lectures attended in the Autumn term of the academic year 2016/2017 and the lecture goal set during the goal-setting experiment in the Spring term of the academic year 2016/2017 - where the former is provided as a baseline for attendance to the treatment group, during the experiment. On the right-hand side (dashed line), it shows the percentage of students achieving their goals weekly - as in, the average percentage of treated students who achieve their attendance goal each week, without considering their attendance at the end of the term. We break down the two variables by the quintile of previous attendance to lectures, calculated based on the percentage of attended lectures in the Autumn term of the academic year 2016/2017. The graph in the bottom panel shows, on the left-hand side (solid line), the percentage of students who are on track to achieve their goal at the end of the Spring term of the academic year 2016/2017, without having failed to do so in any week of the term. On the right-hand side (dashed line), it shows instead the percentage of students who achieve their lecture attendance goal, each week. We break down the two variables by the week of the term, starting from the week when the goal-setting treatment is implemented. We use the data for students who are part of the treatment group during the lab sessions of the goal-setting experiment, for a total of 498 students (we exclude from the distribution those attending during the 4th week of the goal-setting experiment).

Table 1.1: Share and Number of First Year Bachelor Students by Department

Department	Number of Students	% of Students
Art History	15	0.57
Biological Sciences	251	9.58
Business School	402	15.35
Computer Science and Electronic Engineering	232	8.86
Economics	224	8.55
Government	165	6.3
Health and Human Sciences	64	2.44
History	135	5.15
Interdisciplinary Studies	46	1.76
Language and Linguistics	108	4.12
Law	299	11.42
Literature, Film, and Theatre Studies	167	6.38
Maths	60	2.29
Philosophy	70	2.67
Psychology	230	8.78
Sociology	151	5.77
Total	2,619	100

Notes: We report the share of students in each university department. These are the students who enrolled in their bachelor's degrees starting from the academic year 2015/2016. The sample size is 2,619.

Table 1.2: Mean Comparison: Population vs Participants to BOOST2018 vs Participants to Wave 6 vs Sample of Interest

Variables	(1) Population Mean/SE	(2) Participants Mean/SE	(3) January 2017 Mean/SE	(4) Sample Mean/SE
Female	0.505	0.527	0.578	0.576
	(0.010)	(0.011)	(0.015)	(0.015)
Mature students	0.094	0.078	0.069	0.070
	(0.006)	(0.006)	(0.008)	(0.008)
Home Low SES	0.203	0.217	0.217	0.219
	(0.008)	(0.009)	(0.013)	(0.013)
Home High SES	0.346	0.365	0.362	0.369
	(0.009)	(0.011)	(0.015)	(0.015)
Home Missing SES	0.130	0.115	0.100	0.097
	(0.007)	(0.007)	(0.009)	(0.009)
EU SES	0.166	0.161	0.198	0.200
	(0.007)	(0.008)	(0.012)	(0.013)
Overseas SES	0.156	0.142	0.123	0.116
	(0.007)	(0.008)	(0.010)	(0.010)
White	0.582	0.563	0.577	0.586
	(0.010)	(0.011)	(0.015)	(0.015)
Asian	0.165	0.164	0.156	0.155
	(0.007)	(0.009)	(0.011)	(0.011)
Black	0.162	0.181	0.183	0.177
	(0.007)	(0.009)	(0.012)	(0.012)
Other Ethnicity	0.090	0.092	0.084	0.082
	(0.006)	(0.007)	(0.008)	0.009
Public School	0.359	0.383	0.384	0.390
	(0.010)	(0.011)	(0.015)	(0.015)
Private School	0.086	0.080	0.070	0.068
	(0.006)	(0.006)	(0.008)	(0.008)
University	0.017	0.012	0.008	0.009
	(0.003)	(0.003)	(0.003)	(0.003)
Other School	0.195	0.198	0.194	0.193
	(0.008)	(0.009)	(0.012)	(1025)
Top Tariff Quintile	0.144	0.153	0.146	0.146
	(0.007)	(0.008)	(0.011)	(0.011)
Bot Tariff Quintile	0.147	0.151	0.158	0.159
	(0.007)	(0.008)	(0.011)	(0.011)
Missing Tariff Quintile	0.260	0.237	0.247	0.251
	(0.009)	(0.010)	(0.013)	(0.014)
Humanities	0.279	0.279	0.279	0.277
	(0.009)	(0.010)	(0.014)	(0.014)
Science and Health	0.316	0.315	0.327	0.324
	(0.009)	(0.011)	(0.014)	(0.015)
Social Sciences	0.404	0.406	0.394	0.399
	(0.010)	(0.011)	(0.015)	(0.015)
Individuals	2,619	1,871	1,051	1,022

Notes: We report the shares of students belonging to different subcategories of the sociodemographic characteristics available from the university administrative data. In column (1) we show the characteristics of the population. In column (2) we report the characteristics of those who signed up for participation in the survey. In column (3) we have the characteristics of all the participants of the goal-setting experiment, in January 2017, during the Spring term of the 2016/2017 academic year. Finally, we include in column (4) the characteristics of those who participated in the goal-setting experiment during the first 3 weeks only.

Table 1.3: Balance Checks: Treatment vs. Control Group - Demographic Characteristics

Variables	(1) Control Mean/SE	(2) Treatment Mean/SE	(3) t-test of the difference (1)-(2)
Female	0.572	0.577	-0.005
	(0.022)	(0.022)	
Mature	0.076	0.065	0.011
	(0.012)	(0.011)	
Home Low SES	0.229	0.213	0.016
	(0.018)	(0.018)	
Home High SES	0.367	0.364	0.004
	(0.021)	(0.021)	
Home Missing SES	0.102	0.091	0.011
	(0.013)	(0.013)	
EU SES	0.188	0.211	-0.024
	(0.017)	(0.018)	
Overseas SES	0.114	0.121	-0.007
	(0.014)	(0.014)	
White	0.570	0.599	-0.029
	(0.022)	(0.022)	
Asian	0.159	0.148	0.011
	(0.016)	(0.016)	
Black	0.182	0.176	0.006
	(0.017)	(0.017)	
Other Ethnicity	0.089	0.077	0.012
	(0.012)	(0.012)	
Public School	0.403	0.374	0.030
	(0.021)	(0.022)	
Private School	0.053	0.083	-0.030*
	(0.010)	(0.012)	
University	0.011	0.006	0.005
	(0.005)	(0.003)	
Other School	0.206	0.184	0.023
	(0.018)	(0.017)	
Top Tariff Quintile	0.153	0.136	0.017
	(0.016)	(0.015)	
Bot Tariff Quintile	0.148	0.168	-0.020
	(0.015)	(0.017)	
Missing Tariff Quintile	0.254	0.247	0.007
	(0.019)	(0.019)	
Humanities	0.301	0.251	0.050*
	(0.020)	(0.019)	
Science and Health	0.305	0.348	-0.043
	(0.020)	(0.021)	
Social Science	0.394	0.401	-0.007
	(0.021)	(0.022)	
Individuals	528	506	

Notes: We report the average sociodemographic characteristics of the students - available from the university administrative data - based on their assignment to the treatment or the control group during the goal-setting experiment held in January 2017, during the Spring term of the academic year 2016/2017. We calculate the average of these characteristics only for those who participated in the goal-setting experiment during the first 3 weeks. In column (2) we report the characteristics of the treated group. The values displayed in column (3) are for the differences between the two means and the relative t-test. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table 1.4: Balance Checks: Treatment vs. Control Group - Additional Individual Characteristics

Variables	(1) Control Mean/SE	(2) Treatment Mean/SE	(3) t-test of the difference (1)-(2)
First Year Marks	61.254	61.854	-0.601
	(0.466)	(0.537)	
Year 1 Attendance: Spring Term (events)	62.293	63.314	-1.022
	(0.931)	(0.950)	
Year 2 Attendance: Autumn Term (events)	65.783	66.335	-0.552
	(0.871)	(0.904)	
Year 1 Attendance: Spring Term (lectures)	68.911	68.974	-0.063
	(0.980)	(1.011)	
Year 2 Attendance: Autumn Term (lectures)	68.525	68.445	0.080
	(0.941)	(0.969)	
Wave 2 Lab Participation	0.973	0.979	-0.006
	(0.007)	(0.006)	
Treatment Group in Wave 2 Experiment	0.502	0.491	0.011
	(0.021)	(0.022)	
Present Bias	0.194	0.188	0.007
	(0.020)	(0.019)	
Discount Rate	53.561	53.469	0.091
	(1.903)	(1.864)	
Overconfidence	15.233	15.847	-0.614
	(1.175)	(1.311)	
Openness	5.052	5.022	0.031
	(0.058)	(0.057)	
Conscientiousness	4.535	4.512	0.023
	(0.052)	(0.051)	
Extraversion	4.405	4.375	0.030
	(0.063)	(0.062)	
Agreeableness	4.896	4.868	0.028
	(0.057)	(0.056)	
Neuroticism	4.061	4.063	-0.003
	(0.065)	(0.065)	
Competitiveness	51.131	49.609	1.522
	(1.386)	(1.344)	
Planning Efficacy	4.441	4.395	0.046
	(0.047)	(0.047)	
Loss Aversion	0.869	0.833	0.036
	(0.015)	(0.017)	
Risk Aversion	0.251	0.246	-0.005
	(0.038)	(0.039)	
Willingness to Pay for Lectures	13.980	13.237	0.743
	(0.689)	(0.746)	
Willingness to Pay for Classes	14.872	14.920	-0.048
	(0.717)	(0.831)	
Individuals	528	506	

Notes: We report the average individual characteristics of the students - available either from the university administrative records or elicited during previous sessions of the survey - based on their assignment to the treatment or the control group during the goal-setting experiment held in January 2017, during the Spring term of the academic year 2016/2017. We calculate the average of these characteristics only for those who participated in the goal-setting experiment during the first 3 weeks. In column (1) we show the characteristics of the control group. In column (2) we report the characteristics of the treated group. The values displayed in column (3) are for the differences between the two means and the relative t-test. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table 1.5: Impact of the Goal-Setting Treatment: Attendance to Events, Lectures, and Classes

	(1)	(2)	(3)	(4)	(5)	(6)
			Attendance to:			
Variables	Events		Lectures		Classes	
Treatment	1.613*	1.863*	2.113**	2.347**	0.084	0.418
	(0.945)	(1.002)	(1.069)	(1.141)	(1.398)	(1.451)
Mean of the Outcome	61.15	58.58	64.58	62.21	56.80	54.47
Individuals	1,022	1,022	972	972	1,004	1,004
Observations	9,032	7,010	8,449	6,546	8,052	6,327
Stratifying Variables	X	X	X	X	X	X
Baseline	X	X	X	X	X	X
Week Fixed Effects	X	X	X	X	X	X
Period of Interest	Post W16	Post W19	Post W16	Post W19	Post W16	Post W19

Notes: We estimate the effect of the treatment in the weeks following the first and third week of the Spring Term of 2017 ("Post W16" and "Post W19", respectively). The dependent variables we use - all expressed on a scale from 0 to 100 - are the following: the percentage of academic events (columns (1) and (2)), lectures (columns (3) and (4)), or classes (columns (5) and (6)) attended by the students, and are collected through the swipe card system in place at the university. Academic events include both lectures and classes. We estimate the treatment effect using Equation 1.1. We include controls for week fixed effects, the stratifying variables used for the randomization process (sex, age, socioeconomic status, tariff quintile, and department), and the dependent variable's baseline - the percentage of events/lectures/classes attended during the previous term. Standard errors are clustered at the individual level. ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$.

Table 1.6: Impact of the Goal-Setting Treatment: Academic Performance

	(1)	(2)	(3)	(4)	(5)
	Academic Performance:				
Variables	GPA	SPM	SGG	Degree Mark	On Time
Treatment	0.094	-0.338	1.379	0.693	-1.534
	(0.540)	(0.623)	(1.420)	(0.525)	(2.158)
Mean of the Outcome	62.39	97.33	24.81	64.34	86,01
Individuals	982	1,000	1,006	948	1,017
Observations	982	1,000	1,006	948	1,017
Stratifying Variables	X	X	X	X	X
Baseline	X	X	X		

Notes: We estimate the effect of the treatment in the year of the goal-setting intervention. The dependent variables we use - all expressed on a scale from 0 to 100 - are the following: students' Grade Point Average in the second academic year (1), the Share of Passed Modules (SPM) (2), the Share of Good Grades (SGG) (3), Final Degree Mark (4), and whether they graduated on time (5). The dependent variables referring to students' academic performance are collected from students' administrative records. We estimate the treatment effect using Equation 1.2. We include controls for the stratifying variables used for the randomization process (sex, age, socioeconomic status, tariff quintile, and department) and the dependent variable's baseline. That is not available for variables such as final degree mark and whether they graduated on time. Standard errors are clustered at the individual level. ***: $p<0.01$, **: $p<0.05$, *: $p<0.1$.

Table 1.7: Impact of the Goal-Setting Treatment: Well-Being and Time Allocation

	(1)	(2)	(3)	(4)	(5)	(6)
	Well-Being:			Time Allocation:		
Variables	Satisfaction	Interest	Study	Sleep	Work	Exercise
Treatment	-0.126	1.883*	-0.542	-0.046	-0.168	0.251
	(1.120)	(1.096)	(0.640)	(0.154)	(0.131)	(0.468)
Mean of the Outcome	68.80	72.78	3.00	8.09	0.93	0.63
Individuals	836	836	779	853	853	853
Observations	836	836	779	853	853	853
Stratifying Variables	X	X	X	X	X	X
Baseline	X	X	X	X	X	X

Notes: We estimate the effect of the treatment in the year of the goal-setting intervention. The dependent variables we use are the following: Satisfaction with life at the University (1), and Interest in the Field of Study (2) - expressed on a scale from 0 to 100 - as well as students' allocation of studying time (3), sleeping time (4), work time (5), and exercise time (6) during a typical day of the week - expressed in hours per day. These variables are collected during the first two online sessions held in the academic year 2016/2017, in November 2016 and March 2017. We estimate the treatment effect using Equation 1.2. We include controls for the stratifying variables used for the randomization process (sex, age, socioeconomic status, tariff quintile, and department) and the dependent variable's baseline. Standard errors are clustered at the individual level. ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$.

Table 1.8: Impact of the Goal-Setting Treatment: Heterogeneity by Planning Efficacy

	(1)	(2)	(3)	(4)	(5)	(6)
	Attendance to:		Academic Performance:		Well-Being:	
Variables	Lectures	Lectures	GPA	Degree Mark	Satisfaction	Interest
Linear Combination: Treatment + Treatment # 1st P.E. Tercile	4.840***	4.310**	-1.076	0.584	1.037	4.669***
	(1.842)	(1.943)	(0.932)	(0.921)	(1.868)	(1.754)
Linear Combination: Treatment + Treatment # 2nd P.E. Tercile	-0.084	0.640	0.442	1.512	0.902	0.535
	(2.053)	(2.197)	(0.975)	(1.005)	(2.167)	(2.037)
Linear Combination: Treatment + Treatment # 3rd P.E. Tercile	0.529	1.302	0.798	0.578	-2.932	-0.391
	(1.960)	(2.129)	(0.990)	(0.910)	(1.877)	(1.920)
2nd Planning Efficacy Tercile	4.268**	2.831	-0.848	-0.523	-0.940	3.290
	(1.956)	(2.077)	(0.902)	(0.967)	(2.086)	(2.007)
3rd Planning Efficacy Tercile	6.960***	5.568***	-0.127	2.949***	3.558**	3.799**
	(2.014)	(2.166)	(1.028)	(0.911)	(1.797)	(1.865)
Treatment # 2nd P.E. Tercile	-4.924*	-3.670	1.517	0.928	-0.134	-4.135
	(2.770)	(2.946)	(1.329)	(1.362)	(2.879)	(2.710)
Treatment # 3rd P.E. Tercile	-4.311	-3.007	1.873	-0.006	-3.969	-5.060**
	(2.721)	(2.918)	(1.361)	(1.293)	(2.654)	(2.568)
Mean of the Outcome	60.12	58.16	61.40	64.02	67.36	70.53
Individuals	874	874	884	864	836	836
Observations	7,593	5,885	884	864	836	836
Stratifying Variables	X	X	X	X	X	X
Baseline	X	X	X	X	X	X
Week Fixed Effects	X	X				
Period of Interest	Post W16	Post W19	Full Year			

Notes: We estimate the effect of the treatment on different outcomes, based on the students' tercile of planning efficacy. We define the terciles of planning efficacy using the distribution of the variable from the online survey held in November 2016, during the Autumn term of the academic year 2016/2017. The first tercile of planning efficacy is kept as a baseline. The dependent variables we use - all expressed on a scale from 0 to 100 - are the following: Percentage of Lectures Attended (columns (1) and (2)), Grade Point Average in the second academic year (column (3)), Final Degree Mark (column (4)), Satisfaction with Life at the University (column (5)), and Interest in the Field of Study (column (6)). Attendance to lectures is estimated in the weeks following the first and third week of the Spring Term of 2017 ("Post W16" and "Post W19", respectively). Attendance is collected through the swipe card system in place at the university. GPA is derived from students' administrative records. Well-being is self-reported by the students during the first two online sessions held in the academic year 2016/2017, in November 2016 and March 2017. We estimate the treatment effect in columns (1) and (2) using Equation 1.1, and in columns (3), (4), and (5) using Equation 1.2. We include controls for week fixed effects, the stratifying variables used for the randomization process (sex, age, socioeconomic status, tariff quintile, and department), and the dependent variable's baseline. Standard errors are clustered at the individual level. ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$.

Table 1.9: Impact of the Goal-Setting Treatment: Heterogeneity by Loss Aversion

	(1)	(2)	(3)	(4)	(5)	(6)
	Attendance to:		Academic Performance:		Well-Being:	
Variables	Lectures	Lectures	GPA	Degree Mark	Satisfaction	Interest
Linear Combination: Treatment + Treatment # Loss Neutral	3.286	4.401	3.841**	2.454*	1.848	-2.349
	(2.663)	(2.948)	(1.671)	(1.358)	(3.099)	(3.073)
Linear Combination: Treatment + Treatment # Loss Averse	2.028*	2.139*	-0.533	0.411	-0.683	2.516**
	(1.168)	(1.240)	(0.561)	(0.570)	(1.211)	(1.179)
Loss Averse	2.717	3.700	3.278**	1.298	-2.718	-4.518**
	(2.236)	(2.446)	(1.489)	(1.113)	(2.274)	(2.284)
Treatment # Loss Averse	-1.259	-2.302	-4.373**	-2.043	-2.531	4.865
	(2.909)	(3.197)	(1.757)	1.471)	(3.321)	(3.281)
Mean of the Outcome	64.52	62.09	61.21	63.95	70.48	76.92
Individuals	972	972	982	962	836	836
Observations	8,449	6,546	982	962	836	836
Stratifying Variables	X	X	X	X	X	X
Baseline	X	X	X	X	X	X
Week Fixed Effects	X	X				
Period of Interest	Post W16	Post W19	Full Year			

We estimate the effect of the treatment on different outcomes, based on whether a student is loss-averse or not. Loss aversion is defined as someone with an endowment effect larger than 0. It is elicited in the goal-setting experiment held in January 2017, during the Spring term of the academic year 2016/2017. The loss-neutral category is kept as a baseline. The dependent variables we use - all expressed on a scale from 0 to 100 - are the following: Percentage of Lectures Attended (columns (1) and (2)), Grade Point Average in the second academic year (column (3)), Final Degree Mark (column (4)), Satisfaction with Life at the University (column (5)), and Interest in the Field of Study (column (6)). Attendance to lectures is estimated in the weeks following the first and third week of the Spring Term of 2017 ("Post W16" and "Post W19", respectively). Attendance is collected through the swipe card system in place at the university. GPA is derived from students' administrative records. Well-being is self-reported by the students during the first two online sessions held in the academic year 2016/2017, in November 2016 and March 2017. We estimate the treatment effect in columns (1) and (2) using Equation 1.1, and in columns (3), (4), and (5) using Equation 1.2. We include controls for week fixed effects, the stratifying variables used for the randomization process (sex, age, socioeconomic status, tariff quintile, and department), and the dependent variable's baseline. Standard errors are clustered at the individual level. ***: $p<0.01$, **: $p<0.05$, *: $p<0.1$.

Table 1.10: The Interplay between Planning Efficacy and Loss Aversion

Variables	(1) Attendance to: Lectures	(2) Lectures	(3) Academic Performance: GPA	(4) Degree Mark	(5) Well-Being: Satisfaction	(6) Interest
Linear Combination: Treatment + Treatment # 1st P.E. Tercile	4.625***	3.705*	-1.293	0.566	-0.038	5.555***
	(2.025)	(2.162)	(1.031)	(0.997)	(2.024)	(1.852)
Linear Combination: Treatment + Treatment # 2nd P.E. Tercile	-0.838	-0.394	-0.104	0.814	0.967	1.215
	(2.299)	(2.464)	(1.066)	(1.049)	(2.267)	(2.204)
Linear Combination: Treatment + Treatment # 3rd P.E. Tercile	-0.447	0.275	-0.425	0.595	-3.484*	-0.659
	(2.237)	(2.396)	(0.886)	(1.029)	(1.996)	(2.117)
2nd Planning Efficacy Tercile	5.642**	4.202*	-0.881	-0.066	-1.283	3.116
	(2.101)	(2.247)	(0.947)	(0.981)	(2.222)	(2.082)
3rd Planning Efficacy Tercile	8.476***	6.971***	0.944	2.989***	3.443*	3.939**
	(2.163)	(2.309)	(0.896)	(1.004)	(1.991)	(2.048)
Treatment # 2nd P.E. Tercile	-5.464*	-4.099	1.190	0.758	1.005	-4.340
	(3.066)	(3.290)	(1.453)	(1.442)	(3.060)	(2.894)
Treatment # 3rd P.E. Tercile	-5.072*	-3.430	0.868	0.539	-3.446	-6.215**
	(3.017)	(3.231)	(1.357)	(1.436)	(2.849)	(2.806)
Mean of the Outcome	64.89	62.70	62.64	64.83	68.87	73.68
Individuals	756	756	884	864	722	722
Observations	6,572	5,092	884	864	722	722
Stratifying Variables	X	X	X	X	X	X
Baseline	X	X	X	X	X	X
Week Fixed Effects	X	X				
Period of Interest	Post W16	Post W19	Full Year			

Notes: We estimate the effect of the treatment on different outcomes, based on the students' tercile of planning efficacy, for students who are categorized as loss averse. Loss aversion is defined as someone with an endowment effect larger than 0. It is elicited in the goal-setting experiment held in January 2017, during the Spring term of the academic year 2016/2017. We define the terciles of planning efficacy using the distribution of the variable from the online survey held in November 2016, during the Autumn term of the academic year 2016/2017. The first tercile of planning efficacy is kept as a baseline. The dependent variables we use - all expressed on a scale from 0 to 100 - are the following: Percentage of Lectures Attended (columns (1) and (2)), Grade Point Average in the second academic year (column (3)), Final Degree Mark (column (4)), Satisfaction with Life at the University (column (5)), and Interest in the Field of Study (column (6)). Attendance to lectures is estimated in the weeks following the first and third week of the Spring Term of 2017 ("Post W16" and "Post W19", respectively). Attendance is collected through the swipe card system in place at the university. GPA is derived from students' administrative records. Well-being is self-reported by the students during the first two online sessions held in the academic year 2016/2017, in November 2016 and March 2017. We estimate the treatment effect in columns (1) and (2) using Equation 1.1, and in columns (3), (4), and (5) using Equation 1.2. We include controls for week fixed effects, the stratifying variables used for the randomization process (sex, age, socioeconomic status, tariff quintile, and department), and the dependent variable's baseline. Standard errors are clustered at the individual level. ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$.

Table 1.11: Balance Checks: Commitment Contract Takers vs. Potential Takers

Variables	(1) Control Mean/SE	(2) Treatment Mean/SE	(3) t-test of the difference (1)-(2)
Female	0.694	0.698	-0.004
	(0.039)	(0.041)	
Home Low SES	0.181	0.155	0.026
	(0.032)	(0.032)	
Home High SES	0.354	0.357	-0.003
	(0.040)	(0.042)	
Home Missing SES	0.132	0.109	0.023
	(0.028)	(0.027)	
EU	0.208	0.217	-0.009
	(0.034)	(0.036)	
Overseas	0.125	0.163	-0.038
	(0.028)	(0.033)	
Top Tariff Quintile	0.139	0.124	0.015
	(0.029)	(0.029)	
Bottom Tariff Quintile	0.174	0.124	0.050
	(0.032)	(0.029)	
Faculty: Science and Health	0.292	0.403	-0.111
	(0.038)	(0.043)	
Faculty: Social Sciences	0.417	0.364	0.053
	(0.041)	(0.043)	
Previous Attendance: Lectures	70.139	69.821	0.318
	(1.798)	(1.997)	
Year 1 Marks	61.092	62.640	-1.548
	(0.925)	(0.903)	
Planning Efficacy	4.534	4.500	0.034
	(0.085)	(0.097)	
Endowment Effect	0.833	0.814	0.019
	(0.031)	(0.034)	
Risk Aversion	-0.272	-0.257	-0.015
	(0.067)	(0.070)	
Present Biased	0.130	0.198	-0.068
	(0.032)	(0.039)	
Individuals	142	129	

Notes: We report the average demographic and individual characteristics of the students - available either from the university administrative records or elicited during previous waves of the survey - based on their assignment to the treatment or the control group during the goal-setting experiment held in January 2017, during the Spring term of the academic year 2016/2017. We include only students from the treatment group who decided to enter the commitment contract offered to them during the experiment and students who, assigned to the control group during the experiment, stated they would have been interested in entering the same commitment contract if they were offered, during the online session of the survey held in March 2017, during the spring term of the academic year 2016/2017. In column (1) we show the characteristics of the control group. In column (2) we report the characteristics of the treated group. The values displayed in column (3) are for the differences between the two means and the relative t-test. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table 1.12: Impact of the Goal-Setting Treatment and the Commitment Contract

	(1)	(2)	(3)	(4)	(5)	(6)
	Attendance to:		Academic Performance:		Well-Being:	
Variables	Lectures	Lectures	GPA	Degree Mark	Satisfaction	Interest
Treatment	5.618**	6.398***	-0.543	0.456	1.286	3.030
	(2.165)	(2.343)	(1.032)	(0.990)	(2.357)	(2.047)
Mean of the Outcome	66.71	64.22	61.60	63.26	66.85	71.41
Individuals	246	246	262	256	240	240
Observations	2,116	1,639	262	256	240	240
Stratifying Variables	X	X	X	X	X	X
Baseline	X	X	X	X	X	X
Week Fixed Effects	X	X				
Period of Interest	Post W16	Post W19	Full Year			

Notes: We estimate the effect of the treatment on those who decided to enter the commitment contract during the goal-setting experiment, in January 2017. The control group consists of those students who were assigned to the control group during the experiment, and who declared, in the March 2017 online survey, that they would have entered the commitment contract if they were offered that option. The dependent variables we use - all expressed on a scale from 0 to 100 - are the following: Percentage of Lectures Attended (columns (1) and (2)), Grade Point Average in the second academic year (column (3)), Final Degree Mark (column (4)), Satisfaction with Life at the University (column (5)), and Interest in the Field of Study (column (6)). Attendance to lectures is estimated in the weeks following the first and third week of the Spring Term of 2017 ("Post W16" and "Post W19", respectively). Attendance is collected through the swipe card system in place at the university. GPA is derived from students' administrative records. Well-being is self-reported by the students during the first two online sessions held in the academic year 2016/2017, in November 2016 and March 2017. We estimate the treatment effect in columns (1) and (2) using Equation 1.1, and in columns (3), (4), and (5) using Equation 1.2. We include controls for week fixed effects, the stratifying variables used for the randomization process (sex, age, socioeconomic status, tariff quintile, and department), and the dependent variable's baseline. Standard errors are clustered at the individual level. ***: $p<0.01$, **: $p<0.05$, *: $p<0.1$.

Table 1.13: Robustness Exercise: Impact on "Swiping Behavior"

Variables	(1) Spring 2017 vs. Autumn 2016 Attendance to Lectures
Linear Combination: Treatment + Treatment # 1st Size Tercile	3.679***
	(1.016)
Linear Combination: Treatment + Treatment # 2nd Size Tercile	1.209
	(1.211)
Linear Combination: Treatment + Treatment # 3rd Size Tercile	4.812***
	(1.533)
2nd Size Tercile	0.348
	(0.679)
3rd Size Tercile	6.211***
	(0.802)
Treatment # 2nd Size Tercile	-4.264***
	(1.291)
Treatment # 3rd Size Tercile	1.258
	(1.480)
Mean of the Outcome	63.23
Individuals	972
Observations	66,864
Individual Fixed Effects	X
Week Fixed Effects	X

Notes: We estimate the treatment effect on the percentage of lectures attended during the Spring term of the academic year 2016/2017, based on the tercile of the lecture size. Lecture size is calculated based on the number of students expected to attend the lecture, over all the lectures held in the Autumn and Spring terms of the academic year 2016/2017. Attendance is collected through the swipe card system in place at the university. We estimate the effect of the treatment using Equation 1.3. We include controls for week and individual fixed effects. Standard errors are clustered at the individual level. ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$.

Table 1.14: Robustness Exercise: Heterogeneity by Enjoyment of Lectures

Variables	(1) Attendance to Lectures	(2)
Linear Combination: Treatment + Treatment # 1st WTP Tercile	2.532	3.232
	(1.901)	(2.002)
Linear Combination: Treatment + Treatment # 2nd WTP Tercile	1.188	1.327
	(1.787)	(1.886)
Linear Combination: Treatment + Treatment # 3rd WTP Tercile	2.856	2.429
	(1.932)	(2.143)
2nd Willingness to Pay for Lectures Tercile	-2.232	-2.096
	(1.921)	(2.006)
3rd Willingness to Pay for Lectures Tercile	-0.444	-0.663
	(2.004)	(2.101)
Treatment # 2nd WTP Tercile	-1.344	-1.905
	(2.613)	(2.756)
Treatment # 3rd WTP Tercile	0.324	-0.804
	(2.736)	(2.952)
Mean of the Outcome	66.74	63.91
Individuals	972	972
Observations	8,449	6,546
Stratifying Variables	X	X
Baseline	X	X
Week Fixed Effects	X	X
Period of Interest	Post W16	Post W19

Notes: We estimate the effect of the treatment on the percentage of lectures attended, based on the students' tercile of students' enjoyment of lectures - how much money they would want to attend one extra hour of lectures. The variable is elicited in the goal-setting experiment held in January 2017, during the Spring term of the academic year 2016/2017. The first tercile of planning efficacy is kept as a baseline. Attendance to lectures is estimated in the weeks following the first and third week of the Spring Term of 2017 ("Post W16" and "Post W19", respectively). Attendance is collected through the swipe card system in place at the university. We estimate the treatment effect using Equation 1.1. We include controls for week fixed effects, the stratifying variables used for the randomization process (sex, age, socioeconomic status, tariff quintile, and department), and the dependent variable's baseline. Standard errors are clustered at the individual level. ***: p <0.01, **: p <0.05, *: p <0.1.

Table 1.15: Robustness Exercise: Including Week 4 Participants

Variables	(1) Attendance to: Lectures	(2) Lectures	(3) Academic Performance: GPA	(4) Degree Mark	(5) Well-Being: Satisfaction	(6) Interest
Treatment	2.228**	2.808**	-0.414	0.832	-0.056	2.299**
	(1.049)	(1.174)	(1.057)	(0.529)	(1.114)	(1.097)
Mean of the Outcome	64.20	60.35	62.26	64.34	68.35	72.39
Individuals	1018	1018	1,008	989	858	858
Observations	8,847	5,850	1,008	989	858	858
Stratifying Variables	X	X	X	X	X	X
Baseline	X	X	X	X	X	X
Week Fixed Effects	X	X				
Period of Interest	Post W16	Post W20	Full Year			

Notes: We estimate the effect of the treatment on different outcomes, including also students who attended the lab for the experiment during the 4th week of the term. The dependent variables we use - all expressed on a scale from 0 to 100 - are the following: Percentage of Lectures Attended (columns (1) and (2)), Grade Point Average in the second academic year (column (3)), Final Degree Mark (column (4)), Satisfaction with Life at the University (column (5)), and Interest in the Field of Study (column (6)). Attendance to lectures is estimated in the weeks following the first and third week of the Spring Term of 2017 ("Post W16" and "Post W19", respectively). Attendance is collected through the swipe card system in place at the university. GPA is derived from students' administrative records. Well-being is self-reported by the students during the first two online sessions held in the academic year 2016/2017, in November 2016 and March 2017. We estimate the treatment effect in columns (1) and (2) using Equation 1.1, and in columns (3), (4), and (5) using Equation 1.2. We include controls for week fixed effects, the stratifying variables used for the randomization process (sex, age, socioeconomic status, tariff quintile, and department), and the dependent variable's baseline. Standard errors are clustered at the individual level. ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$.

Chapter 2

Beyond Test Scores: the Rank Effect and Non-Cognitive Skills

2.1 Introduction

Decades of research on the impact of peers on academic outcomes have generally identified positive but moderate effects, despite extensive innovations and debate over the appropriate methodological framework.[1] While these findings suggest that direct peer effects may be limited, a parallel literature on students' academic rank has consistently documented much larger impacts, revealing a different channel through which peer groups shape individual outcomes. Building on the "big fish in a little pond" theory (Marsh and Parker, 1984), this body of work argues that students' relative standing within their peer group can have substantial and lasting consequences for their future trajectories.

More specifically, researchers have found that higher academic rank positively influences future academic performance, educational attainment, career choice (Megalokonomou and Zhang, 2024; Carneiro et al., 2023; Denning et al., 2023; Elsner et al., 2021; Pagani et al., 2021; Murphy and Weinhardt, 2020; Elsner and Isphording, 2018, 2017), major

[1] A continuous discussion on the correct functional form for estimating peer effects has highlighted the limitations of the classic linear-in-means model, which may overlook important within-group heterogeneity, ignore selection effects, and fail to capture the dynamic nature of peer influence. While alternative approaches often find more nuanced and sometimes stronger effects, these differences are typically modest. For a comprehensive review, see Sacerdote (2011).

choice (Goulas et al., 2023; Delaney and Devereux, 2021), and future earnings (Denning et al., 2023). We contribute to this growing literature by extending the range of outcomes considered and the mechanisms explored — delving more deeply into the non-cognitive channels through which rank operates and examining its relationship with parental investment. We also assess the long-term impacts across a broad set of educational, socioeconomic, and well-being dimensions.

We study the effects of academic rank using a unique dataset covering the full population of children born between October 1950 and October 1955 in Aberdeen, Scotland, and enrolled in the city's primary schools. The dataset includes nearly 10,000 children and draws on three rich sources of information. First, a comprehensive survey conducted between 1962 and 1964 gathered data on children's academic progression, cognitive performance, and anthropometric indicators through school medical exams and hospital birth records. Non-cognitive skills were measured using the Rutter Questionnaire (Rutter, 1967), a widely validated behavioral assessment completed by teachers. Second, a randomly selected 20% subsample of children was surveyed at home, with parents providing detailed information on family structure, social and health behaviors, parental involvement, and time use. Finally, a follow-up survey conducted in 2001 (with the original participants now aged between 46 and 51 years old) achieved a response rate of nearly 60%, offering detailed information on adult outcomes including education, occupation, income, health, and subjective well-being. The breadth and depth of this dataset allow us to trace the effects of academic rank from childhood well into adulthood.

We construct academic rank based on students' performance on a standardized cognitive test administered around age 9, and define each student's peer group at the school-cohort level — typically involving 30–35 children, small enough to ensure meaningful peer interaction. Our identification strategy exploits quasi-random variation in rank within these groups: we show that group composition is conditionally random, with no evidence of systematic sorting based on observed characteristics. This allows us to interpret within-group rank as plausibly exogenous to underlying ability, parental background, or non-cognitive traits. Our analysis then examines how this measure of rank influences a range of outcomes, from academic performance and behavioral skills in primary school to parental investment and long-term educational and economic trajectories.

We find that academic rank has a substantial and robust impact on students' academic performance, as measured by the results of the 11-plus examination — a high-stakes standardized test taken at the end of primary school, which determined access to grammar schools, the selective secondary track in the UK system. A rise of four positions in a student's rank within their school-cohort group — equivalent to a 10% improvement in relative standing given the average group size — raises 11-plus test scores by roughly 6% of a standard deviation, holding cognitive skills, individual, and peer characteristics constant. Interpreting rank effects linearly is supported by the data: we show that the estimated relationship is approximately linear across the full rank distribution, consistent with findings in Elsner et al. (2021) and Murphy and Weinhardt (2020), and suggestive of a self-preserving dynamic where early rank boosts performance, leaving the rank unchanged. Moreover, the effect is stronger for girls, who not only begin with higher average test scores but also respond more sharply to their relative group position — an important pattern that echoes the gender differences in sensitivity to feedback observed in other peer-related settings (Buser et al., 2023; Lavy, 2019).

To study the effect of academic rank on non-cognitive development, we derive two standardized measures — externalizing and internalizing skills — using factor analysis on teacher responses to the Rutter Questionnaire (Rutter, 1967). This method allows us to reduce dimensionality and recover latent behavioral traits from a broad set of teacher-reported items, following a well-established approach in developmental psychology (e.g., Boyle and Jones, 1985; McGee et al., 1985), and increasingly adopted in economics (e.g., Attanasio et al., 2020b). Externalizing skills capture outward-oriented behaviors such as restlessness, aggression, and difficulty concentrating, often associated with poor impulse control. Internalizing skills, by contrast, reflect inward-focused behaviors such as anxiety, low self-esteem, and social withdrawal, and are commonly linked to emotional regulation and self-perception. While previous work on rank effects has typically focused on narrow, single-trait outcomes such as self-confidence or academic expectations (Elsner et al., 2021;

Murphy and Weinhardt, 2020), our use of composite constructs offers a broader and more structured perspective on non-cognitive development.

We find that a four-position improvement in school-cohort rank — approximately a 10% increase in relative standing — raises internalizing skills by roughly 4.5% of a standard deviation, a statistically significant and robust effect across specifications. In contrast, the effect on externalizing skills is smaller, around 3% of a standard deviation, and less precisely estimated, being statistically significant only at the 10% level. These average effects conceal notable gender heterogeneity. Boys show gains in both skill dimensions as their rank improves, while girls exhibit a much stronger response in internalizing skills but no meaningful change in externalizing skills. However, the effect on externalizing skills appears to be driven by a small number of extreme cases with severe behavioral issues, as it loses statistical significance once these cases are excluded. In contrast, the internalizing skills effect remains robust even when the most extreme cases are removed, both in the overall sample and when disaggregated by gender. Taken together, these findings suggest that academic rank has a broad but uneven impact on children's non-cognitive development — influencing self-concept and emotional adjustment more consistently than behavioral regulation. This interpretation is consistent with psychological evidence linking relative academic standing to perceived competence, self-esteem, and long-term expectations (Creemers et al., 2013; Elsner et al., 2021).

Given the robust impact of rank on non-cognitive development, a natural question is whether these effects arise solely through internal psychological mechanisms or are also shaped by external factors. One potential channel is parental behavior, as parents might respond to their child's rank by adjusting their level of involvement or support. We find little evidence of systematic responses: there are no consistent effects on parental involvement with homework or on the time children spend studying. The only detectable effect is a modest increase in parental awareness of the teacher's identity, hinting at some limited effect on their engagement. Overall, the results suggest that rank effects are driven primarily by children's internal adjustments, rather than changes in parental investment.

To assess the long-term consequences of academic rank, we estimate its effect on adult outcomes using responses from a 2001 follow-up survey, administered nearly 40 years after the original data collection. Despite selective attrition, we show that participation in the follow-up is unrelated to academic rank, allowing for credible identification of long-term effects. Our findings indicate that academic rank in primary school impacts educational attainment. A four-position increase in school-cohort rank raises the probability of attending a grammar school by 4 percentage points — equivalent to a 20% increase relative to the baseline. The same improvement increases the likelihood of completing O-levels and A-levels (which were standardized qualifications typically taken at ages 16 and 18, respectively, and served as key milestones for educational progression in the UK system) by 5% and 12%, respectively. These effects are notably stronger for girls, who appear to benefit more from higher relative rank in terms of access to selective education. Importantly, we do not find an impact of rank on the probability of getting a degree.

However, these educational gains do not consistently extend to the labor market. We estimate a modest impact on earnings: a four-position rank increase raises the probability of earning more than £25,000 per year by 2.2 percentage points (roughly 6% of the baseline), and only for boys. There is no effect on higher income thresholds (£35,000 or £45,000), nor on broad measures of socioeconomic status, such as working in high-status occupations. Rank also appears unrelated to family formation: we find no significant effect on the probability of ever marrying or having children. Likewise, there is no detectable impact on subjective well-being, including self-reported happiness and enjoyment of daily activities. On the other hand, higher-ranked children seem to have fonder memories of their time in primary school, as they are more likely to state they were happy at the time.

Taken together, the results suggest that while early academic rank meaningfully shapes educational trajectories, its influence on economic and personal outcomes is more limited. This pattern is especially striking for girls, whose stronger educational response to rank does not translate into higher earnings or occupational status. These findings likely reflect the historical context of mid-20th-century Scotland, when access to university and professional careers, particularly for women, was far more constrained than today. As a result, rank-driven differences in educational success may have been insufficient to overcome broader structural barriers, limiting the translation of early academic advantage into adult socioeconomic returns.

We contribute to the growing literature on the effect of academic rank in several ways. We introduce a broader measure of non-cognitive skills, grounded in decades of literature in developmental psychology (Narusyte et al., 2017; Klein et al., 2009; Iloeje and Meme, 1992; McGee et al., 1985; Rutter, 1967). With the exception of Pagani et al. (2021), previous studies focus on single self-reported traits, such as self-confidence (Murphy and Weinhardt, 2020) or academic expectations (Elsner et al., 2021; Elsner and Isphording, 2017). Our analysis uses validated composite measures of externalizing and internalizing skills, capturing a richer spectrum of non-cognitive traits (for instance, emotional regulation, impulse control, attention, and social interaction). This approach provides a more comprehensive view of how rank influences children's behavioral and emotional development, and allows us to test not just whether rank matters for non-cognitive skills, but which dimensions it affects and how robust those effects are. Notably, we show that the rank effect on internalizing skills is both sizable and stable, while the effect on externalizing skills is more fragile and driven by a small subset of students with severe behavioral difficulties.

We also offer new evidence on the role of parental behavior in mediating the rank effect. The possibility that parents adjust their involvement based on the child's relative academic standing has been suggested in theory but rarely tested empirically. A recent exception is Megalokonomou and Zhang (2024), who show that parents in China respond to their child's classroom rank by increasing or reducing tutoring investments, interpreting rank as a signal of ability. Our paper complements this work by studying a different type of parental response—day-to-day involvement in the child's education—using survey data collected from mothers in a randomized subsample. We examine whether rank affects parental knowledge of the child's teacher, help with homework, and the amount of time children spend studying at home. The results show little systematic evidence of behavioral adjustment beyond a small increase in their awareness of the teacher's name. This suggests that the effects of rank on non-cognitive and academic outcomes may arise more from how children internalize their relative position than from active parental recalibration of effort, especially in lower-stakes educational environments.

Moreover, we provide rare empirical evidence on the long-term consequences of early academic rank, nearly four decades after the reference point. Few studies beyond Denning et al. (2023), who link rank to earnings 25 years later, have examined whether early relative standing has persistent effects. We find that academic rank has lasting consequences for educational attainment, particularly for girls, whose probability of attending grammar school and completing formal qualifications responds more strongly to rank than that of boys. These educational gains do not translate into higher earnings for women, likely reflecting historical constraints on university access and labor market participation. Overall, by tracing the influence of academic rank across a wide range of adult outcomes — from recalled school experiences to income — we offer a comprehensive account of its long-term relevance.

Finally, by showing that academic rank influences the development of non-cognitive skills and has long-lasting effects on educational and economic outcomes, our findings contribute to the broader literature on the long-term returns to early skill formation. In particular, they complement evidence that interventions targeting cognitive and non-cognitive development in childhood can yield persistent benefits (Sorrenti et al., 2024; Heckman et al., 2013; Heckman and Kautz, 2012), suggesting that relative position in the classroom may be an overlooked yet meaningful mechanism shaping children's life trajectories.

2.2 Data

This study draws on rich data tracking a cohort of students enrolled in all primary schools in Aberdeen, Scotland, in 1962. Our analysis focuses on children attending grades 3 to 7 at the time, for whom we construct a measure of rank based on performance in a standardized test taken at age 9. We define each student's rank within their school-cohort group to capture their position in the local academic distribution. We examine the relationship between this early academic rank and a wide set of outcomes measured over the life course. These include academic performance at age 11, non-cognitive skills assessed during primary school, and parental investment for a subset of students whose parents were interviewed. We also study long-term outcomes, including educational attainment, socioeconomic status, fertility, and both physical and mental health, of those who replied to a follow-up survey approximately four decades later, in 2001.

2.2.1 The Aberdeen Children of the 1950's Survey

We base our analysis on the Aberdeen Child Development Survey, a comprehensive study conducted between 1962 and 1964.[2] Originally designed to investigate the link between anthropometric measures and reading disabilities, the survey was not primarily focused on schooling. Nevertheless, it provides a rich array of information that enables us to track children throughout their primary school years and into adulthood. The dataset includes detailed records of academic progression and achievement, non-cognitive skill measures, physical development indicators, and family background data collected through interviews with a randomly selected subset of parents. In addition, it incorporates follow-up survey data capturing long-term outcomes approximately 40 years later. The data we use are drawn from three distinct sources within the broader study.

The reading survey: It is the core of the initial data collection effort. Conducted in multiple stages, this survey began in December 1962 with the assistance of teachers, who helped students provide detailed demographic information, including date of birth, father's occupation, and family size. Schools also contributed attendance records for the preceding two years and available test scores — any missing scores were supplemented in subsequent rounds. In addition, hospital records were used to extract obstetric and social data gathered during the mothers' pregnancies. By July 1963, further information was collected from schools, including anthropometric measures such as height and weight, taken during routine medical examinations when children entered school (around age 5), and again at ages 9 and 12. The final stage of the Reading Survey took place in

[2]Now known as "The Aberdeen Children of the 1950s"

March 1964, when teachers completed behavioral assessments. These included the Rutter Questionnaire (scale B) (Rutter, 1967), a validated psychological tool for identifying minor behavioral disorders in children.

The family survey: It targeted a random subsample comprising 25% of the children included in the Reading Survey. With an overall response rate of 80%, this follow-up collected information through interviews with the children's mothers or a substitute caregiver. The survey captured additional insights into the children's health and behavioral conditions, how they allocated their time between leisure and school activities, and provided more detailed demographic information about the parents.

The 2001 follow-up survey: Participants from the initial cohort were contacted by mail in 2001 and asked to complete a comprehensive questionnaire covering various aspects of their adult lives, including physical and mental health, psychological well-being, and socioeconomic status. Approximately 60% of those who received the survey returned it with complete responses, allowing us to link early-life information to long-term outcomes.

2.2.2 Sample Definition

Our dataset includes 12,151 observations, each corresponding to a child born in Aberdeen between October 1950 and September 1955 and attending grades 3 to 7 at the time. From this initial pool, we apply a series of restrictions to define our analysis sample. First, we exclude students enrolled in private schools (2.88%) and special schools (9.5%), as these institutions follow different educational and administrative structures.[3] We also drop students attending schools that closed before 1964, when the reading survey ended. Finally, we retain only those students for whom we can reliably identify their cohort as of December 1962. After these exclusions, our working sample consists of 9,969 students across 28 schools. For each specific outcome analysis, we further refine the sample to ensure comparability and meaningful inference. When examining the effect of rank on academic performance, we use all students in our working sample. For the analysis of non-

[3]Private schools refused to have their teachers administer the Rutter Questionnaire, which is problematic for our analysis. We do not know enough about how special schools worked to understand if they could be compared to the other public schools in the sample.

cognitive skills, we focus on students still enrolled in primary school as of March 1964, when teachers completed the Rutter Questionnaire. We therefore exclude the cohorts who had already transitioned to secondary or junior secondary school by that time. This restriction leaves us with a sample of 6,779 students.

2.2.3 Defining School Cohort in the Scottish Education System

Given that our measure of relative rank is defined within school-cohort groups, it is essential to clarify how we construct cohorts in the context of the institutional setting. In principle, a cohort would correspond to all students enrolling in school during the same academic year, beginning from grade 1 at around age 5. However, the school entry policy in place at the time involved two separate intakes per calendar year: one in January and one in August.[4] Children born between October and the end of March were eligible for the January intake, while those born between April and the end of September typically entered school in August. As a result, even within the same grade, students could belong to different intake groups, having started school at different times and progressed separately through the system.

Because children spent two semesters in each grade before advancing to the next, the two intake streams remained effectively distinct despite students being in the same nominal grade. As a result, children who entered school in different semesters followed parallel but separate academic trajectories. This institutional feature is crucial for our analysis, as it implies that grade alone is not sufficient to identify a student's cohort. To construct the rank variable accurately, we define each student's cohort based on both the grade they attended in the reference year of the survey (December 1962) and their expected intake group, as inferred from their date of birth. This procedure leaves us with a total of 10 distinct cohorts. We then calculate each student's rank within their school-cohort group, defined by the combination of the school they attended and their inferred

[4]Based on the statement by the Director of Education to a Town Council of Aberdeen meeting that took place on 3rd October 1960 (Lawlor et al. (2006)) we know that schools had two or more admission dates every academic year. All the students who became five years old before the next admission date had to attend school from the first school day following that admission date.

cohort.

2.2.4 Defining Rank within the School-Cohort Group

We use the standardized outcome of a cognitive skills test taken at age 9 (see subsection 2.2.5) to construct the ranking of students within their school-cohort group. To facilitate comparability across groups of different sizes, we calculate a "percentilized rank" following the approach of Murphy and Weinhardt (2020), which normalizes each student's rank by the total number of students in their group. This approach ensures that the resulting rank measure is bounded between 0 (the lowest-ranked student) and 1 (the highest-ranked student):

$$RANK_{isc} = (n_{isc} - 1)/(N_{sc} - 1) \tag{2.1}$$

Where n_{isc} is the ordinal rank of individual i enrolled in school s, in cohort c. N_{sc} is the size of the school-cohort group to which the student belongs. This transformation expresses each student's relative standing as a percentile within their group, ensuring that rank measures are directly comparable across groups of varying sizes. When ties occur, the mean rank is assigned, preserving the average rank of a group and avoiding arbitrary tie-breaking.

2.2.5 Standardized Tests

We rely on two standardized tests administered during primary school to measure students' cognitive skills and academic achievement. The first is a low-stakes assessment taken around age 9, which we use to construct relative academic rank and capture baseline cognitive ability. The second, known as the 11-plus test, is a high-stakes examination taken at the end of primary school that played a central role in determining secondary school placement.

Age 9 Test: We use performance on the age 9 test as our primary measure of students' cognitive skills. This test not only serves as the basis for constructing our within-group rank measure, but also offers a reliable baseline assessment of individual academic ability. Its design — focused on verbal and numerical reasoning — was specifically intended to capture core cognitive competencies and to facilitate comparisons of student achievement across different regions. In addition, the test was used as a screening tool for reading difficulties.[5] Given the nature and purpose of the test, we refer to the resulting scores as our measure of *cognitive skills* throughout the analysis. Although the test was administered under standardized conditions, the resulting distribution of scores deviates from the standard normal, as shown in Figure 2.1. To assess whether these distributions differ systematically across cohorts, we perform a Kruskal-Wallis test—a non-parametric method that compares the distributions of multiple groups based on the ranks of their values. The results indicate no statistically significant differences in the distributions of cognitive skills across cohorts, suggesting a high degree of comparability in this key measure used to define students' relative academic rank.

11-plus Test: Our second measure of academic performance is based on the results of the 11-plus test, a standardized exam administered at the end of primary school to children aged 11 to 12.[6] This was a high-stakes examination that played a central role in determining admission to grammar schools, with long-term implications for educational attainment and, particularly for girls, labor market outcomes (Clark and Del Bono, 2016). Among the components, the two verbal reasoning tests are consistently available for all cohorts, whereas the overall composite score is missing for the two youngest cohorts — those born between October 1954 and October 1955. Consequently, we use verbal reasoning scores as our primary outcome measure, although results using the composite score are also reported where available. Test scores were standardized using both national and Aberdeen-specific norms by year. Figure 2.2 illustrates the distribution of standardized verbal reasoning scores across cohorts, which closely resemble a standard normal

[5] The test was known as the "Schonell and Adams Essential Intelligence Test (Form B)" (Schonell and Adams, 1940), and consisted of a battery of 100 questions covering verbal reasoning and arithmetic. It was administered such that children in different cohorts would take it at the same relative age: students who entered school in January were tested in November, while those who started in August were tested in May.

[6] Students sat four separate tests: two verbal reasoning tests, one arithmetic test, and one English test. In addition, teachers provided an estimate of each student's likely performance, which was scaled to match the mean and standard deviation of the class's actual results.

distribution and show minimal variation. A Kruskal-Wallis test confirms that the score distributions do not differ significantly across cohorts, supporting the comparability of this outcome measure throughout the sample.

2.3 Empirical Strategy

We aim to estimate the rank effect on future academic achievements (the 11-plus Test), non-cognitive skills, parental investment, and long-term outcomes. Our empirical strategy is informed by the recently developed literature on the rank effect (Murphy and Weinhardt, 2020; Elsner et al., 2021; Carneiro et al., 2023; Denning et al., 2023). We start by describing the identification strategy, clarifying our assumptions, and illustrating the identifying variation we exploit.

2.3.1 Our Empirical Strategy

We have individual children i, enrolled in school $s \in [1, ..., 28]$, in grade $c \in [3, ..., 7]$ (as in, all the children born over the 12 month-period which defines grade assignment, hence from October of year t until October of year $t+1$), that enter school with intake $k \in [1, 2]$ (where 1 stands for the January intake and 2 stands for the August intake). The group is given by each intake, within a grade, within a school. We refer to the combination of grade and intake as cohorts $c \in [1, ..., 10]$, which we will use as our group unit. The rank of student i, enrolled in school s, and cohort c within their school-cohort group, R_{isc}, is a function of the student's cognitive skills A_{isc} and group characteristics $\overline{W}_{sc}$:

$$R_{isc} = f(A_{isc}, \overline{W}_{sc}) \tag{1}$$

We can estimate the within-cohort rank effect using the following equation:

$$Y_{isc} = \alpha R_{isc} + \beta A_{isc} + \lambda_s + \lambda_c + \epsilon_{isc} \tag{2}$$

The variation that allows us to identify the rank effect has two sources:

1) Children with identical cognitive skills, who are in the same school, but belong to different cohorts.

2) Children with identical cognitive skills, who are in the same cohort, but are enrolled in different schools.

For the rank effect to be identified, we need the following conditional independence assumption to hold: $\mathbb{E}[\epsilon_{isc}|R_{isc}, A_{isc}, \lambda_s, \lambda_c] = 0$. The assumption could fail if the outcomes that we study are correlated with factors that are specific to the school-cohort group, or if peer characteristics that affect our outcomes and are different from rank differ across cohort groups within the same school. We will show that this is not the case in our sample.

Our final specification adds a few more elements. We want to make sure that the rank effect coefficient is not capturing any higher-order peer or ability effect.[7] We add a quadratic polynomial of individual cognitive skills $g(A_{isc})$, and include the mean and standard deviation of the group cognitive skills distribution, $\overline{A}_{sc-i}$ and σ_{sc-i} (calculated as the mean and standard deviation of the children's test score within their school-cohort group). We also include a vector of individual characteristics X_i (which accounts for sex, socioeconomic status, month of birth, and number of siblings). The equation we estimate is the following:

$$Y_{isc} = \alpha R_{isc} + \beta g(A_{isc}) + \gamma_1 \overline{A}_{sc-i} + \gamma_2 \sigma_{sc-i} + X_i \delta + \lambda_s + \lambda_c + \epsilon_{isc} \tag{3}$$

Including a quadratic polynomial of individual cognitive skills in the specification allows for a flexible relationship between test scores and later outcomes. This choice is motivated not only by the need to capture higher-order effects, but also to account for potential nonlinearities in how cognitive ability translates into future performance. Omitting this

[7] Denning et al. (2023) provide an insightful discussion on this issue. They compare the estimation of the rank effect to an exclusion restriction. Once the common impact of being in a certain group is accounted for, and the groups we are comparing are sufficiently similar, the remaining variation in student outcomes (given their cognitive skills) stems from peer effects related to rank. While we can corroborate this claim with several robustness checks, it remains an assumption that needs to be discussed on a case-by-case basis.

flexibility could lead to misspecification and, consequently, biased estimates of the rank effect. By incorporating polynomials, we reduce the risk that our results are driven by incorrect assumptions about the functional form of this relationship. In addition, we control for the mean and standard deviation of peers' cognitive skills within each school-cohort group, which ensures comparisons are made among students from groups with similar ability profiles — further isolating the contribution of relative rank from variation attributable to group-level cognitive skill differences.

As we expect the variation in rank to occur at the group level, all our equations are estimated clustering the standard errors at the school-cohort level (or group level).

2.3.2 Evidence on the Validity of the Identifying Assumption

Believing that our model can disentangle the rank effect from other types of peer effects is essential, but not sufficient, for our identification strategy to be successful. A key fact we need to establish is that assignment to groups was quasi-random. The month of birth cutoff is insufficient to ensure that is the case. It makes the variation of cognitive skills across cohorts idiosyncratic but does not account for students potentially sorting into schools. The literature separates between active and passive sorting. The former involves children (or, more realistically, parents) selecting their peer group based on their rank preference. The latter implies that children with certain characteristics could be more likely grouped (if, for instance, school choice is non-random, as in our setting). Considering that children are assigned to a certain intake based on their month of birth, we can safely rule out active sorting. On the other hand, passive sorting remains a concern.

We use two tests to diagnose whether passive sorting exists in our setting. First, we estimate the relationship between individual characteristics and rank in the school-cohort group. We regress individual characteristics on percentile rank, conditioning on a quadratic polynomial of cognitive skills, $g(A_{isc})$, the mean and standard deviation of the group cognitive skills distribution, $\overline{A}_{sc-i}$ and σ_{sc-i}, and school and cohort fixed effects. The rank variables and the controls are constructed using our measure of cognitive skills, the outcome of the age 9 test. The dependent variables we use are sex (boy or girl), socioeconomic status (high or low), height and weight at the time of the first medical exam, birth weight, and number of siblings.[8]

$$X_i = \alpha R_{isc} + \beta g(A_{isc}) + \gamma_1 \overline{A}_{sc-i} + \gamma_2 \sigma_{sc-i} + \lambda_s + \lambda_c + \epsilon_{isc} \tag{4}$$

The top panel of Table 2.1 shows the estimated coefficient α for Equation 4. We do not find any conditional relation between individual characteristics and rank. However, the coefficient for weight is significant at the 10% level, while the one for height is just barely insignificant. These coefficients are not necessarily problematic, and controlling for the two in our main equation would address the sorting concerns. An additional test can provide more information on whether sorting is happening.

Because ranking is a function of the distribution of cognitive skills within the group, an association between individual characteristics and features of the distribution of peer cognitive skills $f(\overline{A}_{sc-i})$ can inform us on whether children with certain characteristics are more likely to end up in certain groups. In particular, we compute the mean, standard deviation, and quartiles of the student-specific leave-out peers' outcome in individual cognitive skills. Again, we include a quadratic polynomial of individual cognitive skills, $g(A_{isc})$, and school and cohort fixed effects as controls.

$$X_i = \gamma f(\overline{A}_{sc-i}) + \beta g(A_{isc}) + \lambda_s + \lambda_c + \epsilon_{isc} \tag{5}$$

The panels of Table 2.1 referring to Equation 5 report estimates of γ. These results rule out any (conditional) relationship between peer cognitive skills and individual characteristics, suggesting that passive sorting is not an issue in our setting. It corroborates our assumption of conditional quasi-random assignment to school-cohort groups.

[8]Female is a binary variable equal to one if the child is a girl. Socioeconomic status (SES) is derived from the two-digit occupational code based on ISCO-58; children whose parent holds a position classified between 1 and 20 — corresponding to "Administrative, Executive, and Managerial Workers" — are coded as high SES. Height and weight refer to measurements taken during the first school medical exam, typically conducted at school entry. Because children could be examined up to a year apart due to staggered intake and absences, we residualize these measures by age at examination. Specifically, we restrict the sample to children examined within a 12-month age window (dropping 1.5% of extreme values), regress height and weight separately on a quadratic polynomial in age (in months), and use the residuals as standardized variables. Birth weight is recorded in pounds, as reported in the children's medical records. Number of siblings reflects the total number of siblings living in the household as of December 1962.

2.4 Non-Cognitive Development: Externalizing and Internalizing Skills

We measure non-cognitive skills using the Rutter Children's Behaviour Questionnaire (Rutter, 1967), which was completed by teachers in March 1964 for each child in their classroom. This well-established behavioral assessment tool consists of 26 items designed to detect minor behavioral issues in children and has been widely adopted in educational and clinical research settings (Narusyte et al., 2017; Klein et al., 2009; Iloeje and Meme, 1992; McGee et al., 1985; Boyle and Jones, 1985; Behar and Stringfield, 1974). Teachers were asked to rate each behavior as "Does not apply," "Somewhat applies," or "Definitely applies" for each child, providing a structured and consistent measure of classroom behavior.[9]

The 26 behaviors captured in the Rutter scale refer to two broad domains: externalizing and internalizing behaviors, following a well-established classification in child psychiatry (Eisenberg et al., 2001; Achenbach and Edelbrock, 1978). Externalizing behaviors are outward-directed and often reflect poor self-regulation, such as inattention, impulsivity, or aggression. Internalizing behaviors, by contrast, are inward-directed and typically relate to emotional states such as anxiety, sadness, or social withdrawal. These two dimensions provide distinct but complementary perspectives on a child's non-cognitive functioning and social-emotional development.

Similarly to Attanasio et al. (2020b), we interpret these behavioral traits in terms of skills rather than symptoms. Externalizing skills reflect a child's ability to regulate impulses, sustain attention, and interact appropriately with others — skills crucial for classroom engagement and cooperation. Internalizing skills, on the other hand, relate to the capacity to channel focus and emotional awareness toward task performance, including

[9] The 26 items in the questionnaire include: "Restless", "Truant", "Fidgety", "Destroys Belongings", "Fights other Children", "Disliked", "Anxious", "Solitary", "Irritable", "Often Unhappy", "Tics", "Sucks Fingers", "Nail Biting", "Trivial Absences", "Disobedient", "Short Attention Span", "Fearful", "Fussy", "Often Lies", "Stealing", "Wet/Soiled Themselves", "Often Aching/in Pain", "Tearful", "Stutters", "Other Speech Difficulty", "Bullies other Children".

self-discipline and self-awareness. Each item on the Rutter scale is scored such that higher values indicate more positive behavioral traits: a score of 0 corresponds to “Definitely Applies,” 1 to “Somewhat Applies,” and 2 to “Does Not Apply.” This scoring convention allows us to interpret higher scores — and positive coefficients in our analysis — as indicating stronger non-cognitive skills.

2.4.1 Extracting Measures of Non-Cognitive Skills: Factor Analysis

We rely on principal (common) factor analysis to isolate two latent constructs underlying our behavioral data: externalizing and internalizing skills. This statistical technique captures the shared variation across multiple observed variables — here, questionnaire items — and estimates scores for unobserved (latent) traits that best explain that variation. Given the ordinal nature of the items in the Rutter scale, we use a polychoric correlation matrix to obtain consistent and unbiased estimates of the relationships between items (Olsson, 1979). To determine the appropriate number of factors, we combine theoretical guidance from the psychology literature (Narusyte et al., 2017; Klein et al., 2009; Iloeje and Meme, 1992; McGee et al., 1985; Boyle and Jones, 1985; Behar and Stringfield, 1974) with empirical criteria.

The scree plot of eigenvalues from the initial factor extraction helps identify the number of relevant factors. Following the rule proposed by Kaiser (1960), we retain factors with eigenvalues greater than one, indicating they explain more variance than any individual item. As shown in Figure 2.3, four factors exceed this threshold, but two clearly dominate in terms of explained variance — consistent with our theoretical expectation of a two-factor structure.

We then refine the item set used to define each factor. This step is based on rotated factor loadings and item-specific uniqueness values, obtained through oblique (quartimin) rotation, which allows for correlation between factors and improves interpretability. We exclude items that exhibit weak association with any factor (factor loading below 0.4) or high idiosyncratic variance (uniqueness above 0.8). Table 2.2 presents the results from both the initial and refined iterations. After dropping non-informative items, 19 remain and load clearly onto one of the two factors, corresponding to externalizing or internalizing skills. A second round of factor analysis confirms the stability of this structure. Together, the two retained factors explain approximately 80% of the total variance in the selected items.

To compute individual-level scores for each skill dimension, we apply the Bartlett method (Hershberger, 2005), which uses maximum likelihood estimation to generate unbiased factor scores. Each child receives a score for externalizing and internalizing skills, which we standardize by cohort to ensure comparability across groups. The reliability of the two indices, as measured by Cronbach's alpha — a statistic that captures internal consistency among items in a scale — is 0.84 for externalizing skills and 0.66 for internalizing skills. These values suggest high reliability for the former and moderate reliability for the latter.

Figure 2.4 presents the standardized distributions of the two skill measures by cohort. Externalizing skills are shown in the top panel, while internalizing skills are displayed in the bottom panel. As expected for an index originally developed to identify mild behavioral difficulties, both distributions are fairly concentrated, with most children clustered in a narrow range. Each distribution displays a long left tail, representing a smaller share of children with more pronounced behavioral challenges. A Kruskal-Wallis test confirms that the distributions of both skill measures differ significantly across cohorts. This likely reflects differences in age at the time of assessment, as children in different cohorts were tested simultaneously in March 1964, but varied in age depending on their intake group. These differences underscore the importance of relying on within-cohort comparisons in all subsequent analyses.

2.4.2 Non-Cognitive Skills in the Rank Effect Literature

A growing body of research has explored how a student's relative position within their peer group - usually school-cohort or classroom - affects non-cognitive development. Existing studies often focus on specific traits such as conscientiousness, self-esteem, self-confidence, or expectations, typically using narrow constructs to test theoretical mechanisms through which rank may operate. While this work provides important insights, it often considers individual traits in isolation, potentially missing the broader impact rank may have across a range of non-cognitive dimensions.

Our approach broadens this perspective by using two well-established constructs from developmental psychology: externalizing and internalizing skills. These composite measures capture clusters of behaviors that reflect children's ability to regulate themselves (externalizing skills) and to form a stable self-concept and emotional understanding (internalizing skills). This structure allows us to consider how rank affects a more complete set of behaviors relevant to learning, social interaction, and self-perception, without committing to a single, tightly defined mechanism.

Several strands of the literature suggest that both types of skills may respond to changes in rank. For instance, improvements in externalizing skills may reflect better impulse control or attentional focus among students who feel pressure to maintain a high position in the classroom hierarchy — a pattern consistent with findings on conscientiousness (Pagani et al., 2021). Conversely, internalizing skills may respond to rank through shifts in self-confidence or perceived academic ability, consistent with work showing that students' expectations, self-beliefs, and aspirations are shaped by their standing relative to peers (Elsner et al., 2021; Murphy and Weinhardt, 2020; Elsner and Isphording, 2017). Moreover, psychological evidence links lower internalizing skill development to damaged self-esteem (Creemers et al., 2013), suggesting a plausible pathway through which rank may influence emotional well-being.

By studying the effect of academic rank on these two broader skill dimensions, we provide a more holistic view of how the learning environments shape non-cognitive development. While narrow constructs such as self-confidence or conscientiousness offer clarity in testing specific mechanisms, they may overlook the fact that relative academic standing likely influences multiple, interrelated aspects of a child's psychological development. We try to move beyond these single-trait approaches.

2.4.3 Adapted Specification for the Estimation of the Rank Effect on Non-Cognitive Skills

A key challenge in estimating the effect of academic rank on non-cognitive skills is the lack of a baseline measure of non-cognitive traits before the formation of rank. While the quasi-random assignment of students to school-cohort groups mitigates concerns about selection into rank, controlling only for cognitive skills may not be sufficient. If non-cognitive traits at school entry are correlated with, or causally related to, early academic performance, omitting them could bias the estimated rank effect.

To address this, we turn to insights from developmental psychology. The literature identifies three interrelated dimensions of child development: cognitive, non-cognitive, and physical (Berk, 2023; Santrock and Feldman, 2020). While difficult to fully disentangle, these domains are strongly correlated. We leverage this correlation by using available proxies from the cognitive and physical domains as baseline controls for the unobserved non-cognitive traits at school entry. In particular, Duckworth et al. (2019) show that physical development is often more closely related to non-cognitive skills than cognitive ability is, further motivating our use of anthropometric indicators.

Guided by this evidence, we augment our baseline specification (Equation 3) with two additional controls capturing early physical development. The first, $j(H_{isc})$ is a quadratic polynomial in residualized height measured at the first school medical exam (as explained in subsection 2.3.2). The second, $k(B_{isc})$, is a quadratic polynomial in birth weight, measured in pounds, provided by the children's medical records. The resulting specification is:

$$Y_{isc} = \alpha R_{isc} + \beta g(A_{isc}) + \gamma_1 \overline{A}_{sc-i} + \gamma_2 \sigma_{sc-i} + \theta_1 j(H_{isc}) + \theta_2 k(B_{isc}) + X_i \delta + \lambda_s + \lambda_c + \epsilon_{isc} \quad (6)$$

Here, Y_{isc} is the standardized score for either externalizing or internalizing skills, R_{isc} denotes a student's relative rank, and $g(A_{isc})$ is a quadratic polynomial in individual cognitive ability. $\overline{A}_{sc-i}$ and σ_{sc-i} are, respectively, the mean and standard deviation of cognitive skills within a school-cohort group, excluding the individual. $j(H_{isc})$ and $k(B_{isc})$ capture physical development. X_i includes additional student-level controls, while λ_s and λ_c are school and cohort fixed effects, respectively.

2.5 Results

We now turn to the empirical analysis of the effects of academic rank. Using our measure of relative standing within school-cohort groups — constructed from students' performance on the age 9 test — we estimate the impact of rank on a wide range of outcomes. These include short-run academic achievement, non-cognitive skill development, parental investment, and long-term educational, economic, and health-related outcomes. By relying on variation in rank that is orthogonal to cognitive skills, our analysis isolates the effect of a student's position in the local academic distribution on later-life trajectories.

2.5.1 The Rank Effect on Academic Performance

We estimate the effect of academic rank on students' academic performance, measured through the high-stakes 11-plus examination. As discussed, this test played a central role in determining secondary school placement and had lasting consequences for educational and labor market outcomes (Clark and Del Bono, 2016). In Table 2.3, we present the results. Columns (1) and (2) report estimates using verbal reasoning test (VRT) scores, which are available for all the cohorts in our sample. Columns (3) and (4) show estimates using the overall 11-plus composite score, which is unavailable for the two youngest cohorts.

Across all specifications, we find a positive and statistically significant effect of rank on academic performance. The magnitude of the estimated coefficient is remarkably similar whether we use the VRT or the overall 11-plus score, with only a minimal decline observed when moving to the composite measure. Moreover, controlling for individual characteristics — gender, socioeconomic status, height, weight, birth weight, and number of siblings — does not substantially alter the estimates, as seen from the comparison between columns (1) and (2) and between columns (3) and (4).

The estimated coefficient of approximately 0.55 implies that, holding cognitive skills and other characteristics constant, a student ranked at the top of their school-cohort group would perform about 60% of a standard deviation better on the 11-plus test than a student ranked at the bottom. A more realistic interpretation emerges if we consider the effect of a 10% increase in relative rank. Given that the average school-cohort group size is 37 students (with a median of 29), moving up roughly 4 positions within the group is associated with an increase of 0.055 standard deviations in test scores, equivalent to about 6% of a standard deviation.

Our findings are consistent with previous estimates in the literature. Murphy and Weinhardt (2020) document similar magnitudes when examining the impact of relative rank on academic achievement in English primary schools, while Elsner et al. (2021) find comparable effects using data from German secondary schools. These results reinforce the conclusion that a student's position in the local academic distribution has meaningful consequences for measured academic success, even after controlling for absolute ability.

We assume that the effect of academic rank on outcomes is linear and control for potential non-linearities in cognitive skills by including a quadratic polynomial. To assess the plausibility of this assumption, we break down the rank distribution into deciles and plot the estimated effects in Figure 2.5. The relationship appears broadly linear across the distribution, for the effects on both the VRT and the 11-plus test, supporting our baseline specification. These patterns also showcase the self-reinforcing nature of the rank effect: early differences in rank lead to performance gains that influence future rank, making the timing of cognitive skill measurement less critical.

2.5.2 The Rank Effect on Non-Cognitive Skills

We next estimate the impact of academic rank on non-cognitive skills, focusing on externalizing and internalizing skills as derived from the Rutter Questionnaire. Rather than capturing a single narrow trait, these measures bundle together a set of behaviors relevant to self-regulation, social interaction, emotional stability, and self-perception, offering a broad perspective on children's non-cognitive development.

The estimates are reported in Table 2.4. In columns (1) and (4), we regress externalizing and internalizing skills, respectively, on school-cohort rank, controlling for school and cohort fixed effects, a quadratic polynomial of individual cognitive skills, and the mean and standard deviation of peers' cognitive skills within the school-cohort group. In columns (2) and (5), we further add controls for individual characteristics — sex, socioeconomic status, number of siblings, and month of birth — to account for potential differences in family background and demographic factors. Finally, in columns (3) and (6), we also control for early physical development through height, weight, and birth weight, as these may correlate with non-cognitive traits at school entry and thus strengthen the robustness of our estimates. As stated in subsection 2.2.2, we include only children who were still in primary school as of March 1964, when the teachers completed the Rutter Test's questionnaire (therefore, those in the 7 youngest cohorts, born between April 1952 and October 1955).

We find a positive and relatively stable effect of academic rank on externalizing skills across specifications, although the estimates are imprecisely measured and statistically significant only at the 10% level. A four-position increase within the school-cohort group (approximately 10% of the average number of students in the group, which is 37) is associated with an improvement in externalizing skills of around 3% of a standard deviation. In contrast, the effect on internalizing skills is larger and estimated with greater precision: the same 10% jump in relative rank corresponds to an increase of approximately 4.5% of a standard deviation in internalizing skills.

As shown in Figure 2.4, the distributions of externalizing and internalizing skills exhibit long left tails, reflecting a small number of children with severe behavioral difficulties. To ensure that our estimated rank effects are not disproportionately driven by these extreme cases, we re-estimate our main specifications after excluding the bottom 3% and 5% of each distribution. The results, reported in Table 2.5, reveal a clear contrast between the two skill dimensions. The effect of rank on externalizing skills is highly sensitive to these exclusions. Removing just the bottom 3% of children leads to a sharp drop in the estimated coefficient and a complete loss of statistical significance. By contrast, the effect on internalizing skills remains robust, with only modest reductions in magnitude. A four-position improvement in rank continues to correspond to an increase of roughly 3.7% of a standard deviation, even when the most severe 5% of cases are excluded, compared to about 4.9% in the full sample. This pattern holds even when disaggregating by gender (see subsection 2.5.5), with the effect for boys — who primarily drove the overall estimate — also becoming insignificant. That does not happen to internalizing skills, which maintain their significant effect for both boys and girls. This indicates that the rank effect on internalizing skills is more robust and generalizable across the broader population of students.[10]

Our findings fit naturally within the emerging literature documenting the influence of relative academic standing on non-cognitive development. Prior studies have tended to focus on single traits such as self-esteem, conscientiousness, or expectations (e.g., Elsner et al., 2021; Pagani et al., 2021; Murphy and Weinhardt, 2020). By employing composite measures of externalizing and internalizing skills, we provide a more comprehensive picture of how rank shapes multiple dimensions of children's psychological and behavioral development. While our results do not point to a single definitive mechanism, the estimates performed after removing kids with severe behavioral issues from the sample make clear that not all effects are equally stable: while the estimated effect on internalizing skills holds across different specifications and sample restrictions, the effect on externalizing skills appears far more fragile and dependent on the presence of a small number of extreme values. This suggests that rank may primarily operate by influencing children's self-concept and emotional self-perception, rather than affecting behaviors related to attention or impulse control. This interpretation is consistent with existing evidence that academic rank affects self-confidence, self-esteem, and educational expectations (Elsner et al., 2021; Murphy and Weinhardt, 2020; Elsner and Isphording, 2017). In this sense,

[10] All our main results, including those on academic performance, long-term educational attainment, and adult outcomes, remain statistically significant and qualitatively similar when excluding the most extreme 3% to 5% of cases in the non-cognitive skills distribution, indicating that our findings are not driven by a small subset of extremely problematic students.

the rank effect appears to shape students' view of themselves within the educational environment.

Finally, Figure 2.6 illustrates that the relationship between academic rank and both externalizing and internalizing skills follows an approximately linear trajectory, mirroring the pattern observed for academic performance. This supports the validity of our linear modeling approach and suggests that the impact of relative position on non-cognitive development is broadly proportional across the rank distribution.

2.5.3 The Rank Effect on Parental Investment

An important hypothesis in the rank effect literature concerns the role of parents in shaping children's outcomes. In principle, a child who stands out academically might elicit greater attention and investment from parents, potentially reinforcing the effects of relative standing. However, direct evidence on the role of parental involvement in amplifying or moderating rank effects remains limited. Murphy and Weinhardt (2020) find that academic rank affects children's self-confidence but report limited evidence that it significantly alters parental behavior. An exception is the study by Megalokonomou and Zhang (2024), which shows that, in the Chinese context - where rank is made salient and precisely revealed by the institutional structure of the education system - parents respond to their child's classroom rank by increasing or reducing tutoring efforts.

To explore this dimension, we use information from the Family Survey, which involved interviewing the mothers of a randomly selected subset of children—approximately one-quarter of the full sample. Among those selected, about 80% of families agreed to participate. As shown in Appendix B.1, the randomization into the Family Survey was successful: children whose families participated in the survey do not differ systematically in observable characteristics from those who did not. Furthermore, when we estimate the effect of academic rank on the probability of participating in the Family Survey, we find no evidence of selection based on rank.

We examine the effect of academic rank on several measures of parental involvement. Specifically, we study the probability that a parent knows the child's teacher by name and expresses satisfaction with the child's academic progress; the type of support provided with homework (ranging from autonomy to active parental help); and the amount of time the child spends on homework, using dummies indicating whether the child spends less than 30/45/60 minutes a day.[11]

The results reveal a very limited relationship between academic rank and parental involvement. As shown in Table 2.6, there is a positive effect of rank on the probability that a parent knows the teacher's name, significant at the 10% level. A four-position improvement in academic rank — approximately a 10% increase within the school-cohort group — is associated with a 2 percentage point increase in the probability of parental awareness, corresponding to about 6% of the baseline probability. However, none of the other estimated effects on parental behavior — whether being satisfied with the child's progress, providing help with homework, or influencing time spent on homework — are statistically significant.

Although insignificant, the signs of the estimated coefficients hint at an interesting pattern. Higher-ranked students appear slightly more likely to do homework autonomously or under minimal supervision, while lower-ranked students are somewhat more likely to receive active parental help. This could suggest a weak substitution mechanism, where parents become more directly involved when children appear to struggle. However, given the lack of statistical significance, we cannot confidently assert that parental investment systematically responds to academic rank. Moreover, it is difficult to disentangle whether differences in time spent on homework reflect changes in parental behavior or greater intrinsic motivation and commitment on the part of the child. Overall, our findings suggest that, in our setting, the estimated impact of rank on academic outcomes is unlikely to be

[11]Parental knowledge and satisfaction are based on whether the parent reports knowing the child's teacher's name and being happy with the child's school progress. Homework help is captured through three variables: (i) whether the child does homework autonomously, (ii) whether the parent checks or supervises the homework, and (iii) whether the parent provides active help. Parents categorize their child's homework habits among the following: "No homework given", "Child does homework on their own", "Parent just sees it is done", "Parent checks homework", or "Parent gives active help". We define children as autonomous as those in the "Child does homework on their own" category. We defined children being checked as those in the "Parent just sees it is done" and "Parent checks homework" categories. We define children receiving help as those in the "Parent gives active help" category. Time spent on homework is reported in intervals (e.g., 0–15 minutes, 15–30 minutes, etc.), and we create dummies indicating whether the child spends less than 30, 45, or 60 minutes per day on homework.

primarily mediated through shifts in parental involvement.

2.5.4 The Rank Effect on Long-Term Outcomes

We next examine the long-term consequences of academic rank, using responses from a follow-up survey conducted by mail in 2001, nearly 40 years after the original data collection. The response rate for the follow-up was approximately 60%. While the existing literature has documented short- and medium-term effects of rank on academic and non-cognitive outcomes, evidence on long-term impacts remains limited. Denning et al. (2023) find persistent effects on earnings 25 years after school, but in general, the literature tends to assume that early impacts on education and non-cognitive skills naturally translate into longer-run differences without direct empirical confirmation.

Comparing observable characteristics, we show in Appendix B.1 that respondents to the 2001 follow-up were more likely to be women, to come from higher socioeconomic backgrounds, to have had better grades and stronger externalizing skills in primary school, and to come from smaller families. However, crucially, there are no significant differences in academic rank between respondents and non-respondents. While sample selection could bias levels of long-term outcomes, the fact that it is unrelated to rank supports the validity of our estimates for studying the causal impact of academic rank. Moreover, we control for a wide range of baseline characteristics in all regressions, further mitigating concerns about differential selection.

We start by analyzing how individuals retrospectively recall their primary school experience.[12] We also study educational attainment.[13]

[12]Primary school experience is captured through three variables. First, self-reported happiness in primary school, based on the question "Were you happy in primary school?" with possible answers being "Very happy". "Fairly happy", "Neither happy or unhappy", "Not very happy", and "Not at all happy". We create a dummy equal to one for respondents reporting being either "Very happy" or "Fairly happy". Second, perceptions of social relationships are assessed through the question "How many friends did you have in primary school compared to other children?", allowing for the answers "More", "Same number", or "Fewer". We construct two dummies for having more or fewer friends. Third, to investigate exposure to bullying, we use a dummy for affirmative responses to "Were you bullied in primary school?".

[13]Educational achievement is measured using several binary indicators: (i) attendance at grammar school (a selective secondary school for higher-achieving students based on the 11-plus exam results), (ii) attainment of O-level qualifications (exams typically taken at age 16), (iii) attainment of A-level qualifications (advanced exams taken at age 18), and (iv) attainment of a university degree.

The results, presented in Table 2.7, show mixed evidence. Rank has a positive and statistically significant effect on the probability of recalling being happy in primary school, significant at the 5% level. A four-position improvement in rank—about a 10% shift in the group—raises the likelihood of reporting happiness by approximately 1.5% relative to the baseline. However, we find no significant effects of rank on the recalled number of friends or on reported experiences of being bullied.

As expected, the effects of academic rank are much more pronounced on educational outcomes. A four-position increase in rank is associated with a 4 percentage point rise in the probability of attending grammar school — an increase of roughly 20% relative to the baseline rate. Similarly, we find positive and significant effects on achieving O-level and A-level qualifications, with a four-position jump corresponding to a 5% and 12% increase in probability, respectively. In contrast, we find no evidence that academic rank affects the likelihood of completing a university degree: the estimated coefficient is close to zero and imprecisely measured. This may reflect historical context, as university attendance was considerably less common during the period when these cohorts reached adulthood, and admission standards may have been stricter or less influenced by early academic differences.

We next examine the impact of academic rank on socioeconomic status, earnings, family formation, and subjective well-being in adulthood.[14] We also consider the probability of having ever married or having children, based on self-reported answers to direct questions about marital and parental status.[15] Finally, we include two measures of subjective well-being: enjoyment of daily activities and general happiness.[16]

Results are reported in Table 2.8. Overall, we find limited evidence that academic rank has a meaningful long-term effect on socioeconomic status or income. While the

[14] Socioeconomic status takes value one if the occupation of an individual is "Legislators, senior officials, and managers", and zero otherwise. It is based on the one-digit occupational code from ISCO-88. Income is reported in brackets, with thresholds at £1,000, £4,000, £8,000, £12,500, £17,500, £25,000, £35,000, and £45,000. We examine the probability of earning more than £25,000, £35,000, and £45,000 per year.

[15] Participants answered "Have you ever been married?" and "Do you have kids?", which we use to define dummy variables representing the probability of marriage and of having kids.

[16] Participants were asked "How do you feel you enjoy daily activities?" and "Do you feel happy in your everyday life?" with possible responses being "More than usual", "Same as usual", "Less than usual", and "Much less than usual". We construct two dummies: one indicating whether the participant enjoys daily activities the same or more than usual, and another for feeling happy the same or more than usual.

estimated coefficients are generally positive, they are imprecisely measured and generally statistically insignificant. We detect a modest effect of rank on the probability of earning more than £25,000 per year, significant at the 10% level: a four-position improvement in school-cohort rank corresponds to a 2.2% increase in the likelihood of earning above this threshold relative to the baseline probability. In 2001, the median household income in the UK was approximately £20,000 (Office for National Statistics, 2003), making the £25,000 threshold a reasonable proxy for identifying individuals in the upper half of the income distribution.

Turning to family outcomes, we find no significant effect of academic rank on the probability of marrying or having children. Similarly, there is no evidence that higher academic rank affects adult subjective well-being, as measured by enjoyment of daily activities or general happiness. In each case, the point estimates are small and statistically insignificant, suggesting that any potential impact of primary school rank on these aspects of adult life is minimal or absent.

Overall, our analysis indicates that while academic rank has persistent effects on educational achievement, its influence on broader long-term outcomes appears to be limited. It is important to acknowledge that the sample used for the 2001 follow-up survey consists largely of individuals who remained in Aberdeen or the surrounding areas, which may attenuate the estimated effects on socioeconomic and life outcomes. The implications of this selection are not entirely clear: if higher-achieving individuals were more likely to migrate out of Aberdeen and subsequently less likely to respond to the follow-up, we may be missing a segment of the population for whom the rank effect could be particularly impactful. On the other hand, negative selection into the follow-up sample — where respondents are relatively less mobile or lower achieving — could therefore contribute to the limited long-term effects we observe.

2.5.5 Gender Heterogeneity

Gender is a natural dimension along which to explore heterogeneity in the rank effect, given the pronounced differences in boys' and girls' cognitive and non-cognitive profiles during childhood. In our sample, girls exhibit stronger cognitive skills on average (mean of 0.09 vs -0.08 for boys), significantly higher externalizing skills (0.18 vs -0.17), and slightly lower internalizing skills (-0.05 vs 0.05). These initial disparities may shape how boys and girls respond to their relative academic standing within their school-cohort groups.

Table 2.9 presents gender-specific estimates of the rank effect. We find that the impact on academic performance is approximately 30% larger for girls than for boys. For non-cognitive outcomes, boys drive the rank effect on externalizing skills, beginning from a lower baseline and exhibiting a stronger response to higher relative rank. In contrast, girls show a substantially larger and more precisely estimated rank effect on internalizing skills, consistent with their lower initial levels of these traits. However, we show in Table 2.10 that the effect on externalizing skills is highly sensitive to the exclusion of extreme cases: when we remove the bottom 3–5% of students with the most severe externalizing behaviors, the estimated impact becomes small and statistically insignificant for both boys and girls. By contrast, the rank effect on internalizing skills remains robust across these sample restrictions, retaining statistical significance and meaningful effect sizes for both genders. This provides further evidence that the overall rank effect is primarily driven by changes in self-concept and emotional adjustment, as captured by internalizing skills, rather than by improvements in behavioral regulation.

These gender differences extend to other outcomes as well. In terms of parental investment, the small but statistically significant overall effect on parental awareness of the teacher's name appears to be driven by girls, though the gender difference is not itself statistically significant. For educational attainment, rank boosts the probability of grammar school attendance for both genders, but the effect is larger for girls (22% vs 14% increase relative to their respective baselines for a 4-position jump within the school-cohort group). Interestingly, the rank effect on completing O-levels is about 50% higher for boys than for girls, while the effect on A-levels is nearly identical. These findings likely reflect the gender norms of the period, where girls who excelled academically were more likely to pursue selective secondary education but faced greater constraints in progressing to university.

Finally, the modest positive effect of rank on adult earnings observed in the full sample appears to be concentrated among boys. For them, a four-position increase in rank leads to a 1.8 percentage point increase in the probability of earning more than £25,000 per year (2.8% of the baseline), statistically significant at the 5% level. This effect is absent for girls, and the gender difference is itself statistically significant. These results suggest that while academic rank has broadly positive effects for both genders, the long-term economic returns may have been more accessible to boys in this historical context, possibly due to gendered labor market structures and educational pathways. For the other outcomes explored in previous sections, gender heterogeneity does not appear to play a major role.

2.5.6 Robustness of the Results: Using an Alternative Identifying Variation

The identifying variation we exploit comes from differences in academic rank between school-cohort groups, leveraging the quasi-random assignment of students to these groups. Our baseline approach controls for school and cohort fixed effects separately, effectively assuming that any unobserved characteristics specific to a school (e.g., teaching quality, peer norms) or cohort (e.g., year-specific shocks, changes in policy) can be controlled for additively. While our balancing tests and randomization checks provide strong evidence that this assumption holds in our setting, it remains a relatively strong restriction, as it implies that no unobserved factor simultaneously influences both the school and cohort components in a way that is correlated with students' academic rank.

To assess the robustness of our findings to this assumption, we re-estimate our main results using school-cohort fixed effects, which absorb all common variation within each school-cohort cell. This approach captures any shared characteristics that might jointly affect all students within a given peer group, controlling for group-specific unobserved heterogeneity. While this specification sacrifices some statistical power by narrowing the scope of identifying variation to differences between rather than within school-cohort groups, it provides a more conservative but stringent test of the rank effect.

Main results from the specification using school-cohort fixed effects are presented in

Table 2.11. Overall, this alternative approach yields slightly larger point estimates for most outcomes, though the differences are small compared to those obtained with separate school and cohort fixed effects. The only notable exception is the coefficient on income, which becomes smaller and statistically insignificant when school-cohort fixed effects are used. However, the gender-specific analysis reveals that the previously identified rank effect on boys' earnings remains robust — both in magnitude and statistical significance — even under this more stringent specification. Taken together, the results indicate that the choice of fixed effects structure has only a minor impact on our estimated effects, lending further credibility to our main findings.

2.6 Conclusion

We examined the consequences of academic rank within school-cohort peer groups, using rich data on the entire population of children enrolled in Aberdeen (Scotland) primary schools in 1962. Exploiting quasi-random variation in peer group composition, we identify the causal impact of a student's relative academic standing, conditional on ability, on a wide range of short- and long-term outcomes. We leverage detailed school records, teacher-reported behavioral assessments, randomized family interviews, and a follow-up survey conducted nearly four decades later to provide a comprehensive view of how rank shapes trajectories from childhood to adulthood.

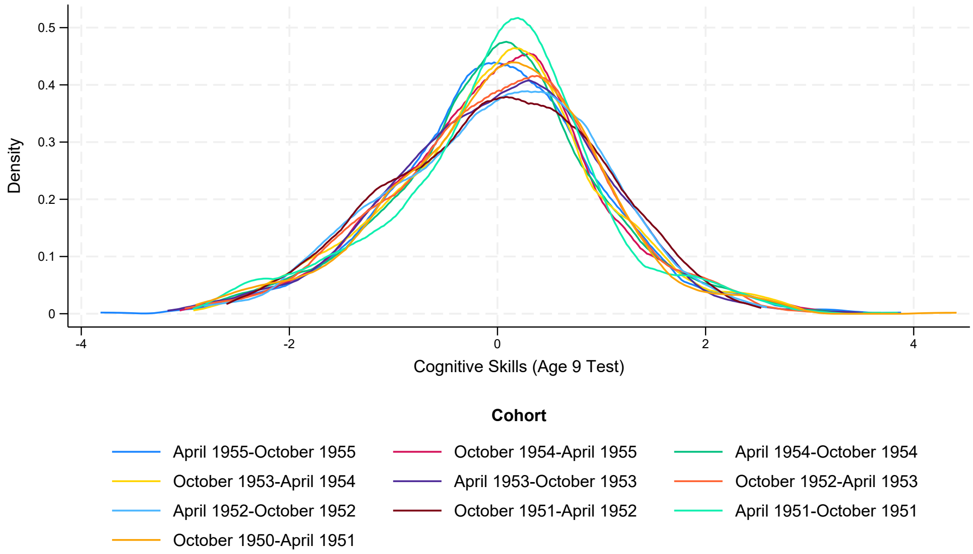
We find that academic rank has a substantial effect on students' academic performance. A 10% improvement in relative position — equivalent to a four-place jump in rank within the average group — raises test scores on the high-stakes 11-plus exam by around 6% of a standard deviation. These effects are larger for girls, who also begin with higher baseline academic achievement. We also show that the same increase in rank improves internalizing skills, such as confidence and self-concept, by around 4.5% of a standard deviation. This effect is robust across specifications and not driven by extreme behavioral cases. In contrast, the impact on externalizing skills (e.g., impulse control, attention) is weaker and sensitive to sample restrictions. These findings suggest that rank shapes how students perceive themselves more than how they regulate their behavior.

Parental investment does not appear to systematically adjust in response to rank. While higher-ranked children are slightly more likely to have parents who know their teacher's name, we detect no meaningful changes in homework support or time allocation. This suggests that rank effects operate primarily through children's internal responses rather than shifts in parental behavior. In the long term, we find that academic rank continues to shape educational attainment, especially for girls. Rank increases the probability of attending grammar school and completing formal secondary qualifications (O-levels and A-levels). However, these educational gains do not translate uniformly into future earnings. Rank improves the probability of having an income above £25,000 only for boys, and we detect no effect on occupational status, marriage, fertility, or adult well-being. The gender asymmetry in long-term outcomes likely reflects the historical context, when educational opportunities for women were expanding, but labor market access remained restricted.

These results make several contributions to the literature. We provide the most detailed evidence to date on the role of academic rank in shaping non-cognitive development. Unlike previous studies that focus on single-trait outcomes such as self-confidence or expectations (e.g., Elsner et al., 2021; Murphy and Weinhardt, 2020; Elsner and Isphording, 2017), we use validated teacher-reported measures and apply factor analysis to capture broader constructs of internalizing and externalizing behavior. Our findings highlight that rank has a particularly robust effect on internalizing traits, aligning with psychological theories that link social comparison to identity formation and self-efficacy.

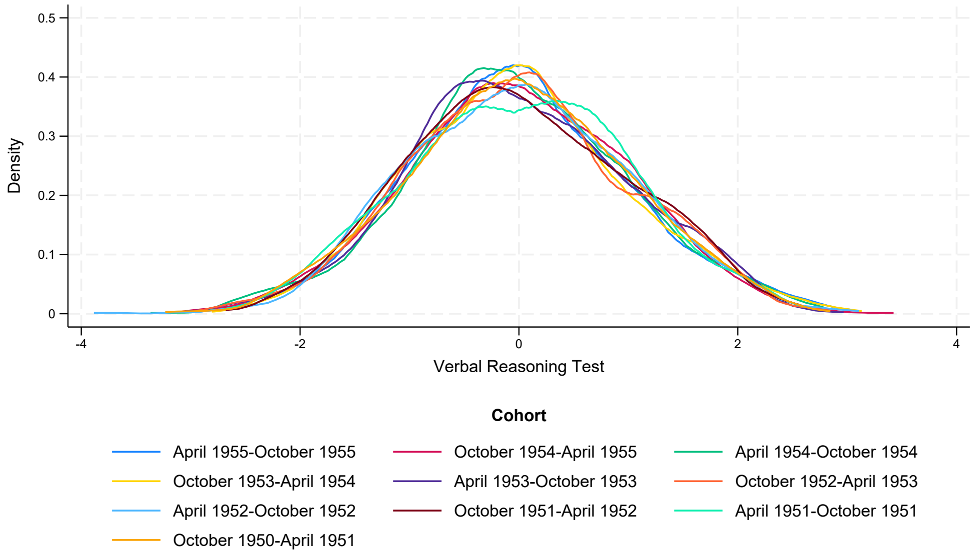
We also find empirical evidence on the long-term consequences of academic rank, nearly 40 years after the reference point. While recent work has established effects on educational choice and earnings in the short run (e.g., Denning et al., 2023; Goulas et al., 2023), we show that these effects persist well into adulthood — but unevenly so. Educational returns are more visible for girls, while income effects are concentrated among boys. These patterns point to important constraints — social, institutional, and historical — on the extent to which academic advantages can translate into broader life outcomes. More broadly, our findings reinforce the idea that ranking within a peer group is not just a reflection of performance but a determinant of future development, with implications for how we structure peer interactions in schools.

Figure 2.1: Distribution of Cognitive Skills, by Cohort



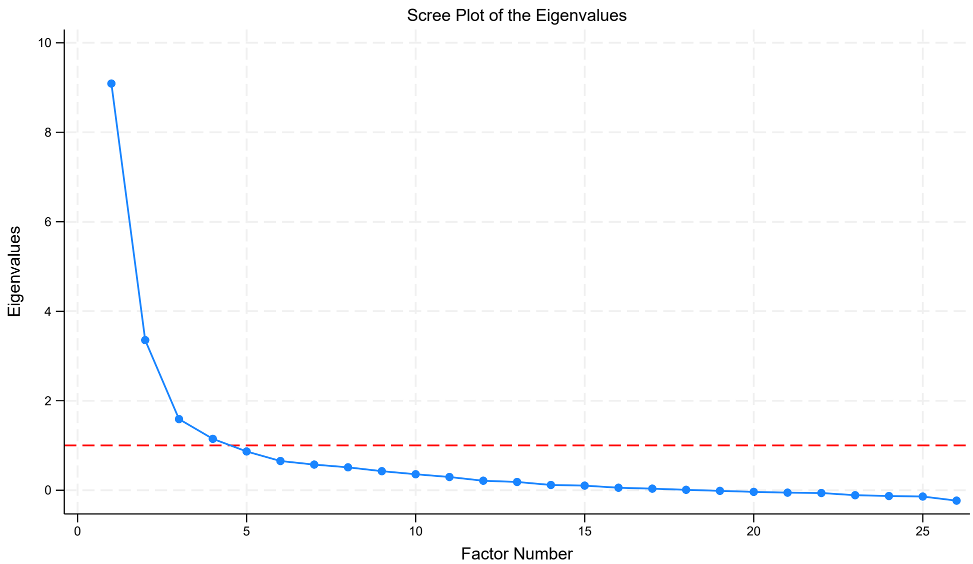
Notes: We show the distribution of students' cognitive skills, proxied by the outcomes of the Age 9 Test, by cohort. The test scores are standardized at the cohort level. We include only children born between October 1950 and October 1955. The number of observations is 9,368.

Figure 2.2: Distribution of the Verbal Reasoning Test, by Cohort



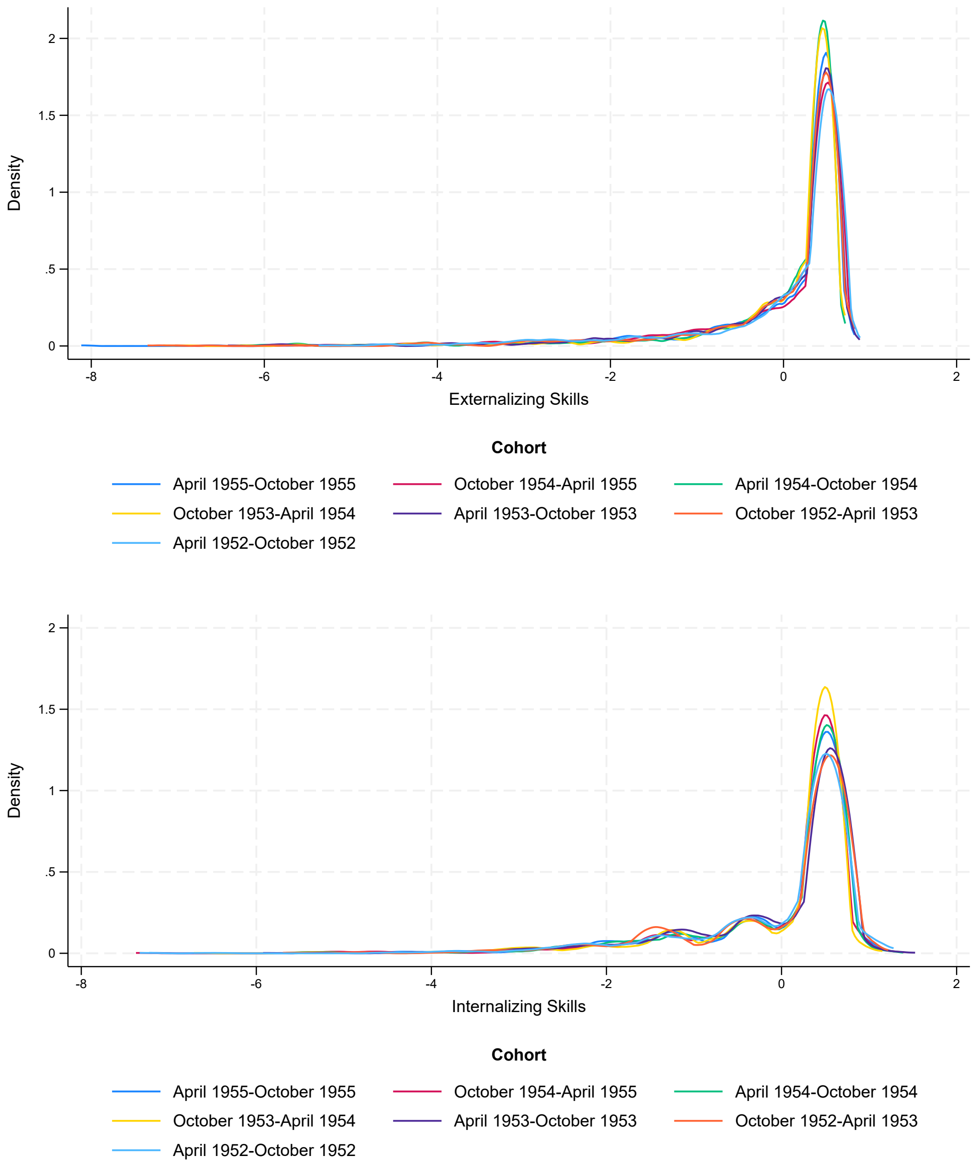
Notes: We show the distribution of the outcomes of the Verbal Reasoning Test, by cohort. The test scores are standardized at the cohort level. We include only children born between October 1950 and October 1955. The number of observations is 9,698.

Figure 2.3: Distribution of Externalizing and Internalizing Skills, by Cohort



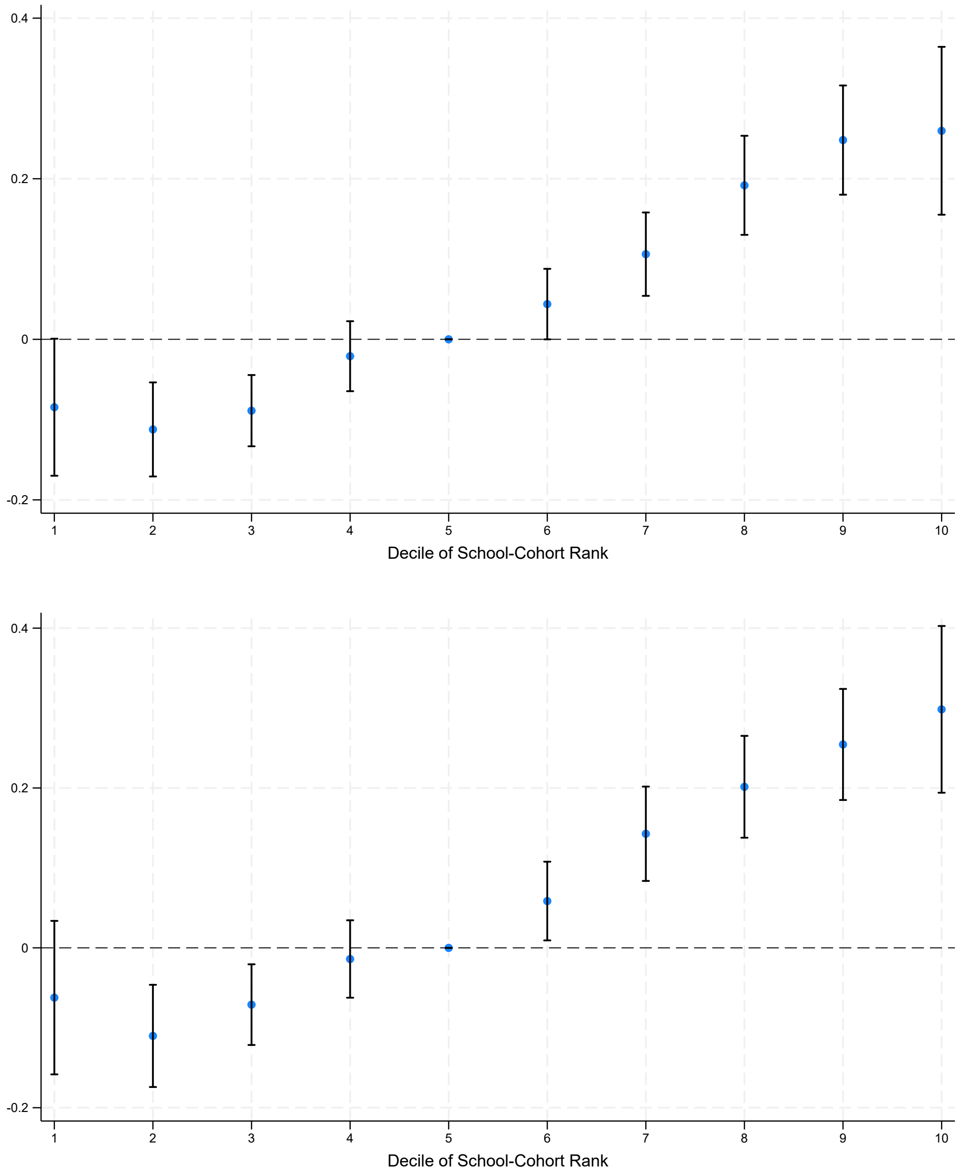
Notes: The graph plots the eigenvalue of each factor estimated through the first iteration of factor analysis. The number of observations is 6,779.

Figure 2.4: Distribution of Externalizing and Internalizing Skills, by Cohort



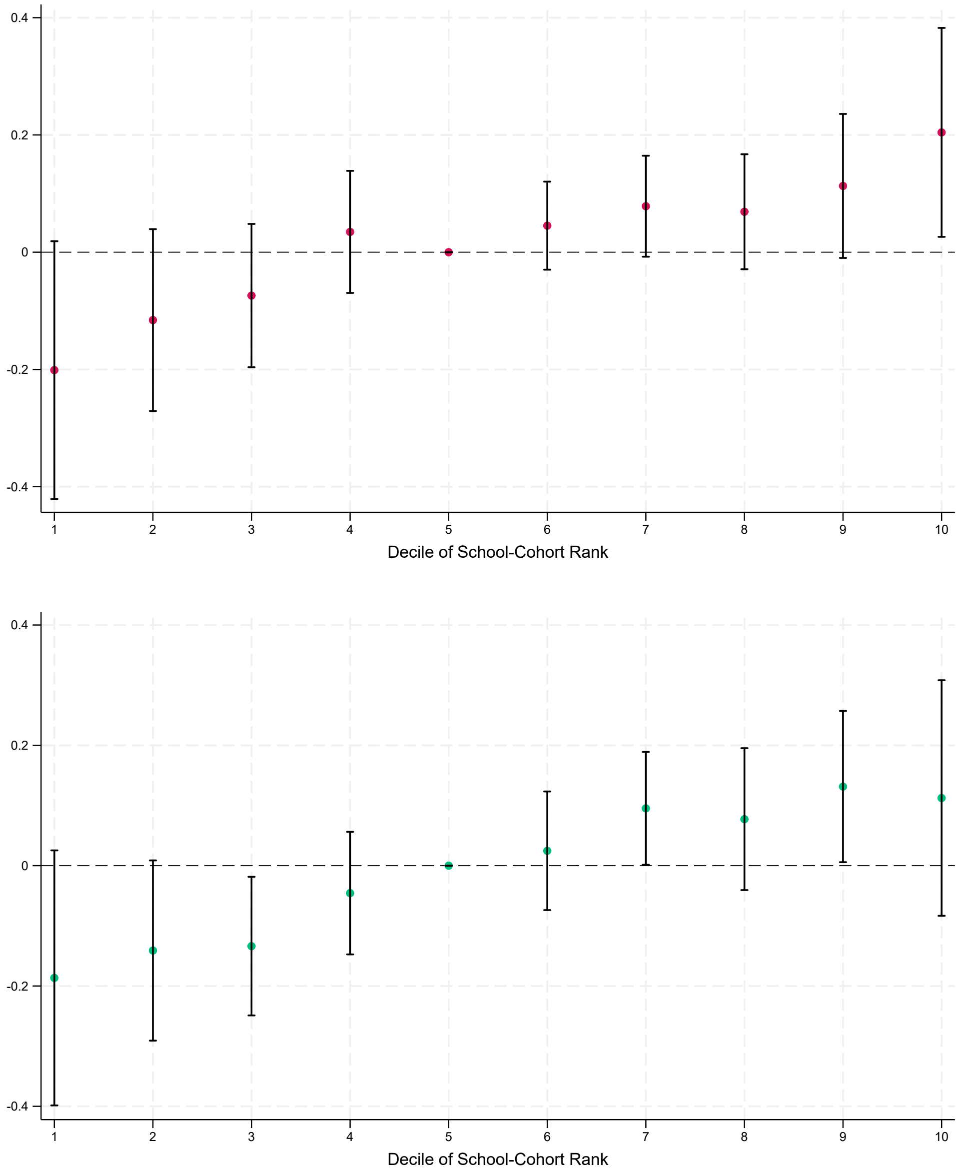
Notes: The graphs show the distribution of externalizing skills in the top panel and internalizing skills in the bottom panel, by cohort. The variables are standardized at the cohort level. We include only children born between April 1952 and October 1955. The number of observations is 6,779.

Figure 2.5: Rank effect on the (standardized) outcome of the Verbal Reasoning and 11-plus tests, by rank decile



Notes: The graph shows the rank effect on the score attained in the Verbal Reasoning Test in the top panel and the 11-plus Test in the bottom panel, by rank decile. We are estimating Equation 3 by replacing the independent variable percentile rank with a categorical variable taking the values of the deciles of school-cohort ranking to which the student belongs, keeping the fifth decile as the baseline.

Figure 2.6: Rank effect on (standardized) externalizing and internalizing skills, by rank decile



Notes: The graph shows the rank effect on externalizing skills in the top panel and internalizing skills in the bottom panel, by rank decile. We are estimating Equation 6 by replacing the independent variable percentile rank with a categorical variable taking the values of the deciles of school-cohort ranking to which the student belongs, keeping the fifth decile as the baseline.

Table 2.1: Balancing Exercise: Individual Characteristics, Rank, and Peer Cognitive Skills

Variables	Woman	High SES	Height	Weight	Birth Weight	Siblings
Equation 4: Conditional relation between individual characteristics and rank						
Percentile Rank	0.018 (0.045)	0.017 (0.027)	0.167 (0.105)	0.165* (0.096)	0.052 (0.097)	-0.089 (0.063)
Equation 5: Conditional relation between individual characteristics and peer quality						
Mean of Peer Cognitive Skills	-0.0001 (0.007)	-0.003 (0.004)	-0.015 (0.023)	-0.024 (0.018)	0.005 (0.017)	0.008 (0.011)
Standard Deviation of Peer Cognitive Skills	-0.005 (0.005)	0.005 (0.003)	-0.019 (0.016)	-0.011 (0.011)	0.001 (0.011)	0.0001 (0.008)
25th Percentile of Peer Cognitive Skills	0.004 (0.007)	-0.004 (0.004)	-0.011 (0.022)	-0.014 (0.016)	-0.001 (0.015)	-0.006 (0.010)
50th Percentile of Peer Cognitive Skills	-0.004 (0.006)	-0.004 (0.003)	0.001 (0.020)	-0.001 (0.015)	0.020 (0.015)	0.010 (0.010)
75th Percentile of Peer Cognitive Skills	0.001 (0.006)	-0.001 (0.003)	-0.010 (0.016)	-0.017 (0.013)	0.006 (0.015)	0.010 (0.010)
Observations	9,698	9,698	9,465	9,458	9,698	9,698

Notes: We estimate the relationship between ranking (Equation 4)/peer cognitive skills (Equation 5) at the school-cohort group level on different characteristics of the students. These characteristics are: the student probability of being a girl, the student probability of coming from an advantaged socioeconomic background (based on the father's occupation), the student height and weight at the time of their first medical exam, the student birth weight (lbs), and the student number of siblings. We include all 10 cohorts of children who attended primary school in Aberdeen in December 1962, who were born between October 1950 and October 1955. Standard errors are clustered at the school-cohort-group level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2.2: Rotated Factor Loadings from the Exploratory Factor Analysis based on the 26 items of the Rutter Questionnaire for Teachers

	Iteration 1		Iteration 2	
Item	Externalizing	Internalizing	Externalizing	Internalizing
Restless	**0.78**	0.00	**0.79**	-0.02
Truant	**0.68**	0.18	**0.67**	0.16
Fidgety	**0.77**	0.01	**0.76**	-0.01
Destroys Belongings	**0.89**	-0.07	**0.89**	-0.07
Fights Others	**0.87**	-0.04	**0.88**	-0.004
Disliked	**0.67**	0.33	**0.68**	0.33
Anxious	-0.16	**0.85**	-0.15	**0.86**
Solitary	0.11	**0.62**	0.12	**0.60**
Irritable	**0.75**	0.04	**0.76**	0.06
Often Unhappy and Miserable	0.22	**0.75**	0.24	**0.76**
Tics	0.39	0.32	-	-
Sucks Finger	0.26	0.25	-	-
Nail Biting	0.24	0.13	-	-
Trivial Absences	0.38	0.34	-	-
Disobedient	**0.87**	-0.12	**0.87**	-0.11
Poor Concentration	**0.57**	0.24	**0.56**	0.20
Afraid	-0.14	**0.85**	-0.12	**0.84**
Fussy over particular child	-0.18	**0.55**	-0.16	**0.58**
Often Lies	**0.86**	0.004	**0.86**	0.01
Stealing	**0.71**	-0.02	**0.70**	0.003
Wet/Soiled Themselves	0.26	0.29	-	-
Often Aching	0.16	**0.53**	0.17	**0.49**
Tearful	0.20	**0.63**	0.21	**0.65**
Stutters	0.20	0.29	-	-
Speech Difficulties	0.19	0.21	-	-
Bullies Others	**0.85**	-0.11	**0.85**	-0.09

Notes: We iterate exploratory factor analysis to decide which items to retain out of the 26 in the Rutter Questionnaire for Teachers. We report the factor loadings and the communities for the oblique rotated total variance matrix. We restrict our sample to children born between April 1952 and October 1955, since we want to include only children who were in primary school when the Rutter Questionnaire was completed (March 1964). In total, we have 6,779 children.

Table 2.3: Rank Effect on the 11-plus Test

Outcome Variables	(1)	(2)	(3)	(4)
	Verbal Reasoning Test		11-plus Test	
Percentile Rank	0.575***	0.579***	0.540***	0.544***
	(0.068)	(0.068)	(0.074)	(0.074)
Mean of the Outcome	0	0	0	0
SD of the Outcome	1	1	1	1
Observations	9,441	9,441	7,575	7,575
School Fixed Effects	X	X	X	X
Cohort Fixed Effects	X	X	X	X
Cognitive Skills	X	X	X	X
Cognitive Skills Squared	X	X	X	X
Mean of Peer Cognitive Skills	X	X	X	X
SD of Peer Cognitive Skills	X	X	X	X
Sex	-	X	-	X
Socioeconomic Status	-	X	-	X
Number of Siblings	-	X	-	X
Month of Birth	-	X	-	X

Notes: We estimate the relationship percentile rank within the school-cohort group and the standardized outcome of the Verbal Reasoning Test, as well as the standardized outcome of the 11-plus Test. Percentile rank is established using our baseline measure of cognitive skills, the outcome of the Age 9 Test. The 11-plus Test consists of 4 components: two Verbal Reasoning Tests, one Algebra Test, and one English Test. We consider the outcome of the Verbal Reasoning Tests first, as it is available for all 10 cohorts of children in our survey; the overall outcome of the 11-plus test is not available for the two youngest cohorts of children in our sample. Our sample consists of 10 cohorts of children who attended primary school in Aberdeen in December 1962, who were born between October 1950 and October 1955. We control for: a categorical variable taking a different value for each school in the sample; a categorical variable taking a different value for each cohort in the sample; a quadratic polynomial of child cognitive skills (based on the outcome of the Age 9 Test); the mean and standard deviation of the cognitive skills of the peers of the students (based on the outcome of the Age 9 Test); a categorical variable taking value 1 if the child is a girl, and 0 if he is a boy; a categorical variable taking value 1 if the child belongs to a family of high socioeconomic status (defined based on the father's occupation), and 0 otherwise; a categorical variable capturing the specific month of birth of the child; and the number of siblings of the child. Standard errors are clustered at the school-cohort level.

Table 2.4: Rank Effect on the Externalizing and Internalizing Skills

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome Variables	Externalizing Skills			Internalizing Skills		
Percentile Rank	0.285*	0.272*	0.276*	0.470***	0.483***	0.487***
	(0.157)	(0.155)	(0.154)	(0.153)	(0.153)	(0.158)
Mean of the Outcome	0	0	0	0	0	0
SD of the Outcome	1	1	1	1	1	1
Observations	6,631	6,516	6,516	6,631	6,516	6,516
School Fixed Effects	X	X	X	X	X	X
Cohort Fixed Effects	X	X	X	X	X	X
Cognitive Skills	X	X	X	X	X	X
Cognitive Skills Squared	X	X	X	X	X	X
Mean of Peer Cognitive Skills	X	X	X	X	X	X
SD of Peer Cognitive Skills	X	X	X	X	X	X
Sex	-	X	X	-	X	X
Socioeconomic Status	-	X	X	-	X	X
Number of Siblings	-	X	X	-	X	X
Month of Birth	-	X	X	-	X	X
Height	-	-	X	-	-	X
Height Squared	-	-	X	-	-	X
Birth Weight	-	-	X	-	-	X
Birth Weight Squared	-	-	X	-	-	X

Notes: We estimate the relationship percentile rank within the school-cohort group and the standardized measures of externalizing and internalizing skills. Percentile rank is established using our baseline measure of cognitive skills, the outcome of the Age 9 Test. The individual measures of externalizing and internalizing skills are estimated using common factor analysis on the 26 items of the Rutter Questionnaire for Teachers, completed in March 1964. We include only children who were still in primary school as of March 1964; therefore, those in the 7 youngest cohorts, born between April 1952 and October 1955. We control for: a categorical variable taking a different value for each school in the sample; a categorical variable taking a different value for each cohort in the sample; a quadratic polynomial of child cognitive skills (based on the outcome of the Age 9 Test); the mean and standard deviation of the cognitive skills of the peers of the students (based on the outcome of the Age 9 Test); a quadratic polynomial of the standardized height measured during the first medical exam in school and a quadratic polynomial of birth weight; a categorical variable taking value 1 if the child is a girl, and 0 if he is a boy; a categorical variable taking value 1 if the child belongs to a family of high socioeconomic status (defined based on the father's occupation), and 0 otherwise; a categorical variable capturing the specific month of birth of the child; and the number of siblings of the child. Standard errors are clustered at the school-cohort level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2.5: Removing Extreme Children with Extreme Behavioral Issues

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome Variables	Externalizing Skills			Internalizing Skills		
No Bottom	0%	3%	5%	0%	3%	5%
Percentile Rank	0.276*	0.084	0.051	0.487***	0.335***	0.256***
	(0.153)	(0.109)	(0.096)	(0.158)	(0.113)	(0.094)
Mean of the Outcome	0	0.13	0.18	0	0.11	0.17
SD of the Outcome	1	0.67	0.56	1	0.76	0.68
Observations	6,516	6,324	6,200	6,516	6,313	6,179
School Fixed Effects	X	X	X	X	X	X
Cohort Fixed Effects	X	X	X	X	X	X
Cognitive Skills	X	X	X	X	X	X
Cognitive Skills Squared	X	X	X	X	X	X
Mean of Peer Cognitive Skills	X	X	X	X	X	X
SD of Peer Cognitive Skills	X	X	X	X	X	X
Sex	X	X	X	X	X	X
Socioeconomic Status	X	X	X	X	X	X
Number of Siblings	X	X	X	X	X	X
Month of Birth	X	X	X	X	X	X
Height	X	X	X	X	X	X
Height Squared	X	X	X	X	X	X
Birth Weight	X	X	X	X	X	X
Birth Weight Squared	X	X	X	X	X	X

Notes: We estimate the relationship percentile rank within the school-cohort group and the standardized measures of externalizing and internalizing skills, progressively excluding the children with the most severe externalizing or internalizing issues. We start from the full sample (columns (1) and (4)), then exclude the bottom 3% (columns (2) and (5)), and finally the bottom 5% (columns (3) and (6)). Percentile rank is established using our baseline measure of cognitive skills, the outcome of the Age 9 Test. We include only children who were still in primary school as of March 1964; therefore, those in the 7 youngest cohorts, born between April 1952 and October 1955. We control for: a categorical variable taking a different value for each school in the sample; a categorical variable taking a different value for each cohort in the sample; a quadratic polynomial of child cognitive skills (based on the outcome of the Age 9 Test); the mean and standard deviation of the cognitive skills of the peers of the students (based on the outcome of the Age 9 Test); a quadratic polynomial of the standardized height measured during the first medical exam in school and a quadratic polynomial of birth weight; a categorical variable taking value 1 if the child is a girl, and 0 if he is a boy; a categorical variable taking value 1 if the child belongs to a family of high socioeconomic status (defined based on the father's occupation), and 0 otherwise; a categorical variable capturing the specific month of birth of the child; and the number of siblings of the child. Standard errors are clustered at the school-cohort level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2.6: Rank Effect on Parental Investment

Outcome Variables	(1) Awareness of Teacher's Name	(2) Happy with Progress	(3) Homework: Autonomous	(4) Homework: Checked	(5) Homework: Helped	(6) Daily Time on Homework: < 30	(7) Daily Time on Homework: < 45	(8) Daily Time on Homework: < 60
Percentile Rank	0.197*	0.083	0.075	0.110	-0.158	0.022	-0.070	-0.043
	(0.117)	(0.104)	(0.111)	(0.136)	(0.126)	(0.090)	(0.131)	(0.116)
Mean of the Outcome	0.68	0.80	0.22	0.42	0.36	0.12	0.47	0.68
SD of the Outcome	0.47	0.48	0.41	0.49	0.50	0.33	0.50	0.47
Observations	1,787	1,770	1,787	1,787	1,787	1,787	1,787	1,787
School Fixed Effects	X	X	X	X	X	X	X	X
Cohort Fixed Effects	X	X	X	X	X	X	X	X
Cognitive Skills	X	X	X	X	X	X	X	X
Cognitive Skills Squared	X	X	X	X	X	X	X	X
Mean of Peer Cognitive Skills	X	X	X	X	X	X	X	X
SD of Peer Cognitive Skills	X	X	X	X	X	X	X	X
Sex	X	X	X	X	X	X	X	X
Socioeconomic Status	X	X	X	X	X	X	X	X
Number of Siblings	X	X	X	X	X	X	X	X
Month of Birth	X	X	X	X	X	X	X	X

Notes: We estimate the relationship percentile rank within the school-cohort group and different outcomes. Percentile rank is established using our baseline measure of cognitive skills, the outcome of the Age 9 Test. These outcomes are derived from the interviews conducted in the context of the Family Survey. The different outcomes are: the probability that the parents know the teacher's name (column (1)); the probability that the parents are happy about the child's progress in school (column (2)); the probability that the child does his homework autonomously (column (3)), with the parent checking them (column (4)), or with the parent's help (column (5)); the probability that the parents reports that the child spends a daily time on homework below 30 minutes (column (6)), below 45 minutes (column (7)), or below 60 minutes (column (8)). Our sample consists of 10 cohorts of children who attended primary school in Aberdeen in December 1962, who were born between October 1950 and October 1955, and who participated in the Family Survey. We control for: a categorical variable taking a different value for each school in the sample; a categorical variable taking a different value for each cohort in the sample; a quadratic polynomial of child cognitive skills (based on the outcome of the Age 9 Test); the mean and standard deviation of the cognitive skills of the peers of the students (based on the outcome of the Age 9 Test); a categorical variable taking value 1 if the child is a girl, and 0 if he is a boy; a categorical variable taking value 1 if the child belongs to a family of high socioeconomic status (defined based on the father's occupation), and 0 otherwise; a categorical variable capturing the specific month of birth of the child; and the number of siblings of the child. Standard errors are clustered at the school-cohort level.

Table 2.7: Rank Effect on Long-Term Outcomes: Primary School Memories and Academic Achievement

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Primary School Memories:				Academic Achievement:			
Outcome Variables	Happy	More Friends	Less Friends	Bullied	Grammar School	O-Level	A-Level	Degree
Percentile Rank	0.128**	0.024	-0.037	0.011	0.398***	0.295***	0.411***	0.004
	(0.057)	(0.033)	(0.041)	(0.069)	(0.054)	(0.065)	(0.058)	(0.047)
Mean of the Outcome	0.81	0.05	0.09	0.24	0.20	0.60	0.33	0.15
SD of the Outcome	0.39	0.21	0.28	0.43	0.40	0.49	0.47	0.36
Observations	6,631	6,516	6,516	6,631	6,516	6,516	6,516	6,516
School Fixed Effects	X	X	X	X	X	X	X	X
Cohort Fixed Effects	X	X	X	X	X	X	X	X
Cognitive Skills	X	X	X	X	X	X	X	X
Cognitive Skills Squared	X	X	X	X	X	X	X	X
Mean of Peer Cognitive Skills	X	X	X	X	X	X	X	X
SD of Peer Cognitive Skills	X	X	X	X	X	X	X	X
Sex	X	X	X	X	X	X	X	X
Socioeconomic Status	X	X	X	X	X	X	X	X
Number of Siblings	X	X	X	X	X	X	X	X
Month of Birth	X	X	X	X	X	X	X	X

Notes: We estimate the relationship percentile rank within the school-cohort group and different outcomes. Percentile rank is established using our baseline measure of cognitive skills, the outcome of the Age 9 Test. These outcomes are derived from the answers to the 2001 Follow-Up survey. The different outcomes are: the probability that the participant recalls being happy during primary school (column (1)); the probability that the participant recalls having more friends than the other children during primary school (column (2)); the probability that the participant recalls having less friends than the other children during primary school (column (3)); the probability that the participant recalls being bullied during primary school (column (4)); the probability that the participant reports having attended a grammar school (column (5)); the probability that the participant reports having achieved O-level education (column (6)); the probability that the participant reports having achieved A-level education (column (7)); the probability that the participant reports having achieved a degree (column (8)). Our sample consists of 10 cohorts of children who attended primary school in Aberdeen in December 1962, who were born between October 1950 and October 1955, and participated in the 2001 mail Follow-Up to the original survey. We control for: a categorical variable taking a different value for each school in the sample; a categorical variable taking a different value for each cohort in the sample; a quadratic polynomial of child cognitive skills (based on the outcome of the Age 9 Test); the mean and standard deviation of the cognitive skills of the peers of the students (based on the outcome of the Age 9 Test); a categorical variable taking value 1 if the child is a girl, and 0 if he is a boy; a categorical variable taking value 1 if the child belongs to a family of high socioeconomic status (defined based on the father's occupation), and 0 otherwise; a categorical variable capturing the specific month of birth of the child; and the number of siblings of the child. Standard errors are clustered at the school-cohort level.

Table 2.8: Rank Effect on Long-Term Outcomes: Socioeconomic Status, Earnings, Fertility, and Well-Being

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Annual Income Above:			Probability of Having:		Well-Being:	
Outcome Variables	SES	£25,000	£35,000	£45,000	Married	Children	Enjoy Day	Happy
Percentile Rank	0.048	0.118*	0.034	0.028	0.018	-0.026	0.052	0.071
	(0.052)	(0.065)	(0.062)	(0.047)	(0.043)	(0.554)	(0.414)	(0.050)
Mean of the Outcome	0.17	0.43	0.24	0.11	0.92	0.86	0.89	0.88
SD of the Outcome	0.37	0.50	0.43	0.32	0.27	0.35	0.31	0.33
Observations	6,631	6,516	6,516	6,631	6,516	6,516	6,516	6,516
School Fixed Effects	X	X	X	X	X	X	X	X
Cohort Fixed Effects	X	X	X	X	X	X	X	X
Cognitive Skills	X	X	X	X	X	X	X	X
Cognitive Skills Squared	X	X	X	X	X	X	X	X
Mean of Peer Cognitive Skills	X	X	X	X	X	X	X	X
SD of Peer Cognitive Skills	X	X	X	X	X	X	X	X
Sex	X	X	X	X	X	X	X	X
Socioeconomic Status	X	X	X	X	X	X	X	X
Number of Siblings	X	X	X	X	X	X	X	X
Month of Birth	X	X	X	X	X	X	X	X

Notes: We estimate the relationship percentile rank within the school-cohort group and different outcomes. Percentile rank is established using our baseline measure of cognitive skills, the outcome of the Age 9 Test. These outcomes are derived from the answers to the 2001 Follow-Up survey, to which roughly 60% of the participants in the original survey responded. The different outcomes are: the probability of being in a high socioeconomic status, which is defined based on the occupation code (column (1)); the probability that the participant reports an annual income above £25,000 (column (2)), above £35,000 (column (3)), or above £45,000 (column (4)); the probability of the participant reports having ever married (column (5)); the probability of the participant reports having ever had children (column (6)); the probability of the participant reports enjoying daily activities (column (7)); the probability of the participant reports being happy (column (8)). Our sample consists of 10 cohorts of children who attended primary school in Aberdeen in December 1962, who were born between October 1950 and October 1955, and participated in the 2001 mail Follow-Up to the original survey. We control for: a categorical variable taking a different value for each school in the sample; a categorical variable taking a different value for each cohort in the sample; a quadratic polynomial of child cognitive skills (based on the outcome of the Age 9 Test); the mean and standard deviation of the cognitive skills of the peers of the students (based on the outcome of the Age 9 Test); a categorical variable taking value 1 if the child is a girl, and 0 if he is a boy; a categorical variable taking value 1 if the child belongs to a family of high socioeconomic status (defined based on the father's occupation), and 0 otherwise; a categorical variable capturing the specific month of birth of the child; and the number of siblings of the child. Standard errors are clustered at the school-cohort level.

Table 2.9: Gender Heterogeneity

Outcome Variables	(1) VRT	(2) Non-Cognitive Skills: Externalizing	(3) Non-Cognitive Skills: Internalizing	(4) Awareness of Teacher's Name	(5) Academic Achievement: Grammar	(6) Academic Achievement: O-Level	(7) Academic Achievement: A-Level	(8) Annual Income: > £25,000
				Rank Effect for Boys				
Percentile Rank	0.497***	0.421**	0.288*	0.166	0.250***	0.343***	0.405***	0.183**
	(0.072)	(0.167)	(0.162)	(0.123)	(0.056)	(0.066)	(0.062)	(0.071)
Mean of the Outcome	-0.04	-0.17	0.05	0.65	0.18	0.34	0.21	0.63
SD of the Outcome	1.01	1.16	0.96	0.48	0.38	0.47	0.41	0.48
Observations	4,902	3,349	3,349	925	2,844	2,844	2,844	2,844
				Rank Effect for Girls				
Percentile Rank	0.645***	0.119	0.603***	0.220*	0.492***	0.232***	0.413***	0.074
	(0.069)	(0.152)	(0.160)	(0.119)	(0.056)	(0.068)	(0.058)	(0.065)
Mean of the Outcome	0.04	0.18	-0.05	0.71	0.22	0.37	0.18	0.25
SD of the Outcome	0.98	0.75	1.03	0.46	0.42	0.48	0.38	0.43
Observations	4,539	3,167	3,167	900	3,031	3,031	3,031	3,031
T-test of the Difference	-0.148***	0.303***	-0.315***	-0.054	-0.242***	0.111***	-0.008	0.109**
	(0.035)	(0.083)	(0.091)	(0.063)	(0.029)	(0.038)	(0.034)	(0.043)
School Fixed Effects	X	X	X	X	X	X	X	X
Cohort Fixed Effects	X	X	X	X	X	X	X	X
Cognitive Skills	X	X	X	X	X	X	X	X
Cognitive Skills Squared	X	X	X	X	X	X	X	X
Mean of Peer Cognitive Skills	X	X	X	X	X	X	X	X
SD of Peer Cognitive Skills	X	X	X	X	X	X	X	X
Sex	X	X	X	X	X	X	X	X
Socioeconomic Status	X	X	X	X	X	X	X	X
Number of Siblings	X	X	X	X	X	X	X	X
Month of Birth	X	X	X	X	X	X	X	X
Height	-	X	X	-	-	-	-	-
Height Squared	-	X	X	-	-	-	-	-
Birth Weight	-	X	X	-	-	-	-	-
Birth Weight Squared	-	X	X	-	-	-	-	-

Notes: We estimate the relationship percentile rank within the school-cohort group and different outcomes, providing the linear combination of the effect of rank based and of the interaction between the rank and the children's gender. Percentile rank is established using our baseline measure of cognitive skills, the outcome of the Age 9 Test. The sample size changes depending on the outcome estimated, as they are derived from different surveys. The outcomes are: the standardized score of the Verbal Reasoning Test, or "VRT" (column (1)); the standardized measures of externalizing and internalizing skills (columns (2) and (3)); the probability that the parents know the teacher's name (column (4)); the probability that the participant reports having attended a grammar school (column (5)); the probability that the participant reports having achieved O-level (columns (6)) or A-level education (column (7)); the probability that the participant reports an annual income above £25,000 (column (8)). We control for: a categorical variable taking a different value for each school in the sample; a categorical variable taking a different value for each cohort in the sample; a quadratic polynomial of child cognitive skills (based on the outcome of the Age 9 Test); the mean and standard deviation of the cognitive skills of the peers of the students (based on the outcome of the Age 9 Test); a quadratic polynomial of the standardized height measured during the first medical exam in school and a quadratic polynomial of birth weight; a categorical variable taking value 1 if the child is a girl, and 0 if he is a boy; we control for a categorical variable taking value 1 if the child belongs to a family of high socioeconomic status (defined based on the father's occupation), and 0 otherwise; we control for a categorical variable capturing the specific month of birth of the child; finally, we control for the number of siblings of the child. Standard errors are clustered at the school-cohort level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2.10: Gender Heterogeneity: Removing Extreme Children with Extreme Behavioral Issues

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome Variables	Externalizing Skills			Internalizing Skills		
No Bottom	0%	3%	5%	0%	3%	5%
	Rank Effect for Boys					
Percentile Rank	0.421**	0.160	0.105	0.288*	0.234*	0.165*
	(0.167)	(0.110)	(0.099)	(0.162)	(0.119)	(0.098)
Mean of the Outcome	-0.17	0.03	0.10	0.05	0.15	0.20
SD of the Outcome	1.16	0.75	0.62	0.96	0.75	0.66
Observations	3,349	3,275	3,183	3,349	3,329	3,261
	Rank Effect for Girls					
Percentile Rank	0.119	0.021	0.006	0.603***	0.423***	0.333***
	(0.152)	(0.113)	(0.098)	(0.160)	(0.115)	(0.097)
Mean of the Outcome	0.18	0.24	0.27	-0.05	0.08	0.13
SD of the Outcome	0.75	0.56	0.49	1.03	0.77	0.69
Observations	3,167	3,187	3,151	3,167	3,125	3,057
School Fixed Effects	X	X	X	X	X	X
Cohort Fixed Effects	X	X	X	X	X	X
Cognitive Skills	X	X	X	X	X	X
Cognitive Skills Squared	X	X	X	X	X	X
Mean of Peer Cognitive Skills	X	X	X	X	X	X
SD of Peer Cognitive Skills	X	X	X	X	X	X
Sex	X	X	X	X	X	X
Socioeconomic Status	X	X	X	X	X	X
Number of Siblings	X	X	X	X	X	X
Month of Birth	X	X	X	X	X	X
Height	X	X	X	X	X	X
Height Squared	X	X	X	X	X	X
Birth Weight	X	X	X	X	X	X
Birth Weight Squared	X	X	X	X	X	X

Notes: We estimate the relationship percentile rank within the school-cohort group and the standardized measures of externalizing and internalizing skills, providing the linear combination of the effect of rank based and of the interaction between the rank and the children's gender, and progressively excluding the children with the most severe externalizing or internalizing issues. We start from the full sample (columns (1) and (4)), then exclude the bottom 3% (columns (2) and (5)), and finally the bottom 5% (columns (3) and (6)). We control for: a categorical variable taking a different value for each school in the sample; a categorical variable taking a different value for each cohort in the sample; a quadratic polynomial of child cognitive skills (based on the outcome of the Age 9 Test); the mean and standard deviation of the cognitive skills of the peers of the students (based on the outcome of the Age 9 Test); a quadratic polynomial of the standardized height measured during the first medical exam in school and a quadratic polynomial of birth weight; a categorical variable taking value 1 if the child is a girl, and 0 if he is a boy; we control for a categorical variable taking value 1 if the child belongs to a family of high socioeconomic status (defined based on the father's occupation), and 0 otherwise; we control for a categorical variable capturing the specific month of birth of the child; finally, we control for the number of siblings of the child. Standard errors are clustered at the school-cohort level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2.11: Robustness Exercise: Using School-Cohort Fixed Effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Non-Cognitive Skills:		Awareness of	Academic Achievement:			Annual Income:
Outcome Variables	VRT	Externalizing	Internalizing	Teacher's Name	Grammar	O-Level	A-Level	> £25,000
Percentile Rank	0.612***	0.287*	0.525***	0.228*	0.432***	0.322***	0.417***	0.092
	(0.068)	(0.158)	(0.158)	(0.132)	(0.055)	(0.069)	(0.060)	(0.068)
Mean of the Outcome	0	0	0	0.68	0.20	0.60	0.33	0.43
SD of the Outcome	1	1	1	0.47	0.40	0.49	0.47	0.50
Observations	9,441	6,516	6,516	1,787	5,744	5,744	5,744	5,744
School-Cohort Fixed Effects	X	X	X	X	X	X	X	X
Cognitive Skills	X	X	X	X	X	X	X	X
Cognitive Skills Squared	X	X	X	X	X	X	X	X
Mean of Peer Cognitive Skills	X	X	X	X	X	X	X	X
SD of Peer Cognitive Skills	X	X	X	X	X	X	X	X
Sex	X	X	X	X	X	X	X	X
Socioeconomic Status	X	X	X	X	X	X	X	X
Number of Siblings	X	X	X	X	X	X	X	X
Month of Birth	X	X	X	X	X	X	X	X
Height	-	X	X	-	-	-	-	-
Height Squared	-	X	X	-	-	-	-	-
Birth Weight	-	X	X	-	-	-	-	-
Birth Weight Squared	-	X	X	-	-	-	-	-

Notes: We estimate the relationship percentile rank within the school-cohort group and different outcomes. Percentile rank is established using our baseline measure of cognitive skills, the outcome of the Age 9 Test. The sample size changes depending on the outcome estimated, as they are derived from different surveys. The outcomes are: the standardized score of the Verbal Reasoning Test, or "VRT" (column (1)); the standardized measures of externalizing and internalizing skills (columns (2) and (3)); the probability that the parents know the teacher's name (column (4)); the probability that the participant reports having attended a grammar school (column (5)); the probability that the participant reports having achieved O-level (columns (6)) or A-level education (column (7)); the probability that the participant reports an annual income above £25,000 (column (8)). We control for: a categorical variable taking a different value for each school-cohort group in the sample; a quadratic polynomial of child cognitive skills (based on the outcome of the Age 9 Test); the mean and standard deviation of the cognitive skills of the peers of the students (based on the outcome of the Age 9 Test); a quadratic polynomial of the standardized height measured during the first medical exam in school and a quadratic polynomial of birth weight; a categorical variable taking value 1 if the child is a girl, and 0 if he is a boy; a categorical variable taking value 1 if the child belongs to a family of high socioeconomic status (defined based on the father's occupation), and 0 otherwise; a categorical variable capturing the specific month of birth of the child; and the number of siblings of the child. Standard errors are clustered at the school-cohort level.

Chapter 3

Breaking Barriers or Reinforcing Gaps? Scientific Education and Gendered Academic Choices

3.1 Introduction

The rising importance of sectors of the economy relying on science, technology, engineering, and mathematics (STEM) brought many advanced countries to implement policies to increase the pool of workers with adequate skills in those fields. It is generally recognized that a shortage of STEM skills exists, particularly in advanced economies (Directorate General for Employment, 2023; Nicholson, 2020; Salesforce, 2020), and the reasons are not fully understood.[1] But increasing the number of STEM graduates goes beyond merely meeting rising workforce demands; it also drives long-term economic growth and fosters

[1] Evidence of STEM graduate shortages is well-documents across advanced economies. The demand for computer and information science professionals is especially high. The U.S. Bureau of Labor Statistics predicts STEM occupations will grow faster than overall employment, especially in computer-related fields, where growth is anticipated to outpace the average by over threefold (Nicholson, 2020). In Europe, by 2020, there were an estimated 756,000 unfilled roles in the ICT sector alone, underlining the magnitude of the demand for these specialized skills (Salesforce, 2020). The Employment and Social Developments in Europe (ESDE) 2023 report by the European Commission (Directorate General for Employment, 2023) highlights persistent STEM worker shortages in the EU, emphasizes that this gap is expected to grow with demographic shifts and as the green and digital transitions advance. It focuses on how the education and training systems struggle to meet the demand for more specialized skills. It points to the lack of STEM graduates as the primary cause of these shortages.

innovation (Iaria et al., 2018; Waldinger, 2016; Moser et al., 2014). By boosting the pool of STEM talent, nations not only fill current skill gaps but also create a foundation for sustained advancement in technology and productivity.

I study whether additional exposure to and training in scientific disciplines during secondary school (specifically, mathematics and physics) affects the probability of applying for a STEM degree. High schools are the stepping stones for a student's higher education choices (Deming et al., 2014; Altonji et al., 2012). In the Italian context, that is the first time students can choose a school curriculum based on the subjects they want to focus on. Moreover, there is a high degree of unexplained and unobservable heterogeneity in how secondary school can develop useful skills to pursue a career in STEM (Ellison and Swanson, 2016).

I exploit a program available to Italian high school students enrolled in *Liceo Scientifico*, the academic track focused on scientific disciplines. The program was known as *Piano Nazionale Informatico*, which translates to National Plan for Informatics (henceforth *PNI*). It involved extra weekly hours of mathematics and physics. At the end of the full course of study, *PNI* students would have attended roughly 40% more hours of mathematics and 75% more hours of physics. The syllabus of both subjects would also be more advanced. In the time window I focus on, 50% of *Liceo Scientifico* in the country offered the *PNI* program.

I use data from the *Anagrafe Nazionale Studenti Universitari*, covering students who completed high school and enrolled in university between the academic years 2010/2011 and 2014/2015. The data allow me to identify the information provided by the student at the time of university enrollment and track their career throughout higher education. I can determine where they went to high school, what they studied, their score in the final high school exam, and their municipality of residence. I observe which university they attended, their degree course, and their achievements until graduation/dropout. Thanks to additional data provided by the Ministry of Education, I can distinguish scientific high schools offering the *PNI* program from the others.

Estimating the causal effect of attending a school offering the *PNI* program presents two main challenges. First, schools that implement the *PNI* track may differ systematically from those that do not, potentially biasing comparisons. Second, students are not randomly assigned to schools, raising concerns about self-selection. The first issue is mitigated by institutional features of the Italian education system that make program adoption unlikely to reflect school quality or a deliberate attempt to attract a particular student profile. The second is addressed through an instrumental variable approach that leverages geographic variation in program availability within commuting distance from students' place of residence.

If schools offering the *PNI* program systematically differ from traditional-only schools — particularly in unobservable ways such as teacher motivation or ability — estimates may conflate the program's effects with broader school-level differences. I discuss how this potential issue is mitigated by the fact that teacher assignment in Italy follows a centralized process based on national rankings, in which placement is determined by seniority and performance on a standardized national exam (Testo Unico Istruzione, 1994). Appointments are made strictly in order of these rankings, minimizing scope for strategic hiring. As a result, the composition of teaching staff is largely shaped by institutional rules rather than school-specific preferences or quality. This quasi-random allocation is especially relevant given that the decision to adopt the *PNI* program was taken internally by the school principal and teaching staff. Since schools do not choose their teachers, it is unlikely that the decision to implement the program reflects unobserved academic advantages.

Descriptive evidence supports the idea that offering the *PNI* program reflects structural conditions rather than unobserved school quality. Because schools could choose to adopt the program in a portion of their classes, implementation was more feasible in larger institutions — likely because it allowed them to introduce *PNI* without requiring the teaching staff to fully abandon the traditional curriculum. Consequently, schools offering the program tend to be larger on average. They are also more commonly located in urban—but not necessarily metropolitan—areas, where a single high school often serves the entire local student population. In contrast, traditional-only schools are more often found in metropolitan settings, where educational supply is fragmented, or in rural areas, where low student numbers make it difficult to support multiple programs. Importantly, there is meaningful overlap in school size across the two groups, with many traditional-only schools falling within the size range of those offering *PNI*.

To address concerns of student self-selection into schools offering the *PNI* program — particularly the possibility that more motivated or higher-ability students might disproportionately enroll in *PNI* schools — I adopt an instrumental variable approach that leverages geographical variation in access to schools offering the program. I construct the instrument as the share of scientific high schools offering *PNI* among those accessible from a student's municipality of residence, where accessibility is defined as requiring a commute of less than one hour. This design reflects the highly localized nature of school choice in Italy: students rarely move away from home to attend high school, and commute times are generally short (11 minutes on average, 10 at the median). Importantly, the availability of the program does not follow Italy's well-known North-South divide, further supporting the view that access is shaped by local school structure rather than broader regional disparities.

The instrument shows a strong first-stage relationship with treatment and varies substantially across municipalities. To assess whether it meets the key condition of independence — that is, whether it is unrelated to unobserved determinants of student outcomes aside from program access — I test its association with a broad set of local characteristics. Conditional on province fixed effects and population size, the instrument is not statistically significantly correlated with labor force participation, employment rates, migrant share, or prevailing gender norms of the students' municipalities of residence. Where statistical significance occurs, the estimated effects are economically negligible, suggesting that the instrument does not proxy for other contextual factors that confound estimates of the program's effect.

One potential concern is that the availability of the *PNI* program may have influenced students' decision to enroll in the scientific track (*Liceo Scientifico*) in the first place. Since my analysis is restricted to students who attended a *Liceo Scientifico*, any effect of the instrument on track choice would introduce selection into the sample, thus violating the exclusion restriction. To address this, I test whether the share of *PNI* schools accessible from a student's municipality is associated with enrollment in the *Liceo Scientifico*. I find no evidence of such a relationship, supporting the assumption that program availability does not distort the sample's composition.

A second concern is that students attending a *PNI* school may have a higher propensity to pursue higher education, potentially introducing selection bias, as my sample is limited to those who make this transition. To assess whether this issue might compromise the validity of my estimates, I test for differences in high school performance — specifically, final diploma marks and the probability of repeating a year — using the instrumental variable approach. The results, for both the overall sample and when disaggregated by gender, indicate no significant differences between treated and untreated students, supporting the assumption that program availability does not distort the sample's composition.

Overall, results for the full sample point to a small and statistically insignificant effect of the *PNI* program on the probability of enrolling in a STEM degree. I also find no evidence of an impact on university performance, measured by the number of credits earned during the first year. However, attending a school offering the program causes a statistically significant reduction in the probability of dropping out of higher education — about 4% relative to the baseline, significant at the 10% level. These aggregate findings conceal important gender differences. For boys, the likelihood of enrolling in a STEM major increases by roughly 8% relative to the baseline and is statistically significant at the 10% level. No significant effects emerge for boys' university performance or dropout. By contrast, I find no significant impact on any of the outcomes considered among girls. Importantly, these estimates capture intention-to-treat effects, reflecting exposure to schools offering the program rather than actual enrollment in the *PNI* program — making the observed magnitudes particularly noteworthy.

The impact of STEM-promoting educational policies often varies by gender, with recent studies suggesting that cultural and socioeconomic context plays a key role in shaping this heterogeneity. In more traditional settings, where social norms may discourage female participation in technical fields, girls often appear less responsive to such interventions — potentially due to differences in preferences or perceived barriers (De Philippis, 2023; Wiswall and Zafar, 2017; Reuben et al., 2017; Zafar, 2013). By contrast, evidence from more egalitarian societies indicates that removing these cultural barriers can substantially increase the effectiveness of STEM-promoting policies for girls (Joensen and Nielsen, 2016). Given Italy's marked socioeconomic and cultural diversity, it provides an ideal context for testing whether increasing exposure to scientific disciplines interacts with existing social norms to either widen or narrow gender disparities in educational choices.

To study how local gender norms shape the impact of the policy, I construct a measure based on voting outcomes from the 1981 abortion referendum in Italy. The referendum proposed a substantial rollback of the abortion rights established in 1978, and while it was ultimately rejected nationally, the support for the conservative reform varied widely across municipalities. I argue it provides a historically grounded and geographically specific measure of cultural attitudes. Drawing on a well-established literature showing that gender norms are highly persistent over time (Alesina et al., 2013; Fernández, 2007), this approach captures the normative environment in which students were socialized. Recent studies have adopted a similar strategy in Switzerland, showing that historical referendum outcomes continue to predict gendered behaviors in occupation choices and entrepreneurship decades later (Kaiser and Mata, 2025; Arni et al., 2024). Consistent with these findings, I show that support for the abortion restrictions in 1981 strongly predicts women-to-men labor force participation ratio and women's involvement in local politics 30 years later. However, compared to labor force-based indicators measured at the time of the intervention — which may conflate cultural attitudes with structural labor market conditions or reflect post-schooling migration — the referendum-based measure offers a more stable and locally anchored proxy for prevailing gender norms.

I divide municipalities into quartiles based on the gender norms index and estimate the treatment effect separately for girls across these groups. The results reveal a striking pattern: girls from the most conservative municipalities experience a significant increase in STEM enrollment when attending a school offering the *PNI* program. The estimated effect represents roughly a 30% increase relative to the baseline — comparable in magnitude to the most successful interventions in the literature. For instance, De Philippis (2023) finds an effect of 7% of the baseline in the general population, but much larger effects among top students, whose probability of enrolling in medicine or engineering increases by 27% and 17%, respectively. Similarly, large effects are found in experiments exposing girls to successful and charismatic role models (Breda et al., 2023; Porter and Serra, 2020). I show that these coefficients are not driven by small or imbalanced comparison groups, nor are they explained by geographic concentration alone. Rather, they appear to reflect meaningful variation in how students respond to the program in different normative contexts. In areas with more traditional gender norms, both boys and girls who pursue university studies tend to be positively selected, as evidenced by the greater number of credits earned in their first year of university. Among boys, the effect of the program is relatively stable across the gender norm distribution, indicating that their academic trajectories are less sensitive to local cultural context. For girls from these more conservative municipalities, however, exposure to enhanced scientific education leads to a substantial increase in STEM enrollment. This suggests that academic reinforcement can be particularly impactful in overcoming cultural barriers, providing a critical push for capable students in environments where traditional expectations might otherwise constrain their choices.

This paper contributes to the growing literature on university field-of-study choice (see Altonji et al. (2016) for a review) by examining how curricular exposure during secondary school — specifically through the enhanced scientific training of the *PNI* program — influences students' decisions to pursue STEM degrees. By focusing on students' major choices rather than general attainment, this work complements existing studies on the long-term effects of high school curricula on educational outcomes (e.g., Altonji et al. (2012); Levine and Zimmerman (1995); Altonji (1995)).

Several recent studies have exploited quasi-natural experiments to evaluate how additional exposure to mathematics and science shapes students' academic and career trajectories. Goodman (2019) examines changes to minimum high school math requirements in the United States, finding modest increases in STEM enrollment. Cortes et al. (2015) study a policy that doubled algebra instruction for low-performing ninth-grade students in Chicago, while Görlitz and Gravert (2018) focus on a German reform increasing instruction time in core subjects. In Denmark, Joensen and Nielsen (2016, 2009) analyze curriculum changes that reduced the cost of selecting advanced math and find long-run returns in both education and earnings. De Philippis (2023) investigates a UK policy reform targeting high-ability students, showing strong impacts on boys' STEM enrollment. All these studies demonstrate the benefits of this additional preparation, at the very least on test scores. My study differs in several important respects. First, I examine a large-scale national policy — Italy's *PNI* program — that, unlike in Görlitz and Gravert, 2018, does not involve a deep restructuring of the schooling system or the teaching programs. Moreover, instead of expanding the available subjects (De Philippis, 2023), the *PNI* program studied here provided a clean change in overall instruction time.

Most importantly, I provide evidence on the interaction between prevailing gender norms and additional exposure to scientific teaching. While previous studies document gender heterogeneity in the effects of increased mathematics instruction — often finding, as I do, that such interventions widen the STEM gender gap (De Philippis, 2023; Morando, 2020) — a more nuanced analysis suggests that this pattern is not universal and may depend critically on the surrounding cultural context. Using a novel proxy for gender attitudes, I show that the *PNI* program can counteract the effect of traditional gender roles. Leveraging a historically grounded proxy for gender attitudes — one that is plausibly insulated from contemporaneous economic conditions — I show that the *PNI* program can mitigate the influence of traditional gender norms on educational choices, possibly with long-lasting consequences. I also contribute to a deeper understanding of when and where STEM interventions are most effective, and for whom.

3.2 Institutional Setting

The institutional framework provides important information for the identification strategy of this study. The first key point concerns the nature of the program itself, which did not offer any formal academic or certification advantage. As such, it was not a prerequisite for pursuing STEM studies in higher education. The second important aspect is the process through which schools adopted the program — a decision shaped by the teaching staff, whose assignment is quasi-random and largely independent of school quality. This institutional structure ensures that variation in access to the program is not systematically driven by school-specific characteristics, supporting the credibility of the identification strategy.

3.2.1 The Italian Secondary School System and Access to Higher Education

At age 14, after completing a common national curriculum through primary and lower secondary school, Italian students face their first major academic choice: they must decide what type of high school to attend, effectively selecting the track and subjects they will focus on for the remainder of their secondary education. Figure 3.1 shows students can choose three different educational paths: academic, technical, or professional. Completing the full cycle of education (five years) in any of these schools formally grants access to higher education.

I focus on the students of a specific academic track program, the *Liceo Scientifico*. This type of school emphasizes scientific subjects such as mathematics, physics, and chemistry while maintaining its humanities-focused core.[2] It is the most selected by those who want to pursue higher education, accounting for 60% of all academic track students (Associ-

[2]Other academic track curricula are: *Liceo Classico* focusing on humanities with subjects like Latin, Greek, and philosophy; *Liceo Linguistico* specializing in foreign languages; *Liceo Pedagogico* focusing on social sciences and psychology. Alternatively, students might attend a technical school (*Istituto Tecnico*) or a vocational school (*Istituto Professionale*). The former provides specialized training in areas such as business, technology, and engineering. The latter teaches skills geared toward immediate employment in fields like healthcare, hospitality, or manufacturing, emphasizing hands-on learning and direct skills training.

azione Almadiploma, 2013). One reason is undoubtedly the versatility of the curriculum, which leaves students' future options open to any higher education path.

Until the 2009/2010 academic year, students enrolling in the Liceo Scientifico could opt for an alternative track to the standard curriculum — the *PNI* program — which offered additional instruction in mathematics and physics. The program remained in place for cohorts graduating up to 2013/2014 and offered students additional training in mathematics and physics, strengthening the scientific focus of the curriculum. It was widely available across the country, with approximately half of all the schools offering the program, making it a significant feature of Italy's secondary education landscape during that period.

3.2.2 History and Structure of the *Piano Nazionale Informatica*

The *Piano Nazionale Informatica* (the National Plan for Informatics), or *PNI*, was introduced in the 60s and 70s as part of a broader educational modernization effort. Recognizing the growing importance of computer science, it aimed to teach some of its basics in schools. Throughout the '80s and '90s, *PNI* saw revisions with two key features persisting: teacher training in using computer science for teaching and the intention to deploy an experimental program, offered as an alternative to the traditional track. Schools began adopting the program in 1993. It increased the focus on mathematics and incorporated some programming elements.[3] A review of the history up until the end of the 20th century was published by Unione Matematica Italiana (2003).

As Table 3.1 shows, the *PNI* program involved more hours of mathematics (highlighted in red) and physics (highlighted in green) compared to the other available. At the end of the five years, a *PNI* student would have attended 875 hours of mathematics and 490 hours of physics, versus the 630 and 280 hours of a student attending the traditional path. That amounts to 40% more hours of mathematics and 75% more hours of physics.

[3] Overall computer science integration remained limited: it fundamentally involved more weekly hours of mathematics and physics.

The programs also covered a wider range of topics. A key feature of the program is that the additional hours were not replacing other subjects, but were added to the traditional schedule.

The "Gelmini Reform" cancelled the *PNI* in 2010. The main rationale of the reform was to standardize the curricula at the national level. The reform's goal of streamlining high school curricula resulted in a new *Liceo Scientifico* syllabus, closely resembling the one of the *PNI* program.

3.2.3 Schools' Adoption of the *PNI* Program

Teachers and principals influenced a school's decision to offer the *PNI* program. According to professors and school principals I interviewed, who were working in *PNI* schools at the time, offering the program depended on a joint decision of the principal and the assembly of the teachers working at the schools (*Collegio dei Docenti*), following a proposal from one of its members.[4] Schools had the flexibility to implement *PNI* across all classes or only in a subset (depending on the number of students choosing that), making it more feasible for larger schools. This structural advantage naturally increased the likelihood of *PNI* adoption in larger schools, where offering it in just a portion of classes was a more practical option.[5]

3.2.4 On Teachers' Hiring

The hiring process for high school teachers in Italian public schools is heavily regulated. It follows a rigid, centralized structure that leads to a quasi-random allocation of teachers across the country. The system is primarily based on competitive examinations (*concorsi*) and national rankings (Eurydice - European Commission, 2024).

Secondary school teachers must go through academic qualifying programs to acquire specific competencies in addition to a master's degree in one of the subjects taught at the

[4] I contacted several principals (current and former) of different Italian schools offering the *PNI* program. I had phone calls with three of them, who provided all the same explanation on the procedural steps required to adopt the program.

[5] See discussion in subsection 3.4.2.

secondary level. A sizeable portion of these programs involve direct and indirect traineeship activities. They end with a final exam testing their writing and oral skills. Passing the final test grants the teaching qualification, allowing access to the national ranking from which schools recruit teachers. Passing the examination allows one to be hired with a permanent contract and start the one-year induction period in their assigned school. The final job placement is hence dictated by the availability of positions rather than personal choice, leading teachers to schools far from their preferred locations. Moreover, seniority is crucial in determining the teachers' position in the ranking (Articles 399 and 400 of the Testo Unico Istruzione (1994)).

Transfers are difficult and governed by a points-based system that prioritizes tenure and seniority. As a result, new teachers often spend years in schools that were not their first choice, further reinforcing the quasi-random nature of initial placements. Moreover, because competitive exams are irregular, and hiring is slow, many teachers remain in limbo for years, cycling through short-term contracts in multiple regions before securing a stable position (Education International, 2023).

3.3 Data and Sample Description

3.3.1 Data Sources

The main data source for student outcomes is individual-level administrative data provided by the Italian Ministry of Education, which includes detailed information on high school and university performance but lacks comprehensive background characteristics. To address this limitation, I supplement this dataset with municipality-level data from the 2011 Census, the 2011 Registry of Local Administrators, and the outcomes of the 1981 abortion referendum, which provide a richer description of local characteristics.

Administrative Student-Level Data. In the "Anagrafe Nazionale Studenti Universitari" the Ministry of Education of Italy records all students enrolled at Italian universities. It provides demographic information collected at the time of their enrollment into higher education and details on their chosen degree course, progression, and outcomes.

Importantly, demographic information is collected when students apply to university, generally during their final year of high school. That allows me to identify the students' municipality of residence at the time of high school and the high school they attended.

School-Level Data. Formally, no registry indicates whether the *PNI* program was available in a certain school. To identify which *Liceo Scientifico* offered the program, I exploit the communications between the Ministry of Education and the school. The mathematics test of the "Esame di Maturità" (a 4-part exam that needs to be passed to be awarded a high school diploma, with the first two portions set at the national level) was different for *PNI* students. High schools needed to request a copy of the test to the Ministry of Education so their students could take the exam. The government recorded which schools requested the test each academic year, allowing me to identify the schools that offered the *PNI* program. A limitation of the data is that they allow the identification of *PNI* schools, but not of *PNI* students. A school offering the *PNI* program could provide both programs (traditional and *PNI*), or the latter only. I collected additional information about the 585 *PNI* schools in my sample by contacting their administrative offices. Roughly 1/3 of them responded to this survey. Overall, around 15% of them provided only the *PNI* program, while the remainder had both. Because of that, I will take an intention-to-treat approach, assuming all schools offer both programs.

Municipality-Level Data. To describe the municipalities in my data I exploit information from the 2011 Italian Census provided by the Italian National Institute of Statistics (*ISTAT*). It contains information on the socioeconomic characteristics of the residents of each municipality. I also use data from the Ministry of the Interior's Registry of Local and Regional Administrators, which provides detailed information on elected officials at the local level. This source allows me to characterize local administrations — particularly, by the gender composition of their elected representatives.

Gender Stereotypes. To measure gender stereotypes at the municipality level, I use the support rate for the 1981 referendum, called to repeal a 1978 which legalized and regulated access to abortion. The vote was called by Catholic and right-wing organizations seeking to substantially restrict the right to abortion, effectively aiming to eliminate access except in cases of severe medical necessity. The proposal was rejected by a wide margin, with only 32% of voters supporting the repeal. Despite the national result, support for the conservative proposal varied widely across municipalities (from 4 to 80 percent of the voters). I will discuss using this variable as a proxy for local gender norms in subsection 3.6.2.

3.3.2 Defining STEM Degrees

Defining STEM Degrees. Similarly to Chise et al. (2021), I define STEM disciplines using the definition provided in Eurostat's Classification of Fields of Education and Training (1999). This relies on the 2013 revision of the International Standard Classification of Education (ISCED 1997), which focuses on fields of education and training, ISCED-F. The equivalent STEM field of study can be classified into three categories. (i) Natural sciences, mathematics, and statistics; (ii) Information and Communication Technologies; (iii) Engineering, manufacturing, and construction. Consistent with the European Union's framework, I do not include either architecture or health studies (such as medical or nursing schools).

3.3.3 Sample Description

I focus on the cohorts of students graduating from *Liceo Scientifico* who achieved their diplomas between 2009/2010 and 2013/2014. I observe only those who, following the end of high school, enter higher education. Finally, I remove all students from schools located in the provinces of *Aosta* and *Alto-Adige*.[6] Schools in *Alto-Adige* are granted special autonomy to teach in German and Italian, introducing differences in the programs offered. In the province *Aosta* the special autonomy of the local administration makes the data on high school students transitioning to higher education unreliable. Because of changes in the administrative boundaries established in 2016, I am unable to link students

[6] In Italy, the hierarchy of administrative units below the State consists of 3 layers. The 20 regions contain each from 1 to 12 provinces (the median is 5), totaling 106. Within each province are located the municipalities, for a total above 7900 in the entire country. Their number per province varies between 6 and 309 (the median is 95).

living in different provinces of *Sardegna* (*Carbonia-Iglesias*, *Medio Campidano*, *Ogliastra*, and *Olbia-Tempio*) to their residency information. I exclude from the sample students who did not enroll in university right after the end of high school and students who repeated one or more years (whether that happened in high school or previously).

3.4 Descriptive Statistics

3.4.1 Sample Description: Students

Over the decades, and especially after the liberalization of access to university in 1969, *Liceo Scientifico* became the predominant academic track choice among Italian students (La Repubblica, 2010). Among the reasons were its greater flexibility and the growing awareness of the increasing importance of scientific disciplines. I show a breakdown of the students enrolled in higher education by cohort and their type of diploma in Table 3.2. Roughly 42% of the students starting their degree over that period studied at the *Liceo Scientifico*.[7]

Table 3.3 shows that *Liceo Scientifico* curriculum is unique in that it maintains a balanced gender composition, whereas other academic tracks tend to have a significantly higher proportion of female students. As expected, graduates from this track are more likely to enroll in STEM programs compared to their peers. However, despite their strong orientation toward science and mathematics, they do not exhibit superior academic performance in high school — however, this metric is largely school-specific and difficult to compare across curricula.[8] Similarly, university performance during the first year does not

[7] I include only students who never failed a year and started their degree right after finishing high school.

[8] The Italian high school diploma exam (*Esame di Stato*) is a comprehensive final assessment taken at the end of upper secondary education. The exam structure combines both nationally standardized and school-specific components. Typically, it includes two written tests — centrally prepared by the Ministry of Education and aligned with the student's curriculum (e.g., scientific, classical), and another designed at the school level — followed by an oral exam. While the exam content partially varies depending on the high school track and school, the evaluation process is conducted locally, by a committee composed of internal and external teachers. Despite periodic reforms, the core structure has consistently aimed to assess a broad set of competencies and serves as a necessary qualification for university admission. The outcome ranges between 60 (pass) and 100 *cum laude*.

differ based on students' background.[9] Even in terms of dropout rates from university, *Liceo Scientifico* graduates do not fare better than others.

In Table 3.4 I focus on the differences between students who attended a school offering the *PNI* program and those graduating from a school offering the traditional program only. The difference in gender balance is small but statistically significant, with *PNI* schools enrolling a slightly higher share of girls. *PNI* students also tend to enroll more frequently in STEM degrees, to finish high school with better grades, and to do better during their first year of higher education. The most striking difference lies in the dropout rate, with students from the traditional program being almost 5 percentage points more likely to drop out of university. In general, *PNI* students seem to do better once they enroll in university.

3.4.2 Sample Description: Schools

Table 3.5 highlights two key features of the setting. As discussed in subsection 3.2.3, the rules governing *PNI* adoption made it more feasible for larger schools to implement the program in just a subset of classes, which helps explain why these schools tend to be larger on average. Mean cohort sizes in *PNI* schools are more than double those of traditional-only schools (97 vs. 43), with a similar contrast in medians (88 vs. 31). Yet, the variation in school size among traditional-only schools is substantial, with a standard deviation of nearly 40 — almost as large as the mean — compared to 56 in *PNI* schools. This dispersion is meaningful: 10% of traditional-only schools have a cohort size above the mean size of *PNI* schools, and 33% fall within one standard deviation of the *PNI* schools' mean. Conversely, almost 18% of *PNI* schools are smaller than the mean size of traditional-only schools, and 45% fall within one standard deviation of that mean. Figure 3.2 offers a visual representation of the overlap in school size across both groups.

[9] I measure first-year performance using the number of credits acquired during the first year of higher education. The European Credit Transfer System (ECTS) is a system used all over Europe and enables you to easily compare study programs and transfer your academic qualifications from one educational institution to another. Passing a university exam grants you ECTS corresponding to the expected workload required for that exam (1 ECTS roughly amounts to 25 to 30 hours). Degrees are structured such that each academic year students are required to earn 60 ECTS.

These patterns indicate that program adoption is not mechanically tied to school size. They also highlight that the identification strategy rests on the presence of substantial common support, avoiding the need for strong out-of-sample extrapolation.

The second key characteristic concerns the demographic context of *PNI* school locations. On average, *PNI* schools are situated in less populated and less densely populated municipalities than their traditional-only counterparts. Specifically, the mean population of municipalities with *PNI* schools is approximately 232,000, compared to 335,000 for those hosting only the traditional track. To further explore this pattern, I classify municipalities into rural, urban, and metropolitan categories based on population thresholds used in Italian administrative practice.[10] This classification reveals that traditional-only schools are disproportionately more represented in metropolitan municipalities, which have an average population of 1.2 million (median: 962,003), than in urban municipalities, which are much smaller on average (mean: 50,000; median: 28,563). The percentage of *PNI* schools located in rural municipalities (population mean: 1,939; median: 1,641) is nearly identical to that of traditional-only schools. The resulting pattern suggests that the *PNI* program was more commonly adopted in small-to-medium-sized urban areas, where one or a few schools typically serve the entire student population. In these settings, schools are more likely to be large enough to implement the program selectively across classes. By contrast, traditional-only schools are more frequently found in large metropolitan centers, where educational supply is fragmented across many smaller institutions, or in rural areas, where limited student numbers make offering multiple curricular tracks more difficult. These dynamics indicate that a municipality's demographic profile indirectly shapes program availability by influencing school size and the structure of local enrollment.

A defining feature of the Italian context is the longstanding North-South divide,

[10]Following a long-standing benchmark in national legislation (e.g., Law 311/2004) and recent academic contributions such as Cattivelli (2021), I define rural municipalities as those with fewer than 5,000 residents, and urban ones as those above that threshold. Metropolitan municipalities are instead identified according to national law establishing the 14 official metropolitan cities (*città metropolitane*), which include major urban centers in the country. Law 56/2014 formally established the metropolitan cities and includes the 14 largest urban areas in Italy: Bari, Bologna, Cagliari, Catania, Florence, Genoa, Messina, Milan, Naples, Palermo, Reggio Calabria, Rome, Turin, and Venice. This classification reflects not only population size but also administrative and institutional status.

which cuts across a wide range of economic and social dimensions. For example, Lombardy's GDP per capita reaches 127% of the EU average, while Calabria's stands at just 56%, reflecting persistent regional disparities in development and opportunity (Fernández-Villaverde et al., 2023; Giuntella, 2022). However, when studying the availability of the *PNI* program, this familiar territorial pattern does not appear to hold. I assess the regional presence of *PNI* schools and the share of students who attend them in Table 3.6.[11] The larger size of *PNI* schools is particularly evident in regions like *Trentino* and *Abruzzo*, where the share of schools offering *PNI* is well below the national average, yet the proportion of students enrolled in *PNI* programs is significantly higher. In contrast, Central Italy stands out for having a high share of *PNI* schools. However, the distribution of *PNI* availability appears relatively uniform across most regions. *Basilicata* emerges as an exception, being the only region where *PNI* and Traditional schools have a similar size profile. More broadly, apart from specific cases such as *Sardegna* and *Trentino*, there are no major regional gaps in access, suggesting that the *PNI* program was widely available across the country, suggesting that other factors — such as local school structure and demographic conditions — played a more influential role in shaping access to the program.

3.5 Empirical Strategy

To estimate the impact of differential exposure to scientific subjects for student i who attended school s on outcomes Y, I consider the following linear model:

$$Y_{is} = \beta_0 + \beta_1 T_{is} + \beta_2 G_{is} + \beta_3 S_s + \beta_4 P_{im} + \lambda_{cp} + \epsilon_{is} \tag{3.1}$$

[11] I employ the classification of the Italian National Institute of Statistics (ISTAT) to group the regions. ISTAT divides Italy into geographical macro-areas to facilitate regional analysis and comparisons. This division allows for a more nuanced understanding of Italy's economic, social, and demographic characteristics, recognizing the diversity between different parts of the country. The macro-areas—typically Northwest (*Val d'Aosta, Piemonte, Liguria, Lombardia*), Northeast (*Friuli Venezia Giulia, Trentino Alto-Adige, Veneto, Emilia Romagna*), Center (*Toscana, Umbria, Marche, Lazio*), and South and Islands (*Campania, Abruzzo, Molise, Basilicata, Puglia, Calabria, Sicilia, Sardegna*). The classification reflects historical, economic, and cultural differences that affect various aspects of life, such as income distribution, employment, education, and healthcare access.

The outcomes are the probability of enrolling in a STEM degree, performance during the first year in higher education (proxied by the number of ECTS attained), and probability of dropout. T_{is} represents a binary indicator of treatment status, i.e. a student attended a school offering the *PNI* program. The objective is to obtain an unbiased estimate of β_1.

A key concern in identifying the causal effect of the *PNI* program is the possibility that schools offering it differ systematically from those that do not in ways that are unobservable but correlate with student outcomes. For instance, if *PNI* schools had stronger leadership, more motivated teachers, or a generally better academic environment, failing to account for this could lead to a biased estimate of the program's true effect. However, the Italian institutional framework helps to address these identification challenges. As I explained in subsection 3.2.3 and subsection 3.2.4, teacher assignments are quasi-random, and the decision to implement the *PNI* program was based on a collective agreement between the principal and teaching staff, rather than a targeted policy for high-performing schools.

One remaining issue is that the *PNI* program is generally offered by larger schools, as they could implement it in only a subset of classes. Since school size may be associated with other factors — such as access to resources or broader peer networks — that influence students' outcomes, controlling for school size or incorporating it into the identification strategy is essential to isolate the program's effect. I do this by including in Equation 3.1 S_s, a categorical variable controlling for the tercile of the number of students enrolled. To capture time-varying differences across schools' geographical areas, I include λ_{pc}, a set of dummy variables accounting for the combination of the school's province and the students' cohorts. I interact the cohort indicator with the province, which is the second layer in Italy's administrative hierarchy above the municipality (below the regions).

Endogeneity at the student level poses an additional identification challenge, as students who enroll in the *PNI* program may differ systematically from those who do not — particularly in terms of unobserved motivation, ability, or family support, all of which may also influence their likelihood of pursuing a STEM degree. To account for observable heterogeneity, I include controls for gender G_{is} and for whether the municipality where the student resides is categorized as rural, urban, or metropolitan P_{im}. As I explained in subsection 3.4.2, this classification follows the Italian administrative practice. Rural municipalities are defined as those below 5,000 residents (Law 311/2004), while metropolitan areas are the major urban centers in the country, which are granted special institutional status (Law 56/2014). While the student-level dataset does not offer a wide set of individual-level characteristics, these variables help capture important differences in both student background and local schooling context. In particular, accounting for the municipality population is essential: *PNI* programs were more likely to be offered in urban but not necessarily metropolitan areas, where school size tends to be large enough to support partial program implementation. By contrast, large cities often host a fragmented network of smaller schools, many of which did not adopt *PNI*. Moreover, small towns may only have one school, which might or might not offer the program, regardless of size. Controlling for population size thus helps account for variability in school availability and structure, reducing the risk that local context drives both program exposure and academic outcomes.

Nonetheless, these controls alone cannot fully address students freely sorting to different schools. To isolate a plausibly exogenous variation in school choice, I instrument T_{is} with the share of available schools offering the *PNI* program to students living in a given municipality. The identification strategy is valid if (i) the instrument is as good as randomly assigned (instrumental independence), (ii) it affects the outcome only through its effect on the probability of choosing a *PNI* school (exclusion), (iii) it has a non-zero relationship with the treatment (relevance). I provide evidence for these assumptions in the next sections.

3.5.1 Instrumental Variable

The logic behind my instrument is intuitive. Geographical proximity is an important determinant of the education institution of choice in many contexts (Mandic et al., 2023; Laverde, 2022; Agarwal and Somaini, 2019). The same holds in Italy, where students rarely leave their parental home, especially before finishing secondary education. Boarding schools are rare, as only 8 exist in the country (as of 2024). In my sample, the median commute time for a student to reach their school is less than 9 minutes.[12] Figure 3.3 shows, for each *Liceo Scientifico* in the country, the share of students who commute for less than a certain time. Each school is represented by the share of students traveling less than - in the case of the blue short-dashed line - 15 minutes every day. The graph provides a key insight: in the vast majority of the schools, the share of students commuting for less than 30 minutes to attend is above 90%.

Establishing the importance of distance supports the construction of an instrument based on school availability. I use the ratio between the number of high schools that offered the *PNI* program and the total number of high schools (of the *Liceo Scientifico* type) available to the residents of a given municipality. I exploit a data-inspired approach to define the catchment area, observing all the possible combinations of the students' municipality of residence and their high school of choice. Those allow me to infer which schools are in the feasible catchment area. The underlying assumption is that all the schools "available" to the residents of a given municipality were attended by at least one student (which is why I use all the data of students graduating from a *Liceo Scientifico*). The school's location is defined using the street address, while the students' residence is defined using the centroid of the municipality of residence (reported at the time they enrolled at the university). I refine the instrument by removing from the set of available schools those for which the commute lasted longer than 60 minutes. I plot the instrument on the map of Italian municipalities in Figure 3.4.[13] The distribution of the instrument across municipalities does not reveal any immediately recognizable geographic pattern, suggesting that access to *PNI* schools is not systematically clustered in specific areas.

My instrument is therefore defined at the municipality level (with 7,042 observations, one for each municipality in the sample). On average, around 50% of the accessible schools offered the *PNI* program (the standard deviation is 0.34). I plot the share for each mu-

[12] I use the API services provided by Google Maps to calculate the commute time. I estimate the duration by assuming the students are traveling by car.

[13] As mentioned in subsection 3.3.3, the municipalities located in the provinces of *Aosta* and *Alto-Adige* are excluded, while those located in the provinces of *Carbonia-Iglesias*, *Medio Campidano*, *Ogliastra*, and *Olbia-Tempio* are not available.

nicipality in Figure 3.4. The distribution of the instrument is not uniform, showing peaks around 0, 0.5, and 1, meaning that municipalities cluster into three distinct groups: those with no access, partial access, and full access to the program. This suggests that in municipalities where no schools offer the program (instrument = 0), the probability of treatment is substantially lower, while in municipalities where all schools offer it (instrument = 1), treatment probability is much higher. For municipalities around the middle peak (instrument $\simeq$ 0.5), treatment assignment remains partial, meaning access to the program is available but not universal. Importantly, the instrument shows considerable variation. 60% of the municipalities are covered just partially. Even without considering those with no access to *PNI* schools, roughly 25% fare below the 0.5 threshold.

A strong relationship between the instrument and the endogenous variable is essential for a successful identification strategy, as it eliminates possible issues that can arise under weak instruments. To address the relevance of the instrument, I estimate the following first stage:

$$T_{is} = \gamma_0 + \gamma_1 Z_m + \gamma_2 G_{is} + \gamma_3 S_s + \gamma_4 P_{im} + \lambda_{cp} + \varepsilon_{is} \tag{3.2}$$

Where *PNI* choice T_{is} is predicted by the instrument Z_m, the availability of *PNI* schools for the students living in a given municipality.

3.5.2 Evaluating Instrumental Validity

As previously discussed, potential threats to identification in this setting arise from two key sources: systematic differences between schools that offer the *PNI* program and students' selection into the program. Understanding the mechanisms determining which schools implemented the *PNI* track is thus critical. In this regard, I showed how features of the Italian institutional framework, especially related to teachers' hiring, make staff composition quasi-random and reduce the possibility that *PNI* adoption reflected underlying school quality.

More broadly, the Italian education system is designed to promote equal access, as

Article 34 of the Italian Constitution states. Universal education and equal opportunity are established as fundamental rights (Italian Parliament, 1948). Following this principles, 94% of high schools are state-funded, and follow a standardized syllabus. Part of the final national exam also consists of a test prepared by the Ministry of Education, and that is common across all schools in the country. The location and endowment of schools are planned based on demographic needs, as determined through agreements between regional authorities, the Ministry of Education, and the Ministry of Finance (Article 53(1) of the Testo Unico Istruzione (1994)). This institutional design makes it unlikely that external factors, such as local labor market demand for STEM workers, systematically influenced the adoption of the *PNI* program.

Nevertheless, while the institutional framework of the Italian school system helps mitigate concerns of endogenous program placement, it is important to acknowledge that observable differences between *PNI* and non-*PNI* schools do exist. I argue that these differences are not arbitrary but stem directly from the process required to implement the *PNI* program (outlined in subsection 3.2.3), which was highly dependent on student enrollment numbers. In larger schools, it was more feasible to offer the program in just a subset of classes, making adoption more likely. By contrast, smaller schools, which are common both in metropolitan areas — where school size varies considerably — and in more rural communities, were often less equipped to accommodate the additional specialization. In medium-sized cities, a single school often served the entire area, resulting in larger student bodies and thus a higher probability of offering the program. For this reason, school size is an important confounder, and I explicitly account for it in my main specification to ensure that differences in access are not simply reflecting underlying variation in school scale.

Given that school size — and thus *PNI* availability — is strongly related to the local context, it is crucial to ensure that the instrumental variation in *PNI* access is not driven by underlying differences across municipalities. In particular, if the share of *PNI* schools within a municipality is systematically associated with observable characteristics such as labor market factors or prevailing gender norms, then the estimated effects of the program could be confounded by local conditions, rather than capturing the causal impact of program exposure. To address this concern, I test whether the instrument behaves as good as randomly assigned by examining its relationship with a rich set of municipality-level characteristics. While random assignment is not strictly necessary for identification, demonstrating that any residual correlations are economically negligible helps support the validity of the exclusion restriction and the robustness of the identification strategy. I estimate the following equation:

$$W_m = \alpha_0 + \alpha_1 Z_m + \alpha_2 P_m + \lambda_p + \zeta_m \tag{3.3}$$

Where W_m describes different socio-economic characteristics of the municipalities.[14] P_m is a categorical variable representing whether the municipality is classified as rural, urban, or metropolitan. λ_p accounts for time-invariant factors determined by a student's province of residence.[15]

In the two panels of Table 3.7, columns (1), (5), and (7), present estimates of the relationship between the instrument and key municipality-level characteristics, such as labor force participation (both overall and female-only) and the share of foreign-born residents, controlling for province fixed effects. While the coefficients are statistically significant, their economic relevance is minimal. A change in the instrument from 0% to 100% is associated with changes of less than one percentage point in these outcomes — corresponding to less than 10% of a standard deviation, or just above that in the case of female labor force participation. Similarly, column (15) shows that the relationship between the instrument and support for abortion restrictions corresponds to a change of roughly 13% of a standard deviation — again, modest in magnitude.

When I additionally control for whether a municipality is classified as rural, urban, or metropolitan, only the relationship between the instrument and support for abortion re-

[14]These variables are: labor force participation and employment, for both the overall population and females; the share of foreign-born residents; the share of members of the municipality executive committee who are women as of 2011; the support for restricting access to abortion during the 1981 abortion referendum, as a measure of gender norms at the municipality level.

[15]I do not include cohort-by-province fixed effects as in the main specification because municipality characteristics are measured only at one point in time (2011), and therefore do not vary by cohort.

strictions remains statistically significant. However, its magnitude becomes economically negligible. Even in the most extreme case — comparing two otherwise similar municipalities within the same province and classification (e.g., both urban), where one has no *PNI* schools and the other has full coverage — the implied change in support for abortion restrictions is just 7% of a standard deviation, or 2% of the mean. Furthermore, I find no significant association between the instrument and the share of women appointed by the mayor to the municipal executive council. This supports the interpretation that the instrument is not systematically related to gender norms, alleviating concerns about potential bias.

Having examined the factors driving which schools offered the *PNI* program, I now turn to the question of who enrolled in it — that is, whether student selection into *PNI* schools could introduce endogeneity in estimating the program's effects.

A necessary condition for the validity of the empirical design is that the more demanding scientific syllabus of the *PNI* program does not systematically change enrollment to *Liceo Scientifico*. If the availability of the program alters students' decision to enter the scientific curriculum — particularly by deterring those with lower academic confidence — then my estimation sample would be selectively composed, and the model would be misspecified, as it only includes students enrolled in the *Liceo Scientifico*. More critically, this would violate the exclusion restriction, as the instrument would affect the outcome not solely through treatment exposure (i.e., attending the *PNI*), but also by shaping curriculum selection itself. I do not expect that to be the case, as the treatment is defined at the school level, which assumes all students enrolling in a *PNI* school received the treatment. In reality, the vast majority (85%) of the schools offering the program allowed their students to choose between the two options, and constructed the classes accordingly. This suggests minimal displacement effects, as opting for the traditional program was possible in most cases. I also provide evidence of that by estimating the following equation:

$$W_{is} = \alpha_0 + \alpha_1 Z_m + \lambda_{cp} + \zeta_{is} \tag{3.4}$$

Where I regress the dummy variable W_{is} - indicating whether a student chooses *Liceo Scientifico* over a different academic-track school - on the instrument Z_i and cohort-by-province fixed effects λ_{cp}. This serves as a reduced-form check on whether the program's presence affected the composition of students selecting into the scientific curriculum.[16] Table 3.8 shows that the relationship is absent whether or not I include the province of residence, cohort, or province of residence-by-cohort fixed effects.

While I cannot observe students who do not enroll in university — limiting my ability to directly test whether *PNI* and non-*PNI* schools differ in terms of transition rates — I can assess potential selection by examining whether students who attend the *PNI* program differ in key academic outcomes during high school. These checks offer indirect evidence on whether systematic differences in ability or performance may bias the estimates. Using Equation 3.3, I estimate the relationship between high school-related outcomes and attending a school offering the *PNI* program. These are high school graduation mark and whether a student had to repeat at least one year.

I present the results in Table 3.9. The findings indicate that while OLS estimates suggest statistically significant differences — *PNI* attendance is associated with modestly higher high school grades and a substantially lower probability of repeating a year — these effects vanish under 2SLS estimation, becoming statistically insignificant across all specifications. This divergence between OLS and 2SLS results is likely attributable to selection into treatment. The first-stage relationship between the instrument and treatment remains strong, with consistently large coefficients and robust F-statistics. Overall, the evidence indicates that any systematic differences between *PNI* and non-*PNI* students are small and effectively addressed through the instrumental variable approach. This holds both in the full sample and in gender-specific estimates.

Moreover, as discussed earlier, differences in school quality are difficult to exploit strategically in the Italian context. The scope for meaningful differentiation in curricular content or staffing quality is narrow: teacher assignments are governed by centralized,

[16]This test is grounded on the notion that transition rates from academic-track schools to tertiary education were the same (Associazione Almadiploma, 2013). Because of that, my estimates do not represent the probability of attending a *Liceo Scientifico* conditional on going to university and the other controls.

quasi-random procedures, and all schools follow a nationally consistent syllabus. While a school's reputation could, in principle, influence family decisions, such factors are unlikely to cause a violation of the exclusion restriction here. When students in my sample chose their high school, no public ranking or systematic performance comparison existed to guide them. The first national tool enabling comparisons across schools — *Eduscopio*, an initiative by *Fondazione Agnelli* — was launched only in 2014, several years after the students in this study had already enrolled (Fondazione Agnelli, 2024).

Another potential source of bias relates to residential sorting — when families strategically choose to live in municipalities with greater access to *PNI* schools, anticipating academic advantages. Such behavior would represent a violation of the exclusion restriction, as the instrument could affect outcomes not only through increased treatment exposure but also through selection into municipalities based on unobserved preferences for STEM education. However, this concern appears limited in the Italian context. Students typically choose their high school after completing lower secondary education, around age 13–14, and internal mobility at this stage is uncommon. Unlike in other countries, boarding or distance high schools are virtually nonexistent in Italy — as of 2024, only eight operate nationwide. Existing research on internal migration patterns highlights that family moves are primarily driven by economic and employment opportunities, not by proximity to specific schools (Bonifazi and Heins, 2000; Bonifazi, 2013). Moreover, the Italian National Institute of Statistics (*ISTAT*) reports that migration mainly involves individuals aged 15–39, suggesting that moves often occur after high school decision (ISTAT, 2018).

It is important to note that, given the nature of the identification strategy, the estimated coefficients should be interpreted as intention-to-treat (ITT) effects — capturing the impact of program availability, rather than actual enrollment. This reflects the average effect of being exposed to a higher probability of attending a *PNI* program, rather than the treatment effect on compliers specifically.

3.6 Results

I start by estimating the effect of attending a *Liceo Scientifico* offering the *PNI* program on major choice, looking specifically at the probability of enrolling in a STEM degree. I then show the impact of the treatment on second-order effects, such as the probability of dropping out of university and performance during the first year of higher education (expressed by the logarithmic transformation of the number of credits - ECTS - attained during the first year of enrollment, as more credits imply the completion of more exams). I exploit a linear 2SLS model to estimate the β_1 coefficient.

I estimate the effect of attending a *PNI* school in Table 3.10. The OLS estimates (shown in the odd-numbered columns) suggest no relationship between program exposure and the probability of enrolling in a STEM degree, but indicate a negative association with university dropout and a positive one with first-year performance. The first-stage relationship is strong across all specifications, with a coefficient of approximately 0.57. This implies that moving from a municipality with no schools offering the *PNI* program to one where all schools do increases the probability of treatment by nearly 60 percentage points. Given that the instrument has a mean of 0.52 and a standard deviation of 0.34, a one-standard-deviation increase corresponds to a 17 percentage point rise in treatment probability. The size of the first-stage F-statistic — 170 for STEM and dropout, and 160 for performance — consistently exceeds conventional thresholds, mitigating concerns about weak instruments.

Turning to the 2SLS estimates (even-numbered columns), I find a positive but statistically insignificant effect on both STEM enrollment and first-year academic performance. However, there is a small negative effect on the probability of dropping out of university, significant at the 10% level, suggesting that students exposed to the *PNI* program are somewhat less likely to leave higher education prematurely. Notably, the estimation of performance effects relies on a slightly reduced sample, as first-year credit data is often missing for students who drop out. Specifically, while only about 1% of non-dropout students have missing credits, the figure rises to 19% among those who drop out. This selective attrition could partially mask a potential positive impact of the *PNI* program on academic performance, as students who might have exited the system without this early preparation remain enrolled longer.

3.6.1 Gender Heterogeneity and Policy Implications

The interaction between the program and students' gender has important policy implications. Due to stereotypes about male-dominated careers, girls often choose majors leading to less paid jobs (Zafar, 2013). I estimate the impact of *PNI* separately for boys and girls and present the results in Table 3.11.

It is immediately evident how the results for the overall sample masked a substantial gender heterogeneity. For boys, the probability of enrolling in a STEM major increases by roughly 2 percentage points, representing a nearly 8% variation relative to the sample mean — statistically significant at the 10% level. I do not find any effect on the probability of dropout from higher education or performance in the first year of studies. I do not observe effects for girls across any of the outcomes, as the coefficients remain both economically and statistically insignificant. This suggests that while the *PNI* program may have influenced boys' educational trajectories, it did not generate measurable changes for female students. I show in Appendix C.1 that these results are robust to changes in the definition of school-cohort size (using quintiles instead of terciles), municipality population size (using terciles of the municipality population instead of separating between rural, urban, or metropolitan), and the inclusion of students who repeated at least one year of high school. Just like in the study by De Philippis (2023), the program effectively widens the gender gap in STEM enrollment.

A growing body of research suggests that the effectiveness of policies aimed at strengthening math and science education — and the gender heterogeneity in their effects — depends crucially on prevailing gender norms. In more traditional environments, such policies tend to benefit boys disproportionately. For example, De Philippis (2023), studying the UK, attributes the lack of an effect on girls to persistent gender differences in preferences, consistent with the findings of Reuben et al. (2017), Wiswall and Zafar (2017), and Zafar (2013) which document systematic differences in preferences as one of the key reasons of the STEM gender gap. By contrast, in more gender-equal societies, such as Denmark, Joensen and Nielsen (2016) document positive effects of increased exposure to mathematics on both boys and girls, suggesting that cultural norms can either constrain or enable girls' responsiveness to STEM-related interventions. The Italian setting — marked by persistent and traditional gender norms — provides an ideal case to test whether increasing exposure to scientific disciplines can be effective in narrowing gender disparities where they are most entrenched.

3.6.2 The Interplay between Attending a *PNI* School and Local Gender Norms

Italy has long been marked by persistent economic disparities linked to geographical location (Giuntella, 2022). According to ISTAT (2018), regional differences also extend to attitudes toward traditional gender roles and the acceptance of gender-based violence. These differences are especially relevant when studying educational and career choices, which are shaped by the social context in which students are raised. To capture local variation in gender norms, I rely on the results of the 1981 Italian abortion referendum, where citizens voted on a proposal to significantly restrict access to abortion rights granted just a few years earlier. Although the proposal was ultimately rejected, support varied widely across municipalities, reflecting persistent geographic differences in attitudes toward gender roles. This historical measure provides a geographically and culturally anchored proxy for traditional gender norms — one that is especially relevant in the Italian context, where high school choice is highly localized and adolescent migration is rare. This choice is grounded in the literature on the persistence of cultural traits across generations and their influence on contemporary behavior (Becker et al., 2016; Alesina et al., 2013; Fernández, 2007), as well as recent studies set in Switzerland that employ referendum outcomes to measure local gender attitudes (Kaiser and Mata, 2025; Arni et al., 2024).

Compared to commonly used indicators such as the female-to-male labor force participation ratio (Meluzzi, 2024; Jayachandran, 2015; Fogli and Veldkamp, 2011; Fernández and Fogli, 2009; Fortin, 2005), the referendum-based measure offers two key advantages: it is not mechanically linked to local labor market conditions that may directly influence STEM enrollment, and it provides a more stable reflection of the cultural environment in which students were socialized. I validate the measure (in Appendix C.2, where I provide a more comprehensive discussion) by showing it is negatively and significantly associated with female labor force participation and share of women elected in city councils (and subsequently nominated in administrative positions) 30 years later, even after accounting for municipality size and province fixed effects — supporting its interpretation as a distinct and meaningful proxy for gender norms.

Finally, I estimate the impact of the treatment on girls by level of gender attitudes. To do so, I split the sample into quartiles — calculated at the municipality level. The association between the local support for increasing abortion restrictions and the municipality population that existed in 1981 remains evident and is displayed in the number of observations decreasing when moving from the second to the third and fourth quartiles. The results in Table 3.12 reveal a highly significant effect on STEM enrollment among girls in the most conservative municipalities, where exposure to the *PNI* program appears to have a substantial impact. In these areas, the effect size reaches almost 7 percentage points, a striking increase of 32% of the baseline. For boys, except those in the third quartile of gender norms, the estimated STEM responses across gender-norm quartiles are of similar size. They range from 2.5 to 3.6 percentage points, but are much less precise. Moreover, unlike for girls, there is no clear pattern in the relationship between boys' outcomes and the local normative environment.

The magnitude of my estimated ITT effect — roughly a 30% increase over the baseline STEM enrollment rate for girls in conservative areas — stands out relative to the existing literature. For example, Goodman (2019) evaluates a broad policy shift in high school math requirements and finds a much smaller ITT effect, approximately 3.5% of the baseline - likely reflecting the fact that the policy was not specifically targeted at high-achieving students. Similarly, De Philippis (2023) finds an ITT effect of 7% of the general population's baseline. However, the effect is far more pronounced among top students, whose probability of choosing a medicine or engineering degree increases by 27% and 17% relative to the baseline, respectively. The literature on exposure to role models reports substantially larger effects. Following targeted exposure to inspirational or relatable figures, Porter and Serra (2020) and Breda et al. (2023) find an increase in economics and STEM enrollment of around 27% and 28% over the baseline, respectively. Taken together, these comparisons suggest that the size of my estimated effect, while large, is plausible given the context and the specific subgroup it affects.

I also examine the variation behind the magnitude of this effect. I begin by geographically locating girls in the top quartile of conservative gender norms. While over half come from *Lombardia* or *Veneto*, *Trentino Alto-Adige* also contributes disproportionately relative to its national share of the female student population. Several regions from the South are sizeably represented too, such as *Campania*, *Puglia*, *Calabria*, and *Sicilia*. Since identification is based on within-cohort-by-province comparisons, I analyze patterns of STEM enrollment at that level. A potential concern is that some cohort-by-province cells may include very few students or feature strong imbalances between treatment and control groups. To address this, I conduct robustness checks restricting the sample to cells with at least 10 girls and with treatment-to-control ratios no lower than 0.25 and 0.35, respectively. While these restrictions substantially reduce the sample size, the magnitude of the estimated effects remains relatively stable, dropping to roughly 23% of the baseline probability of choosing a STEM degree and statistically significant only at the 10% level, when I look at the most restrictive comparison - where the sample size is reduced by a factor of 3. This suggests that the results are not driven by a small number of extreme or poorly balanced comparisons. A detailed analysis of this robustness exercise is provided in Appendix C.3.

The persistence of the effect across northern and southern regions complicates a purely economic interpretation. While labor market opportunities may play a role in regions like *Lombardia*, *Veneto*, and *Trentino*, the presence of similarly large effects in *Campania* and (especially) *Calabria* suggests another dynamic at play. One hypothesis is that girls from these more traditional areas — who may face stronger barriers to university participation

— are positively selected when they do enroll in higher education. Among them, exposure to the enhanced math and science curriculum of the *PNI* program may act as a critical nudge toward choosing a STEM major. This interpretation is supported by baseline outcomes, as girls from municipalities with the most conservative gender norms perform well during their first year of higher education and exhibit low dropout rates — reinforcing the idea of positive selection. A similar pattern holds for boys: those from more traditional areas also show slightly better baseline outcomes. However, unlike for girls, there is no clear relationship between boys' outcomes and the prevailing gender attitudes in their municipality.

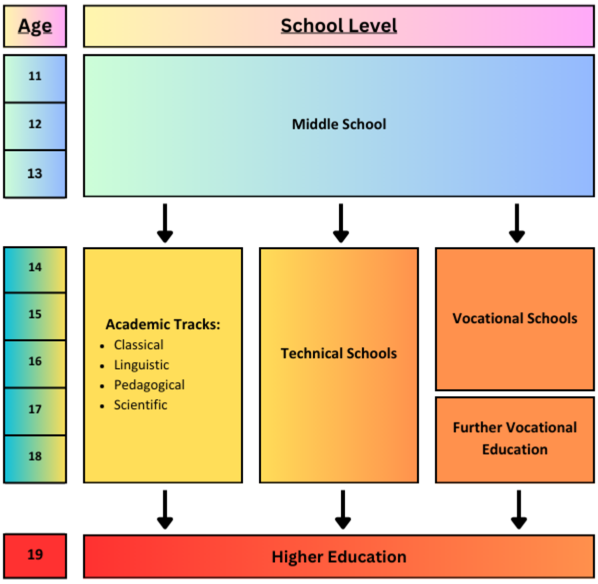
3.7 Conclusion

This paper studies the effect of strengthened scientific instruction on students' educational trajectories by examining the Italian *PNI* program, which introduced more advanced mathematics and physics coursework in a subset of high schools. The program had a modest average effect on the likelihood of enrolling in a STEM degree, with no measurable impact on dropout or first-year academic performance. I document that the program contributes to widening the gender gap in STEM enrollment: while male students respond positively to the treatment, female students, on average, do not. However, this pattern masks an additional, crucial, layer of heterogeneity.

The central contribution of the paper lies in identifying the role of prevailing gender norms in shaping the policy's effectiveness. To do so, I construct a novel proxy for local gender attitudes, based on municipality-level support for the 1981 abortion referendum — a historically rooted measure grounded in the documented persistence of cultural norms over time (Kaiser and Mata, 2025; Arni et al., 2024; Alesina et al., 2013; Fernández, 2007). I show that, although both boys and girls from more traditional municipalities appear positively selected into university, the effect of the *PNI* program on girls is particularly pronounced in these settings. The estimates suggest that greater exposure to scientific curricula significantly increases the probability that girls from more conservative backgrounds choose a STEM major — pointing to the potential of curricular interventions to mitigate the influence of restrictive cultural attitudes.

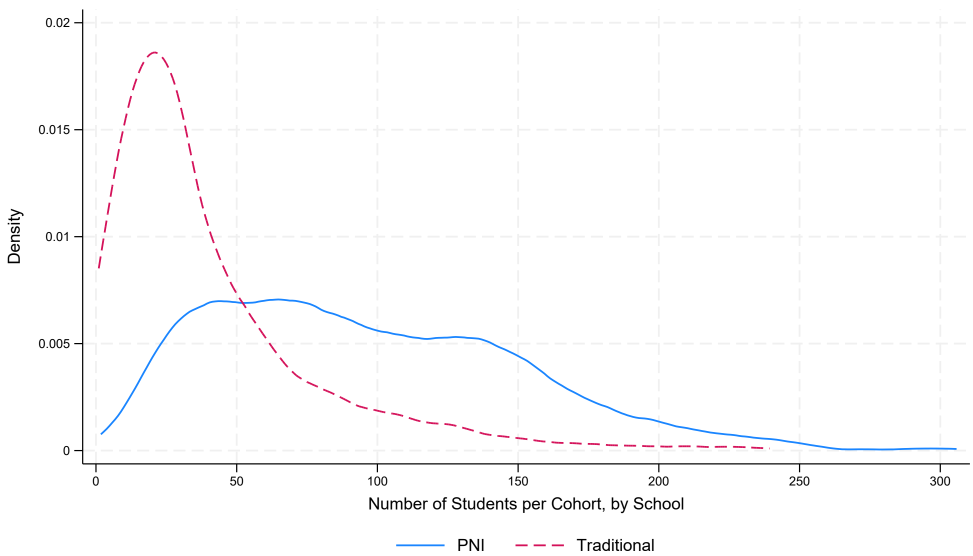
These findings suggest that policies aimed at strengthening scientific education can do more than broaden access to STEM fields: they can actively challenge gendered expectations and promote more equitable academic choices. While the average effect on female students is negligible, I show that in contexts where traditional norms are more deeply entrenched, such interventions can have transformative effects. The magnitude of the impact is comparable to that of those highly successful interventions involving exposure to female role models (Breda et al., 2023; Porter and Serra, 2020), underscoring the value of early, curriculum-based strategies. Conversely, the absence of an effect in less traditional areas — where gender gaps are already narrower — suggests that other factors, such as individual preferences or institutional constraints, may play a more central role in those contexts.

Figure 3.1: School Progression in Italy



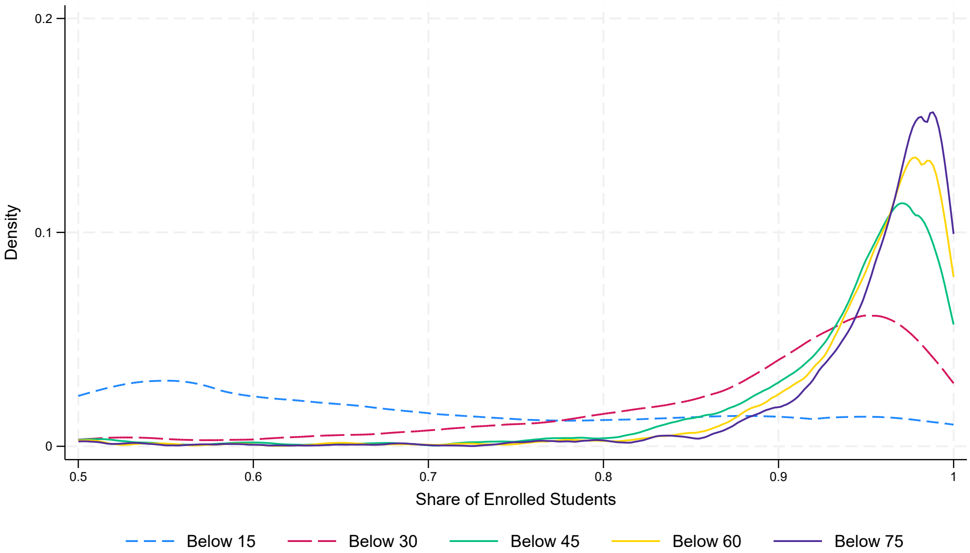
Notes: The figure illustrates Italian students' potential (higher) secondary education choices. The focus is on the options available to students who choose the academic track. These were the options available before the *Gelmini* Reform, which entered into force in the academic year 2010/2011. My sample focuses on students who enrolled in the previous five academic years and therefore were exposed to this set of choices.

Figure 3.2: Distribution of School-Cohort Size: *PNI* vs. Traditional Schools



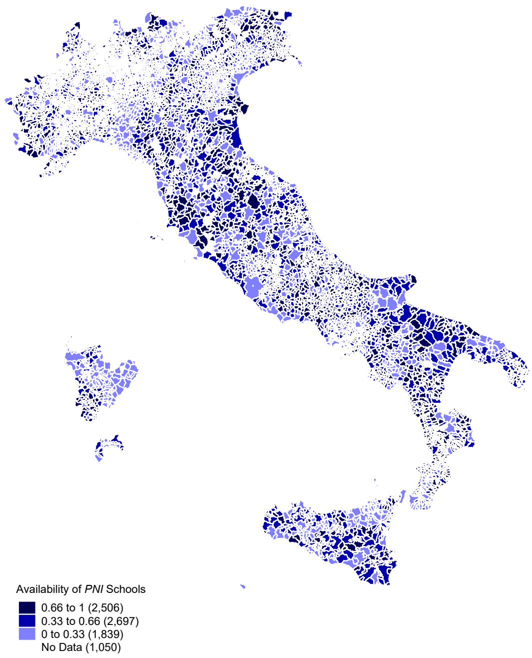
Notes: The graph shows the distribution of school-cohort size - as in, the number of students who, in a given academic year and from a given school, enroll in higher education - by school, separating between schools offering the *PNI* program and those offering only the Traditional one. Only students who attended *Liceo Scientifico*, who enrolled in higher education between the academic years 2010/2011 and 2014/2015, did not repeat one or more years, and enrolled immediately in higher education are included in the cohorts. I also exclude all students and schools from the provinces of *Aosta*, *Alto-Adige*, *Carbonia-Iglesias*, *Medio Campidano*, *Ogliastra*, and *Olbia-Tempio*. The sample size is of 1,152 schools.

Figure 3.3: Share of Students of each *Liceo Scientifico* by Duration of their Commute



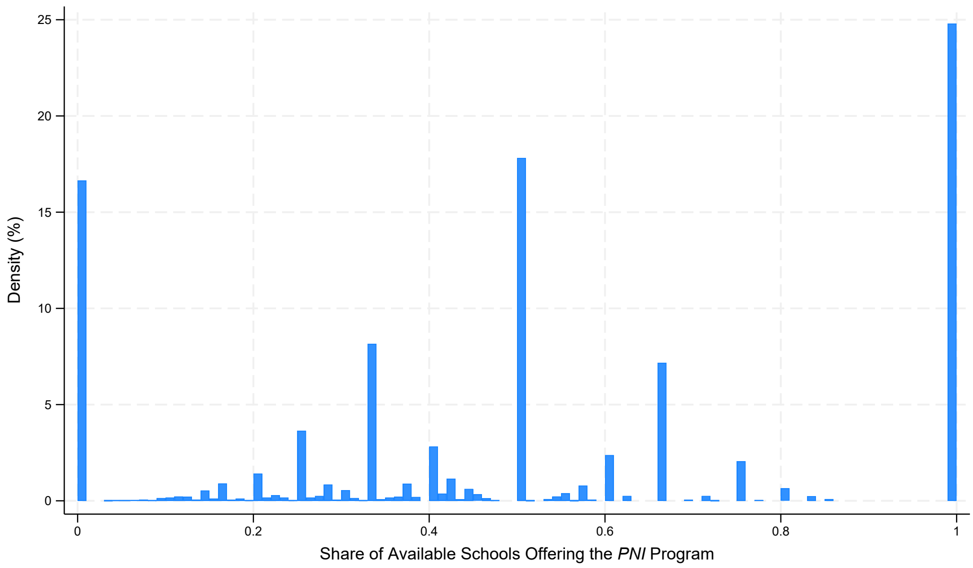
Notes: The graph shows the share of students enrolled in all the academic track scientific high schools in Italy by the duration of their commute to school. Commute time is calculated using the Google Maps API, simulating a trip by car. I include the students who achieve a diploma at *Liceo Scientifico* (the scientific curriculum of the academic-track high schools) and then enroll in higher education. I exclude all students and schools from the provinces of *Aosta*, *Alto-Adige*, *Carbonia-Iglesias*, *Medio Campidano*, *Ogliastra*, and *Olbia-Tempio*. The sample size is 1,152.

Figure 3.4: The Instrumental Variable: Map of the Availability of *PNI* Schools



Notes: The graph maps the between-municipalities variation of the instrumental variable, the share of schools offering a *PNI* program available to residents of a certain municipality. It is calculated by linking each school to all the municipalities of residence of the students enrolled, excluding those with a commute longer than 60 minutes (estimated using Google Maps API as the time required to drive from the student municipality of residence to the street address of the school). The number of observations per category is reported in parenthesis. I exclude all municipalities and schools from the provinces of *Aosta*, *Alto-Adige*, *Carbonia-Iglesias*, *Medio Campidano*, *Ogliastra*, and *Olbia-Tempio*. The sample size is 7,042 municipalities. That does not include the 692 for which there are no data (189 of those located in the provinces I exclude).

Figure 3.5: The Instrumental Variable: Availability of *PNI* Schools



Notes: The graph shows the distribution of the instrumental variable, the share of schools offering a *PNI* program available to residents of a certain municipality. The numerator of this variable is the number of *PNI* schools to which students could have enrolled, depending on their municipality of residence. The denominator consists of the total number of academic track scientific high schools available to residents of that same municipality. Only schools requiring a commute below 60 minutes are included in the numerator and denominator. I exclude all students and schools from the provinces of *Aosta*, *Alto-Adige*, *Carbonia-Iglesias*, *Medio Campidano*, *Ogliastra*, and *Olbia-Tempio*. The sample size is 7,042.

Table 3.1: Timetables for Traditional and *PNI Liceo Scientifico*

	Traditional					PNI				
Subjects	I	II	III	IV	V	I	II	III	IV	V
Italian	4	4	4	3	4	4	4	4	3	4
Latin	4	5	4	4	3	4	5	4	4	3
Mathematics	5	4	3	3	3	5	5	5	5	5
Physics	-	-	2	3	3	2	3	3	3	3
Natural Sciences, Chemistry, and Geography	2	2	3	3	2	2	2	3	3	2
Foreign Language and Literature	3	4	3	3	4	3	4	3	3	4
History	3	2	2	2	3	3	2	2	2	3
Philosophy	-	-	2	3	3	-	-	2	3	3
Art History	2	2	2	2	2	2	2	2	2	2
Physical Education	2	2	2	2	2	2	2	2	2	2
Religion	1	1	1	1	1	1	1	1	1	1
Total	25	27	28	29	30	28	31	31	31	32

Notes: The table shows the differences in weekly timetable between the Traditional and the *PNI* program of *Liceo Scientifico*, for each of the 5 years of secondary education.

Table 3.2: University Students by Diploma Type and Academic Year

Academic Year	2010/2011	2011/2012	2012/2013	2013/2014	2014/2015	Total
Scientific Studies	0.410	0.418	0.418	0.411	0.408	0.413
Classical Studies	0.176	0.175	0.174	0.165	0.162	0.170
Linguistic Studies	0.061	0.060	0.059	0.060	0.065	0.061
Pedagogical Studies	0.071	0.073	0.071	0.076	0.072	0.073
Non-Academic Tracks	0.282	0.274	0.279	0.288	0.284	0.283
Number of Students	200,909	197,100	191,946	190,248	191,978	972,181

Notes: The table shows the share of students graduating from high school who enroll in higher education, by academic year and type of high school. I separate between each of the academic-track high schools(*Liceo Scientifico*, *Liceo Classico*, *Liceo Linguistico*, and *Liceo Pedagogico*) and technical or vocational curricula. Students who failed one year or more are excluded, as well as those who do not immediately enter higher education after achieving their diploma.

Table 3.3: University Students' Characteristics by Diploma Type

	High School Curriculum				
Variables	Scientific	Classical	Linguistic	Pedagogical	Non-Academic
Percentage of Girls	0.496	0.706	0.867	0.911	0.467
	(0.500)	(0.456)	(0.340)	(0.285)	(0.499)
High School Mark	80.245	82.080	80.984	79.617	79.343
	(11.802)	(11.900)	(11.379)	(11.287)	(11.469)
STEM Enrollment	0.244	0.144	0.073	0.083	0.177
	(0.430)	(0.351)	(0.261)	(0.276)	(0.382)
First Year ECTS	41.783	42.573	44.147	41.032	38.290
	(17.734)	(18.608)	(18.714)	(18.396)	(18.937)
Dropout Rate	0.452	0.443	0.430	0.494	0.572
	(0.498)	(0.497)	(0.495)	(0.500)	(0.495)
Number of Students	401,536	165,596	59,175	70,540	275,334

Notes: The table shows student characteristics by the type of high school attended for students who enrolled in higher education between 2010/2011 and 2014/2015. I separate between each of the academic-track high schools(*Liceo Scientifico*, *Liceo Classico*, *Liceo Linguistico*, and *Liceo Pedagogico*) and technical or vocational curricula. Students who failed one year or more are excluded, as well as those who do not immediately enter higher education after achieving their diploma. High School Mark is expressed on a scale from 60 to 100 (a score below 60 implies failing to pass the exam). ECTS are granted upon exam completion, and students can acquire a maximum of 60 during each academic year. I consider dropouts students who officially abandon their degree and those who remain enrolled but have not concluded their studies as of the academic year 2022/2023. The total number of observations is 978,754. Standard deviations are in parentheses.

Table 3.4: Characteristics of *Liceo Scientifico* Graduates, by *PNI* Status

Variables	*Liceo Scientifico* Traditional	*Liceo Scientifico* *PNI*	P-value of the Difference
Share of Girls	0.488	0.499	0.000
	(0.500)	(0.500)	
Diploma Score	79.757	80.474	0.000
	(11.955)	(11.724)	
STEM Enrollment	0.233	0.249	0.000
	(0.423)	(0.433)	
First Year ECTS	40.539	42.346	0.000
	(18.040)	(17.565)	
Dropout Rate	0.484	0.438	0.000
	(0.500)	(0.496)	
Number of Students	118,191	279,928	

Notes: The table shows student characteristics by the type of high school attended for students who enrolled in higher education between 2010/2011 and 2014/2015. I separate between students who graduated from a *Liceo Scientifico* offering the *PNI* program or the traditional program only. I perform a t-test to assess the difference, with the null hypothesis stating that the means are equal. Students who failed one year or more are excluded, as well as those who do not immediately enter higher education after achieving their diploma. High School Mark is expressed on a scale from 60 to 100 (a score below 60 implies failing to pass the exam). ECTS are granted upon exam completion, and students can acquire a maximum of 60 during each academic year. I consider dropouts students who officially abandon their degree and those who remain enrolled but have not concluded their studies as of the academic year 2022/2023. I exclude all students and schools from the provinces of *Aosta*, *Alto-Adige*, *Carbonia-Iglesias*, *Medio Campidano*, *Ogliastra*, and *Olbia-Tempio*. The total number of observations is 400,420. Standard deviations are in parentheses.

Table 3.5: Characteristics of the High Schools, by *PNI* Status

Variables	*Liceo Scientifico* Traditional	*PNI*	P-value of the Difference
Characteristics of the High Schools			
Average Number of Students (by cohort)	43	97	0.000
Median Number of Students (by cohort)	31	88	
	(39.241)	(56.223)	
Average Commute Time (in minutes)	11	9	0.000
Median Commute Time (in minutes)	10	8	
	(8.114)	(7.284)	
Demographic Characteristics of the Schools' Municipalities			
Average Population	334,758	231,882	0.000
Median Population	42,602	33,669	
	(671,256)	(580,884)	
Average Population (residents per square KM)	1,794	1,268	0.000
Median Population (residents per square KM)	893	591	
	(2,178)	(1,776)	
Percentage in Rural Municipalities	3.33	3.49	0.864
	(17.89)	(18.31)	
Percentage in Urban Municipalities	72.50	81.61	0.000
	(44.68)	(38.75)	
Percentage in Metropolitan Municipalities	24.17	14.90	0.000
	(42.85)	(35.64)	
Number of Schools	575	577	

Notes: The table shows the characteristics of the *Liceo Scientifico* high schools in the sample and the municipalities where they are located, over the academic years 2010/2011 and 2014/2015. I also provide the percentage of schools by type of municipality (along the rural, urban, and metropolitan categories). I compare the high schools offering the *PNI* program and those offering only the traditional program. I perform a t-test to assess the difference, with the null hypothesis stating that the means are equal. Standard deviations are in parentheses.

Table 3.6: Regional Distribution of *PNI* Schools and *PNI* Students

	PNI Schools			*PNI* Students		
Region	Traditional	*PNI*	Total	Traditional	*PNI*	Total
Piemonte	53.85	46.15	78	32.96	67.04	24,417
Lombardia	54.02	45.98	174	30.10	69.90	51,759
Trentino Alto-Adige	75.00	25.00	12	50.23	49.77	2,652
Veneto	54.76	45.24	84	32.76	67.24	27,777
Friuli Venezia Giulia	38.10	61.90	21	18.92	81.08	7,396
Liguria	43.48	56.52	23	21.52	78.48	8,874
Emilia Romagna	46.03	53.97	63	25.98	74.02	22,856
Toscana	37.31	62.69	67	17.52	82.48	22,017
Umbria	25.00	75.00	16	21.82	78.18	5,995
Marche	35.48	64.52	31	22.07	77.93	10,073
Lazio	48.72	51.28	117	24.94	75.06	40,947
Abruzzo	60.00	40.00	30	26.63	73.37	11,192
Molise	30.00	70.00	10	17.13	82.87	2,849
Campania	60.00	40.00	150	40.35	59.65	57,493
Puglia	44.71	55.29	85	28.89	71.11	34,651
Basilicata	45.00	55.00	20	43.63	56.37	5,139
Calabria	32.69	67.31	52	24.04	75.96	17,638
Sicilia	54.05	45.95	111	28.17	71.83	38,219
Sardegna	71.43	28.57	21	57.38	42.62	6,175
Total	575	577	1,152	118,191	279,928	398,119
Percent	50.57	49.43	100	29.69	70.31	100

Notes: The table shows the percentage of the high schools offering the *PNI* program and the number of students attending them, by region. I include only students who graduated from a *Liceo Scientifico*, who enrolled in higher education between the academic years 2010/2011 and 2014/2015, who did not fail one year or more years, and who did not immediately enter higher education after achieving their diploma. I also exclude all students and schools from the provinces of *Aosta*, *Alto-Adige*, *Carbonia-Iglesias*, *Medio Campidano*, *Ogliastra*, and *Olbia-Tempio*.

Table 3.7: Balancing Exercise: *PNI* Availability and Municipality Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome Variable	Labor Force Participation		Employment		Foreign-Born Residents	
Share of Available *PNI*	-0.005**	0.0003	0.001	-0.001	-0.004***	-0.002
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Mean of the Outcome	0.50	0.50	0.89	0.89	0.06	0.06
SD of the Outcome	0.06	0.06	0.06	0.06	0.04	0.04
Observations	7,042	7,042	7,042	7,042	7,042	7,042
Province Fixed Effects	X	X	X	X	X	X
Municipality Type		X		X		X

	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Outcome Variable	Women's LFP		Women's Employment		Women in ECs		Abortion Restrictions	
Share of Available *PNI*	-0.008***	-0.003	-0.001	-0.003	0.009	0.006	0.015***	0.008**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.008)	(0.008)	(0.003)	(0.003)
Mean of the Outcome	0.40	0.40	0.87	0.87	0.17	0.17	0.34	0.34
SD of the Outcome	0.07	0.07	0.08	0.08	0.19	0.19	0.11	0.11
Observations	7,042	7,042	7,042	7,042	6,926	6,926	7,018	7,018
Province Fixed Effects	X	X	X	X	X	X	X	X
Municipality Type		X		X		X		X

Notes: I estimate the relationship between the characteristics of the municipalities of residence of the students in my sample and the share of available schools offering the *PNI* program (Equation 3.3). The share of available *PNI* is the ratio between the number of *Liceo Scientifico* schools offering the *PNI* option and the number of *Liceo Scientifico* schools within a 60-minute commute from the municipality of residence of the student. The dependent variables showed in the top panel are: the share of residents (above 15 years of age) who are in the labor force; the share of residents in the labor force who are employed; the share of foreign-born residents (of any age). The dependent variables showed in the bottom panel are: the share of female residents (above 15 years of age) who are in the labor force; the share of female residents in the labor force who are employed; the share of elected women administrators in the municipality executive committee (ECs), which is appointed by the elected major and carries out day-to-day governance; and the share of voters in the municipality who supported the restricting access to abortion in the 1981 abortion referendum. I control for province fixed effects and a categorical variable indicating whether a municipality is considered rural, urban, or metropolitan. I exclude all municipalities from the provinces of *Aosta*, *Alto-Adige*, *Carbonia-Iglesias*, *Medio Campidano*, *Ogliastra*, and *Olbia-Tempio*. Standard errors are clustered at the municipality level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3.8: Relationship Between *PNI* Availability and Alternative Academic Track Choices

Outcome Variable	(1)	(2)	(3)
	Probability of Choosing *Liceo Scientifico*		
Share of Available *PNI*	0.027	0.030	0.030
	(0.029)	(0.024)	(0.024)
Mean of the Outcome	0.59	0.59	0.59
SD of the Outcome	0.49	0.49	0.49
Observations	704,150	704,149	704,146
Province Fixed Effects		X	X
Cohort Fixed Effects		X	X
Cohort-Province Fixed Effects			X

Notes: I estimate the relationship between the probability of choosing to attend a *Liceo Scientifico* and the share of available schools offering the *PNI* program (Equation 3.4). The share of available *PNI* is the ratio between the number of *Liceo Scientifico* schools offering the *PNI* option and the number of *Liceo Scientifico* schools within a 60-minute commute from the municipality of residence of the student. I progressively add categorical variables controlling for province and cohort fixed effects (column (2)), and then for their interaction (column (3)). The association is estimated at the student level, including all students enrolled in academic track schools (*Liceo Classico*, *Liceo Linguistico*, *Liceo Pedagogico*, and *Liceo Scientifico*). I include students who graduated between the academic years 2010/2011 and 2014/2015, did not repeat one or more years of high school, and enrolled immediately in higher education. I exclude all students and schools from the provinces of *Aosta*, *Alto-Adige*, *Carbonia-Iglesias*, *Medio Campidano*, *Ogliastra*, and *Olbia-Tempio*. Standard errors are clustered at the school level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3.9: 2SLS Estimates: High School Outcomes

	(1)	(2)	(3)	(4)
Outcome Variable	Log of High School Mark		Probability of Repeating a Year	
Specification	OLS	2SLS	OLS	2SLS
PNI	0.005***	0.006	-0.022***	0.002
	(0.002)	(0.005)	(0.005)	(0.008)
First-Stage	-	0.570***	-	0.567***
	-	(0.038)	-	(0.035)
F-statistic	-	172.14	-	189.31
Mean of the Outcome	80.30	80.30	0.10	0.10
SD of the Outcome	11.80	11.80	0.30	0.30
Observations	398,119	398,119	441,786	441,786
Cohort-Province Fixed Effects	X	X	X	X
Gender	X	X	X	X
Tercile of School-Cohort Size	X	X	X	X
Type of Municipality	X	X	X	X

I estimate the relationship between the logarithmic transformation of the final high school mark (which is received after completing the final high school exam and ranges between 60 and 101), whether a student repeated at least one year in high school, and attending an academic track scientific high school (*Liceo Scientifico*) offering the *PNI* program using Equation 3.1. I use an OLS estimator in the odd-numbered columns and a 2SLS estimator in the even-numbered ones. I instrument *PNI* assignment using the share of available *PNI*, which is the ratio between the number of *Liceo Scientifico* schools offering the *PNI* option and the number of *Liceo Scientifico* schools within a 60-minute commute from the municipality of residence of the student. I show the coefficient of the first-stage Equation 3.2 regression and the relative F-statistic. I control for cohort-by-province fixed effects, gender, a categorical variable representing the tercile number of students in the school-cohort group, and a categorical variable indicating whether a municipality is considered rural, urban, or metropolitan. Only students who attended *Liceo Scientifico* and graduated between the academic years 2010/2011 and 2014/2015 are included. Additionally, when I estimate the effect on high school mark I consider only those who did not repeat one or more years and who enrolled immediately in higher education; when I estimate the effect on the probability of repeating one year, I only include students who enrolled immediately in higher education. I also exclude all students and schools from the provinces of *Aosta*, *Alto-Adige*, *Carbonia-Iglesias*, *Medio Campidano*, *Ogliastra*, and *Olbia-Tempio*. Standard errors are clustered at the school level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3.10: 2SLS Estimates: Probability of Choosing a STEM Degree

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome Variable	STEM		Dropout		Performance	
Specification	OLS	2SLS	OLS	2SLS	OLS	2SLS
PNI	0.005	0.010	-0.021***	-0.019*	0.031***	0.007
	(0.004)	(0.008)	(0.004)	(0.010)	(0.005)	(0.014)
First-Stage	-	0.570***	-	0.570***	-	0.566***
	-	(0.038)	-	(0.038)	-	(0.038)
F-statistic	-	172.14	-	172.14	-	160.32
Mean of the Outcome	0.25	0.25	0.45	0.45	41.86	41.86
SD of the Outcome	0.43	0.43	0.50	0.50	17.70	17.70
Observations	398,119	398,119	398,119	398,119	360,970	360,970
Cohort-Province Fixed Effects	X	X	X	X	X	X
Gender	X	X	X	X	X	X
Tercile of School-Cohort Size	X	X	X	X	X	X
Type of Municipality	X	X	X	X	X	X

Notes: I estimate the relationship between the outcomes of interest and attending an academic track scientific high school (*Liceo Scientifico*) offering the *PNI* program using Equation 3.1. The outcomes are: probability of choosing a STEM degree ("STEM"), the probability of dropping out from higher education ("Dropout") and students' performance during their first year of university ("Performance", which is proxied by the number of credits, ECTS, acquired during the first academic year. ECTS are granted upon exam completion, and students can acquire a maximum of 60 during each academic year). I use an OLS estimator in the odd-numbered columns and a 2SLS estimator in the even-numbered ones. I instrument *PNI* assignment using the share of available *PNI*, which is the ratio between the number of *Liceo Scientifico* schools offering the *PNI* option and the number of *Liceo Scientifico* schools within a 60-minute commute from the municipality of residence of the student. I show the coefficient of the first-stage Equation 3.2 regression and the relative F-statistic. I control for cohort-by-province fixed effects, gender, a categorical variable representing the tercile number of students in the school-cohort group, and a categorical variable indicating whether a municipality is considered rural, urban, or metropolitan. Only students who attended *Liceo Scientifico*, who enrolled in higher education between the academic years 2010/2011 and 2014/2015, did not repeat one or more years, and enrolled immediately in higher education are included. I also exclude all students and schools from the provinces of *Aosta*, *Alto-Adige*, *Carbonia-Iglesias*, *Medio Campidano*, *Ogliastra*, and *Olbia-Tempio*. Standard errors are clustered at the school level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3.11: 2SLS Estimates: Heterogeneity by Gender

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome Variable	STEM		Dropout		Performance	
Subsample	Boys	Girls	Boys	Girls	Boys	Girls
PNI	0.021*	0.003	-0.022	-0.017	0.009	0.005
	(0.012)	(0.010)	(0.014)	(0.012)	(0.020)	(0.015)
First-Stage	0.561***	0.578***	0.561***	0.578***	0.555***	0.574***
	(0.039)	(0.038)	(0.039)	(0.038)	(0.039)	(0.038)
F-statistic	225.62	179.11	225.62	179.11	208.90	170.00
Mean of the Outcome	0.26	0.23	0.48	0.42	40.68	43.01
SD of the Outcome	0.44	0.42	0.50	0.49	18.03	17.29
Observations	201,055	197,058	201,055	197,058	178,840	182,227
Cohort-Province Fixed Effects	X	X	X	X	X	X
Tercile of School-Cohort Size	X	X	X	X	X	X
Type of Municipality	X	X	X	X	X	X

Notes: I estimate the relationship between the outcomes of interest and attending an academic track scientific high school (*Liceo Scientifico*) offering the *PNI* program using Equation 3.1, for boys and girls separately. The outcomes are: probability of choosing a STEM degree ("STEM"), the probability of dropping out from higher education ("Dropout") and students' performance during their first year of university ("Performance", which is proxied by the number of credits, ECTS, acquired during the first academic year. ECTS are granted upon exam completion, and students can acquire a maximum of 60 during each academic year). I use a 2SLS estimator in all columns. I instrument *PNI* assignment using the share of available *PNI*, which is the ratio between the number of *Liceo Scientifico* schools offering the *PNI* option and the number of *Liceo Scientifico* schools within a 60-minute commute from the municipality of residence of the student. I show the coefficient of the first-stage Equation 3.2 regression and the relative F-statistic. I control for cohort-by-province fixed effects, a categorical variable representing the tercile number of students in the school-cohort group, and a categorical variable indicating whether a municipality is considered rural, urban, or metropolitan. Only students who attended *Liceo Scientifico*, who enrolled in higher education between the academic years 2010/2011 and 2014/2015, did not repeat one or more years, and enrolled immediately in higher education are included. I also exclude all students and schools from the provinces of *Aosta*, *Alto-Adige*, *Carbonia-Iglesias*, *Medio Campidano*, *Ogliastra*, and *Olbia-Tempio*. Standard errors are clustered at the school level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3.12: 2SLS Estimates: *PNI* and Local Gender Norms

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Outcome Variable	STEM		Dropout		Performance		STEM		Dropout		Performance	
Subsample	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
	1st Quartile of Conservative Gender Norms (Approval Rate: 21%)						2nd Quartile of Conservative Gender Norms (Approval Rate: 29%)					
PNI	0.030	-0.027	-0.037	-0.019	0.011	-0.009	0.036*	0.002	-0.027	-0.024	0.010	-0.010
	(0.026)	(0.022)	(0.029)	(0.024)	(0.044)	(0.035)	(0.020)	(0.017)	(0.025)	(0.020)	(0.031)	(0.026)
First-Stage	0.453***	0.471***	0.453***	0.471***	0.445***	0.470***	0.590***	0.622***	0.590***	0.622***	0.583***	0.613***
	(0.061)	(0.058)	(0.061)	(0.058)	(0.061)	(0.058)	(0.070)	(0.065)	(0.070)	(0.065)	(0.069)	(0.065)
F-statistic	108.85	84.24	108.85	84.24	96.27	78.72	129.23	95.67	129.23	95.67	117.19	88.04
Mean of the Outcome	0.26	0.24	0.47	0.42	41.38	43.65	0.26	0.24	0.49	0.43	39.04	41.24
SD of the Outcome	0.44	0.42	0.50	0.49	17.86	17.20	0.44	0.43	0.50	0.50	17.95	17.19
Observations	61,260	60,029	61,260	60,029	54,992	55,545	75,567	70,400	75,567	70,400	66,177	64,520
Cohort-Province Fixed Effects	X	X	X	X	X	X	X	X	X	X	X	X
Tercile of School-Cohort Size	X	X	X	X	X	X	X	X	X	X	X	X
Type of Municipality	X	X	X	X	X	X	X	X	X	X	X	X

	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
Outcome Variable	STEM		Dropout		Performance		STEM		Dropout		Performance	
Subsample	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
	3rd Quartile of Conservative Gender Norms (Approval Rate: 36%)						4th Quartile of Conservative Gender Norms (Approval Rate: 49%)					
PNI	-0.001	-0.010	-0.002	-0.010	0.014	-0.0003	0.025	0.067***	-0.009	-0.015	-0.022	0.023
	(0.021)	(0.019)	(0.024)	(0.021)	(0.037)	(0.026)	(0.026)	(0.023)	(0.032)	(0.027)	(0.043)	(0.031)
First-Stage	0.612***	0.624***	0.612***	0.624***	0.610***	0.622***	0.622***	0.582***	0.622***	0.582***	0.592***	0.574***
	(0.070)	(0.067)	(0.070)	(0.067)	(0.070)	(0.066)	(0.066)	(0.065)	(0.066)	(0.065)	(0.060)	(0.065)
F-statistic	62.47	52.27	62.47	52.27	60.47	50.91	86.44	81.07	86.44	81.07	84.29	78.55
Mean of the Outcome	0.27	0.23	0.47	0.42	41.09	43.32	0.26	0.21	0.44	0.39	43.66	46.22
SD of the Outcome	0.44	0.42	0.50	0.49	18.18	17.42	0.44	0.41	0.50	0.49	17.94	16.99
Observations	42,475	43,378	42,475	43,378	37,915	40,356	20,997	22,425	20,997	22,425	19,002	21,047
Cohort-Province Fixed Effects	X	X	X	X	X	X	X	X	X	X	X	X
Tercile of School-Cohort Size	X	X	X	X	X	X	X	X	X	X	X	X
Type of Municipality	X	X	X	X	X	X	X	X	X	X	X	X

Notes: I estimate the relationship between the outcomes of interest and attending an academic track scientific high school (*Liceo Scientifico*) offering the *PNI* program using Equation 3.1, for boys and girls separately, based on the quartile of the distribution of conservative gender norms of their municipality of residence. Gender norms are calculated using the share of votes in favor of the conservative norm in the 1981 abortion referendum. The outcomes are: probability of choosing a STEM degree ("STEM"), the probability of dropping out from higher education ("Dropout") and students' performance during their first year of university ("Performance", which is proxied by the number of credits, ECTS, acquired during the first academic year. ECTS are granted upon exam completion, and students can acquire a maximum of 60 during each academic year). I use a 2SLS estimator in all columns. I instrument *PNI* assignment using the share of available *PNI*, which is the ratio between the number of *Liceo Scientifico* schools offering the *PNI* option and the number of *Liceo Scientifico* schools within a 60-minute commute from the municipality of residence of the student. I show the coefficient of the first-stage Equation 3.2 regression and the relative F-statistic. I control for cohort-by-province fixed effects, a categorical variable representing the tercile number of students in the school-cohort group, and a categorical variable indicating whether a municipality is considered rural, urban, or metropolitan. Only students who attended *Liceo Scientifico*, who enrolled in higher education between the academic years 2010/2011 and 2014/2015, did not repeat one or more years, and enrolled immediately in higher education are included. I also exclude all students and schools from the provinces of *Aosta*, *Alto-Adige*, *Carbonia-Iglesias*, *Medio Campidano*, *Ogliastra*, and *Olbia-Tempio*. Standard errors are clustered at the school level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Conclusion

This thesis studies the determinants of educational outcomes at three critical stages of the academic lifecycle. It examines how student behavior, peer interactions, and curricular exposure shape both short- and long-term educational trajectories, with a particular focus on understanding the role of individual characteristics and contextual factors.

The first chapter investigates the impact of goal-setting interventions on student attendance in higher education. Using data from a randomized controlled trial at a UK public university, we find that treated students, who were asked to set an attendance goal, attended roughly one additional lecture over the term. While this increase in attendance did not translate into improved academic performance, it positively affected students' interest in their field of study. These findings highlight both the potential and limitations of goal-setting interventions for improving student engagement. On one hand, these strategies are easily scalable, relatively inexpensive, and particularly effective for students with weaker self-regulation and planning abilities. On the other hand, they do not necessarily lead to better academic outcomes, suggesting that the link between attendance and performance may be more complex than commonly assumed. This aligns with broader evidence that first-order behavioral changes, like increased attendance, do not always translate into second-order effects, such as improved grades or graduation rates.

The second chapter focuses on the consequences of relative academic rank, a form of peer effect determined by students' position within their peer group. It draws on a population study conducted in Aberdeen, Scotland, covering all children born in the city between 1950 and 1955. We find that a higher rank within the school-cohort group significantly improves both subsequent test scores and internalizing skills — a set of non-cognitive traits that include self-esteem and emotional stability, which other studies link to long-term educational and labor market success. In particular, rank positively influences performance on the high-stakes 11-plus exam, educational attainment, and adult earnings. However, these effects are strongly gendered: girls benefit more in terms of educational attainment, while boys experience larger gains in future earnings. This pattern likely reflects the broader historical constraints on women's access to higher education and skilled employment, highlighting the long-term consequences of early academic position.

The third chapter studies the impact of a program increasing the scientific content of secondary school curricula on students' educational choices, focusing on the gender gap in STEM degree enrollment. Using data from all Italian students in higher education, I exploit an instrumental variable approach to account for students' selection into the program, leveraging quasi-random variation in program availability to estimate its effects. The results indicate no significant average effects on STEM enrollment, dropout, or academic performance during the first year in higher education. However, the findings reveal substantial gender heterogeneity. While boys exposed to the program are significantly more likely to pursue STEM degrees, the average effect for girls is close to zero. Notably, the impact for girls varies significantly with local gender norms: girls exposed to the program are significantly more likely to choose STEM majors if they were raised in more conservative municipalities. These findings contribute to the growing literature on the interaction between cultural norms and educational choices, highlighting the potential for curricular interventions to counteract restrictive gender norms and reduce disparities in STEM participation, which could, in turn, impact the gender gap in earnings.

Bibliography

Acemoglu, D. and Autor, D. (2011). Skills, tasks, and technologies: Implications for employment and earnings. In Card, D. and Ashenfelter, O., editors, *Handbook of Labor Economics*, volume 4, pages 1043–1171. Elsevier.

Achenbach, T. M. and Edelbrock, C. S. (1978). The classification of child psychopathology: A review and analysis of empirical efforts. *Psychological Bulletin*, 85(6):1275–1301.

Agarwal, N. and Somaini, P. J. (2019). Revealed Preference Analysis of School Choice Models. NBER Working Papers 26568, National Bureau of Economic Research, Inc.

Alesina, A., Giuliano, P., and Nunn, N. (2013). On the origins of gender roles: Women and the plough *. *The Quarterly Journal of Economics*, 128(2):469–530.

Altonji, J. (1995). The effects of high school curriculum on education and labor market outcomes. *Journal of Human Resources*, 30(3):409–438.

Altonji, J., Arcidiacono, P., and Maurel, A. (2016). The analysis of field choice in college and graduate school. In *Handbook of the Economics of Education*, volume 5, chapter Chapter 7, pages 305–396. Elsevier.

Altonji, J., Blom, E., and Meghir, C. (2012). Heterogeneity in human capital investments: High school curriculum, college major, and careers. *Annual Review of Economics*, 4(1):185–223.

Angrist, J., Lang, D., and Oreopoulos, P. (2009). Incentives and services for college achievement: Evidence from a randomized trial. *American Economic Journal: Applied Economics*, 1(1):136–63.

Ariely, D. and Wertenbroch, K. (2002). Procrastination, deadlines, and performance: Self-control by precommitment. *Psychological Science*, 13(3):219–224. PMID: 12009041.

Arni, P., Lalive, R., and Wolter, S. C. (2024). Social norms and gendered occupational choices of men and women: Evidence from job applications in switzerland. *Industrial Relations*, 63(2):123–156.

Ashraf, N., Karlan, D., and Yin, W. (2006). Tying Odysseus to the Mast: Evidence From a Commitment Savings Product in the Philippines*. *The Quarterly Journal of Economics*, 121(2):635–672.

Associazione Almadiploma (2013). Indagine diplomati 2013.

Attanasio, O., Blundell, R., Conti, G., and Mason, G. (2020a). Inequality in socio-emotional skills: A cross-cohort comparison. *Journal of Public Economics*, 191:104171.

Attanasio, O., Ãureo de Paula, and Toppeta, A. (2020b). The Persistence of Socio-Emotional Skills: Life Cycle and Intergenerational Evidence. Working Papers 2020-066, Human Capital and Economic Opportunity Working Group.

Augenblick, N., Niederle, M., and Sprenger, C. (2015). Working over Time: Dynamic Inconsistency in Real Effort Tasks *. *The Quarterly Journal of Economics*, 130(3):1067–1115.

Barro, R. J. (2001). Human capital and growth. *American Economic Review*, 91(2):12–17.

Becker, S. O., Boeckh, K., Hainz, C., and Woessmann, L. (2016). The empire is dead, long live the empire! long-run persistence of trust and corruption in the bureaucracy. *The Economic Journal*, 126(590):40–74.

Behar, L. and Stringfield, S. (1974). A behavior rating scale for the preschool child. *Developmental Psychology*, 10(5):601–609.

Berk, L. E. (2023). *Development Through the Lifespan*. Pearson Education, 10th edition.

Bertrand, M. (2020). Gender in the labor market: The role of equal opportunity and expectations. *Annual Review of Economics*, 12(1):153–180.

Blau, F. D. and Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, 55(3):789–865.

Bonifazi, C. (2013). *L'Italia delle migrazioni*. Il Mulino.

Bonifazi, C. and Heins, F. (2000). Le migrazioni interne degli italiani: Geografie, caratteristiche e flussi migratori. *Bollettino del Centro Studi per il Mezzogiorno e le Migrazioni*, 20:19–35.

Boyle, M. H. and Jones, S. C. (1985). Selecting measures of emotional and behavioral disorders of childhood for use in general populations. *Journal of Child Psychology and Psychiatry*, 26(1):137–159.

Brade, R., Himmler, O., and Jäckle, R. (2018). Normatively Framed Relative Performance Feedback – Field Experiment and Replication. MPRA Paper 88830, University Library of Munich, Germany.

Breda, T., Grenet, J., Monnet, M., and Van Effenterre, C. (2023). How effective are female role models in steering girls towards stem? evidence from french high schools. *The Economic Journal*, 133(653):1773–1809.

Bruhn, M. and McKenzie, D. (2009). In pursuit of balance: Randomization in practice in development field experiments. *American Economic Journal: Applied Economics*, 1(4):200–232.

Buser, T., Niederle, M., and Oosterbeek, H. (2023). Gender, competitiveness, and career choices: Causal evidence from a choice experiment. *The Quarterly Journal of Economics*, 138(1):189–243.

BusinessEurope (2023). New survey of european companies highlights critical labour and skills shortages. https://www.businesseurope.eu/publications/

new-survey-of-european-companies-highlights-critical-labour-and-skills-shortages/. Accessed April 2025.

Cadena, B. C. and Keys, B. J. (2015). Human capital and the lifetime costs of impatience. *American Economic Journal: Economic Policy*, 7(3):126–153.

Card, D. (1999). The causal effect of education on earnings. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 3 of *Handbook of Labor Economics*, chapter 30, pages 1801–1863. Elsevier.

Carneiro, P., Cruz-Aguayo, Y., Salvati, F., and Schady, N. (2023). The effect of classroom rank on learning throughout elementary school: Experimental evidence from ecuador. *Journal of Labor Economics*, 0(ja):null.

Castleman, B. L. and Page, L. C. (2015). Summer nudging: Can personalized text messages and peer mentor outreach increase college going among low-income high school graduates? *Journal of Economic Behavior & Organization*, 115:144–160.

Cattivelli, V. (2021). Institutional methods for the identification of urban and rural areas—a review for italy. In Bisello, A., Vettorato, D., Ludlow, D., and Baranzelli, C., editors, *Smart and Sustainable Planning for Cities and Regions*, pages 187–207, Cham. Springer International Publishing.

Chise, D., Fort, M., and Monfardini, C. (2021). On the intergenerational transmission of stem education among graduate students. *The B.E. Journal of Economic Analysis and Policy*, 21(1):115–145.

Clark, D. and Del Bono, E. (2016). The long-run effects of attending an elite school: Evidence from the united kingdom. *American Economic Journal: Applied Economics*, 8(1):150–76.

Clark, D., Gill, D., Prowse, V., and Rush, M. (2020). Using Goals to Motivate College Students: Theory and Evidence From Field Experiments. *The Review of Economics and Statistics*, 102(4):648–663.

Cohodes, S. R. and Goodman, J. S. (2014). Merit aid, college quality, and college completion: Massachusetts' adams scholarship as an in-kind subsidy. *American Economic Journal: Applied Economics*, 6(4):251–85.

Colucci, D., Franco, C., and Valori, V. (2024). The endowment effect with different possession times and types of items. *Journal of Behavioral and Experimental Economics*, 110:102216.

Cornwell, C., Lee, K. H., and Mustard, D. (2005). Student responses to merit scholarship retention rules. *Journal of Human Resources*, 40(4):895–917.

Cortes, K. E., Goodman, J., and Nomi, T. (2015). Intensive math instruction and educational attainment: Long-run impacts of double-dose algebra. *Journal of Human Resources*, 50(1):108–158.

Creemers, D. H., Scholte, R. H., Engels, R. C., Prinstein, M. J., and Wiers, R. W. (2013). Damaged self-esteem is associated with internalizing problems. *Frontiers in Psychology*, 152(4).

Damgaard, M. T. and Nielsen, H. S. (2018). Nudging in education. *Economics of Education Review*, 64:313–342.

De Paola, M., Scoppa, V., and Nisticò, R. (2012). Monetary incentives and student achievement in a depressed labor market: Results from a randomized experiment. *Journal of Human Capital*, 6(1):56 – 85.

De Philippis, M. (2023). Stem graduates and secondary school curriculum. *Journal of Human Resources*, 58(6):1914–1947.

Deckman, M. and McTague, J. (2015). Religion makes the difference: Conservative religious women and political engagement in the united states. *Politics and Religion*, 8(1):22–47.

Delaney, J. M. and Devereux, P. J. (2021). High school rank in math and english and the gender gap in stem. *Labour Economics*, 69:101969.

Delavande, A., Bono, E., Holford, A., Sen, S., and Lesic, V. (2023). Expectations about the productivity of effort and academic outcomes: Evidence from a randomized information intervention. *SSRN Electronic Journal*.

Delavande, A., Bono, E. D., and Holford, A. (2022). BOOST2018: The Ground-Breaking Study of Student Life, 2015-2020: Secure Access. UK Data Service[Data Collection].

DellaVigna, S. and Malmendier, U. (2006). Paying not to go to the gym. *American Economic Review*, 96(3):694–719.

Deming, D. J., Hastings, J. S., Kane, T. J., and Staiger, D. O. (2014). School choice, school quality, and postsecondary attainment. *American Economic Review*, 104(3):991–1013.

Denning, J. T., Murphy, R., and Weinhardt, F. (2023). Class Rank and Long-Run Outcomes. *The Review of Economics and Statistics*, 105(6):1426–1441.

Dinkelman, T. and Martínez, C. A. (2014). Investing in schooling in chile: The role of information about financial aid for higher education. *Review of Economics and Statistics*, 96(2):244–257.

Directorate General for Employment (2023). Employment and social developments in europe 2023. Technical report, European Commission and Directorate-General for Employment, Social Affairs and Inclusion.

Dobkin, C., Gil, R., and Marion, J. (2010). Skipping class in college and exam performance: Evidence from a regression discontinuity classroom experiment. *Economics of Education Review*, 29(4):566–575.

Dobronyi, C. R., Oreopoulos, P., and Petronijevic, U. (2017). Goal setting, academic reminders, and college success: A large-scale field experiment. NBER Working Papers 23738, National Bureau of Economic Research, Inc.

Duchini, E. and Van Effenterre, C. (2022). School schedule and the gender pay gap. *Journal of Human Resources*.

Duckworth, A. L., Quirk, A., Gallop, R., Hoyle, R. H., Kelly, D. R., and Matthews, M. D. (2019). Cognitive and noncognitive predictors of success. *Proceedings of the National Academy of Sciences*, 116(47):23499–23504.

Education International (2023). Italy: Teachers strike for fair pay and job security amid austerity. Technical report, Education International.

Eisenberg, N., Cumberland, A., Spinrad, T. L., Fabes, R. A., Shepard, S. A., Reiser, M., Murphy, B. C., Losoya, S. H., and Guthrie, I. K. (2001). The relations of regulation and emotionality to children's externalizing and internalizing problem behavior. *Child Development*, pages 1112–1134.

Ellison, G. and Swanson, A. (2016). Do schools matter for high math achievement? evidence from the american mathematics competitions. *American Economic Review*, 106(6):1244–77.

Elsner, B. and Isphording, I. E. (2017). A big fish in a small pond: Ability rank and human capital investment. *Journal of Labor Economics*, 35(3):787–828.

Elsner, B. and Isphording, I. E. (2018). Rank, sex, drugs, and crime. *Journal of Human Resources*, 53(2):356–381.

Elsner, B., Isphording, I. E., and Zölitz, U. (2021). Achievement Rank Affects Performance and Major Choices in College. *The Economic Journal*, 131(640):3182–3206.

Eurydice - European Commission (2024). Conditions of service for teachers working in early childhood and school education - italy. Technical report, European Commission.

Fernández, R. (2007). Women, work, and culture. *Journal of the European Economic Association*, 5(2-3):305–332.

Fernández, R. and Fogli, A. (2009). Culture: An empirical investigation of beliefs, work, and fertility. *American Economic Journal: Macroeconomics*, 1(1):146–177.

Fernández-Villaverde, J., Laudati, D., Ohanian, L., and Quadrini, V. (2023). Accounting for the duality of the italian economy. Accessed: March 28, 2025.

Fogli, A. and Veldkamp, L. (2011). Nature or nurture? learning and female labor force participation. *Econometrica*, 79(4):1103–1138.

Fondazione Agnelli (2024). Eduscopio. Technical report, Fondazione Agnelli. Accessed: 2024-10-27.

Fortin, N. M. (2005). Gender role attitudes and the labour-market outcomes of women across oecd countries. *Oxford Review of Economic Policy*, 21(3):416–438.

Fröhlich, S. and Ruedin, D. (2023). Gender attitudes and political participation: What abortion preferences reveal about gender norms. *Politics, Groups, and Identities*. Forthcoming.

Gächter, S., Johnson, E., and Herrmann, A. (2022). Individual-level loss aversion in riskless and risky choices. *Theory and Decision*, 92.

Giuntella, O. (2022). The origins of italy's north-south divide. LSE Economic History Blog, accessed March 2025.

Goldberg, L. R. (1992). The Development of Markers for the Big-Five Factor Structure. *Psychological Assessment*, 4(1):26–42.

Goldin, C. and Katz, L. F. (2008). *The race between education and technology*. Harvard University Press.

Goodman, J. (2019). The labor of division: Returns to compulsory high school math coursework. *Journal of Labor Economics*, 37(4):1141–1182.

Goulas, S., Griselda, S., and Megalokonomou, R. (2023). Compulsory class attendance versus autonomy. *Journal of Economic Behavior and Organization*, 212:935–981.

Grossman, M. (2006). Education and nonmarket outcomes. In Hanushek, E. A. and Welch, F., editors, *Handbook of the Economics of Education*, volume 1, pages 577–633. Elsevier, Amsterdam.

Görlitz, K. and Gravert, C. (2018). The effects of a high school curriculum reform on university enrollment and the choice of college major. *Education Economics*, 26(3):321–336.

Hanushek, E. A. and Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature*, 46(3):607–668.

Heckman, J., Pinto, R., and Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, 103(6):2052–86.

Heckman, J. J. and Kautz, T. (2012). Hard evidence on soft skills. *Labour Economics*, 19(4):451–464. European Association of Labour Economists 23rd annual conference, Paphos, Cyprus, 22-24th September 2011.

Henry, G. T., Rubenstein, R., and Bugler, D. T. (2004). Is hope enough? impacts of receiving and losing merit-based financial aid. *Educational Policy*, 18(5):686–709.

Hershberger, S. L. (2005). *Factor Score Estimation.* John Wiley & Sons, Ltd.

Himmler, O., Jäckle, R., and Weinschenk, P. (2019). Soft commitments, reminders, and academic performance. *American Economic Journal: Applied Economics*, 11(2):114–42.

Hoxby, C. and Turner, S. (2013). Expanding college opportunities for high-achieving, low income students. *Stanford Institute for Economic Policy Research Discussion Paper.*

Iaria, A., Schwarz, C., and Waldinger, F. (2018). Frontier Knowledge and Scientific Production: Evidence from the Collapse of International Science*. *The Quarterly Journal of Economics*, 133(2):927–991.

Iloeje, S. O. and Meme, J. (1992). Rutter's Behaviour Scale (B2) for Children (Teacher's Scale): Validation and Standardization for Use on Nigerian Children. *Journal of Tropical Pediatrics*, 38(5):235–239.

Islam, A., Kwon, S., Masood, E., Prakash, N., Sabarwal, S., and Saraswat, D. (2020). When Goal-Setting Forges Ahead but Stops Short.

ISTAT (2018). Gender stereotypes and sexual violence in the italian population. Technical report, Italian National Institute of Statistics (ISTAT). Accessed: 2024-10-10.

ISTAT (2018). Internal mobility and international migrations. Technical report, Italian National Institute of Statistics (ISTAT).

Italian Parliament (1948). *Constitution of the Italian Republic*. Gazzetta Ufficiale, Rome, Italy. Accessed: 2024-10-10.

Jayachandran, S. (2015). The roots of gender inequality in developing countries. *Annual Review of Economics*, 7:63–88.

Jefferson, T. and Taplin, R. (2011). An investigation of the endowment effect using a factorial design. *Journal of Economic Psychology*, 32(6):599–907.

Jensen, R. (2010). The (perceived) returns to education and the demand for schooling. *Quarterly Journal of Economics*, 125(2):515–548.

Joensen, J. and Nielsen, H. (2009). Is there a causal effect of high school math on labor market outcomes? *Journal of Human Resources*, 44(1).

Joensen, J. and Nielsen, H. (2016). Mathematics and gender: Heterogeneity in causes and consequences. *Economic Journal*, 126(593):1129–1163.

Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1990). Experimental tests of the endowment effect and the coase theorem. *Journal of Political Economy*, 98(6):1325–1348.

Kahneman, D. and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2):263–291.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and psychological measurement*, 20(1):141–151.

Kaiser, J. P., Koch, A. K., and Nafziger, J. (2021). Self-Set Goals Are Effective Self-Regulation Tools – Despite Goal Revision.

Kaiser, U. and Mata, J. (2025). Persistent gender attitudes and women entrepreneurship.

Kapoor, S., Oosterveen, M., and Webbink, D. (2021). The price of forced attendance. *Journal of Applied Econometrics*, 36(2):209–227.

Kaur, S., Kremer, M., and Mullainathan, S. (2015). Self-control at work. *Journal of Political Economy*, 123(6):1227–1277.

Klein, J. M., Goña Salves, A., and Silva, C. F. (2009). The rutter children behaviour questionnaire for teachers: from psychometrics to norms, estimating caseness. *PsicoUSF*, 14:157 – 165.

Koch, A. K. and Nafziger, J. (2011). Self-regulation through goal setting*. *The Scandinavian Journal of Economics*, 113(1):212–227.

La Repubblica (2010). Rigorosi ma non irragionevoli le nuove norme non c'entrano nulla.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, 112(2):443–477.

Lavecchia, A. M., Liu, H., and Oreopoulos, P. (2016). Behavioral economics of education: Progress and possibilities. In *Handbook of the Economics of Education*, volume 5, pages 1–74. Elsevier.

Laverde, M. (2022). Distance to schools and equal access in school choice systems. Technical report, Boston College.

Lavy, V. (2019). Gender differences in willingness to guess and the implications for test scores. *Journal of Political Economy*, 127(6):2754–2790.

Lawlor, D. A., Smith, G. D., and Ebrahim, S. (2006). Season of birth and childhood intelligence: findings from the aberdeen children of the 1950s cohort study. *British Journal of Educational Psychology*, 76(3):481–499.

Leuven, E., Oosterbeek, H., and van der Klaauw, B. (2010). The effect of financial rewards on students' achievement: Evidence from a randomized experiment. *Journal of the European Economic Association*, 8(6):1243–1265.

Levine, P. and Zimmerman, D. (1995). The benefit of additional high-school math and science classes for young men and women. *Journal of Business and Economic Statistics*, 13(2):137–49.

Liu, Y. (2019). Effort provision under present bias: Optimal goal-setting as a commitment device. Technical report, Humboldt University of Berlin.

Mandic, S., Sandretto, S., Hopkins, D., Wilson, G., Kidd, G., and García Bengoechea, E. (2023). School choice, distance to school and travel to school patterns among adolescents. *Journal of Transport & Health*, 33:101704.

Marsh, H. W. and Parker, J. W. (1984). Determinants of student self-concept: Is it better to be a relatively large fish in a small pond even if you don't learn to swim as well? *Journal of Personality and Social Psychology*, 47(1):213–231.

McGee, R., Williams, S., Bradshaw, J., Chapel, J. L., Robins, A., and Silva, P. A. (1985). The rutter scale for completion by teachers: Factor structure and relationships with cognitive abilities and family adversity for a sample of new zealand children. *Journal of Child Psychology and Psychiatry*, 26(5):727–739.

McKenzie, D. (2012). Beyond baseline and follow-up: The case for more T in experiments. *Journal of Development Economics*, 99(2):210–221.

Megalokonomou, R. and Zhang, Y. (2024). How good am i? effects and mechanisms behind salient rank. *European Economic Review*, 170:104870.

Meluzzi, F. (2024). The college melting pot: Peers, culture and women's job search.

Morando, G. (2020). Mathematics specialization at high school and undergraduate degree choice: Evidence from england. *Educational Evaluation and Policy Analysis*, 0(0):01623737241255348.

Moretti, E. (2004). Workers' education, spillovers, and productivity: Evidence from plant-level production functions. *American Economic Review*, 94(3):656–690.

Moser, P., Voena, A., and Waldinger, F. (2014). German jewish Émigrés and us invention. *American Economic Review*, 104(10):3222–55.

Murphy, R. and Weinhardt, F. (2020). Top of the Class: The Importance of Ordinal Rank. *The Review of Economic Studies*, 87(6):2777–2826.

Narusyte, J., Ropponen, A., Alexanderson, K., and Svedberg, P. (2017). Internalizing and externalizing problems in childhood and adolescence as predictors of work incapacity in young adulthood. *Social Psychiatry and Psychiatric Epidemiology*, 52(9):1159–1168.

Nicholson, J. R. (2020). New digital economy estimates.

Norris, P., Inglehart, R., Tworzecki, H., et al. (2021). Measuring populism worldwide. *West European Politics*, 44(4):752–790.

O'Donoghue, T. and Rabin, M. (1999). Doing it now or later. *American Economic Review*, 89(1):103–124.

Office for National Statistics (2003). The effects of taxes and benefits on household income, 2001/02. https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth.

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, 44(4):443–460.

Oreopoulos, P. and Petronijevic, U. (2013). Making college worth it: A review of the returns to higher education. *The Future of Children*, 23(1):41–65.

Oreopoulos, P. and Salvanes, K. G. (2011). Priceless: The nonpecuniary benefits of schooling. *Journal of Economic Perspectives*, 25(1):159–184.

Pagani, L., Comi, S., and Origo, F. (2021). The Effect of School Rank on Personality Traits. *Journal of Human Resources*, 56(4):1187–1225.

Patterson, R. W. (2018). Can behavioral tools improve online student outcomes? Experimental evidence from a massive open online course. *Journal of Economic Behavior & Organization*, 153(C):293–321.

Porter, C. and Serra, D. (2020). Gender differences in the choice of major: The importance of female role models. *American Economic Journal: Applied Economics*, 12(3):226–54.

Reuben, E., Wiswall, M., and Zafar, B. (2017). Preferences and biases in educational choices and labour market expectations: Shrinking the black box of gender. *The Economic Journal*, 127(604):2153–2186.

Romer, D. (1993). Do students go to class? should they? *Journal of Economic Perspectives*, 7(3):167–174.

Rudd, B., Patel, K., Levy, N., and Dhatariya, K. (2013). A survey of the implementation of the nhs diabetes guidelines for management of diabetic ketoacidosis in the intensive care units of the east of england. *Journal of the Intensive Care Society*, 14(1):60–64.

Rutter, M. (1967). A children's behaviour questionnaire for completion by teachers: Preliminary findings. *Journal of Child Psychology and Psychiatry*, 8(1):1–11.

Sacerdote, B. (2011). Peer effects in education: How might they work, how big are they, and how much do we know thus far? *Handbook of the Economics of Education*, 3:681–703.

Salesforce (2020). The impact of equality and values driven business.

Santrock, J. W. and Feldman, R. S. (2020). *Child Development*. McGraw-Hill Education, 19th edition.

Schonell, F. J. and Adams, R. H. (1940). *The essential intelligence test*. Edinburgh, London: Oliver & Boyd.

Sorrenti, G., Zölitz, U., Ribeaud, D., and Eisner, M. (2024). The Causal Impact of Socio-Emotional Skills Training on Educational Success. *The Review of Economic Studies*, page rdae018.

Stinebrickner, R. and Stinebrickner, T. R. (2008). The effect of credit constraints on the college drop-out decision: A direct approach using a new panel study. *The American Economic Review*, 98(5):2163–2184.

Testo Unico Istruzione (1994). *Decreto Legislativo 1994, 297.* Gazzetta Ufficiale.

Thaler, R. and Benartzi, S. (2004). Save more tomorrow™: Using behavioral economics to increase employee saving. *Journal of Political Economy*, 112(S1):S164–S187.

Unione Matematica Italiana (2003). Il piano nazionale per l'informatica. *Bollettino dell'Unione Matematica Italiana*, 6-A(3):441–461. La Matematica nella Società e nella Cultura.

van Lent, M. (2019). Goal setting, information, and goal revision: A field experiment. *German Economic Review*, 20(4):e949–e972.

van Lent, M. and Souverijn, M. (2020). Goal setting and raising the bar: A field experiment. *Journal of Behavioral and Experimental Economics (formerly The Journal of Socio-Economics)*, 87(C):S2214804319306184.

Waldinger, F. (2016). Bombs, brains, and science: The role of human and physical capital for the creation of scientific knowledge. *The Review of Economics and Statistics*, 98(5):811–831.

Wertenbroch, K. (1998). Consumption self-control by rationing purchase quantities of virtue and vice. *Marketing Science*, 17(4):317–337.

Wiswall, M. and Zafar, B. (2017). Preference for the Workplace, Investment in Human Capital, and Gender*. *The Quarterly Journal of Economics*, 133(1):457–507.

Zafar, B. (2013). College major choice and the gender gap. *Journal of Human Resources*, 48(3):545–595.

Appendices

Appendix for Chapter 1: Boosting Attendance through Goal Setting: Evidence from a Randomized Experiment

A.1 Placebo Check

In this appendix, we present the results of the placebo checks conducted to assess baseline balance between the treatment and control groups. By estimating treatment effects on pre-intervention outcomes — when no treatment had yet occurred — we can verify that any post-treatment differences are unlikely to be driven by pre-existing trends or imbalances. The absence of statistically significant differences across attendance, academic performance, well-being, and time use provides additional support for the validity of our experimental design.

We check that no treatment effect on attendance (to events, lectures, and classes) exists in the term preceding the experiment, estimating Equation 1.1. Again, the dependent variable is attendance in the Autumn Term of 2016. The results in columns (1) to (6) in Table A.1 are reassuring, as future-treated students exhibit no statistically significant difference from those in the control group.

Similarly, we repeat the check for outcomes realized at the academic-year level rather than at the term level, estimating Equation 1.2. Because data are not at term-by-individual rather than at the week-by-individual level, accounting for week fixed effects is no longer necessary. Columns (7) to (9) in Table A.1 show that no outcome related to academic performance is statistically different for students in the treatment or control group. We cannot control for the baseline of the three academic performance variables, as we lack a baseline to do so (they are measured only once per academic year).

Finally, in Table A.2 we show that there are no differences between the treatment and control group in terms of self-reported well-being and allocation of time during a typical week, during the academic year preceding the goal-setting experiment.

A.2 Description and Validation of Planning Efficacy and Loss Aversion Measures

To interpret the effects of our intervention through the lens of the theoretical model presented in Koch and Nafziger, 2011, it is essential to capture the behavioral traits the model identifies as central — namely, present bias and loss aversion. These traits shape how individuals respond to goals, particularly in contexts involving effort and self-control. In this appendix, we describe how we construct and validate measures of these traits.

A.2.1 Planning Efficacy as Present Bias in the Effort Domain

We provide supporting evidence for the use of planning efficacy as an effort-based proxy for present bias in the context of our study. As discussed in the main text, present bias is a key behavioral trait in our theoretical framework. Given that the behavior targeted by the intervention—lecture attendance—requires sustained effort rather than financial decision-making, we seek a measure that captures time inconsistency in the effort domain.

To this end, we construct a planning efficacy index based on four self-reported items, capturing students' tendencies to plan ahead and manage their time effectively. To validate this index, we examine its relationship with both prior academic behavior and proxies for forward-looking effort. Table A.3 reports correlations and OLS estimates (conditioning on stratification variables) between planning efficacy - measured in the online survey performed in Autumn 2016, one term before the goal-setting experiment - and a set of relevant outcomes. We find that students with higher planning efficacy are significantly more likely to have attended lectures in the previous term and to have achieved higher grades, consistent with reduced procrastination and stronger academic self-regulation. Moreover, they are more likely to have booked their lab session earlier and attended it sooner after signing up—behaviors requiring advanced planning. The association between planning efficacy and these effort-based outcomes is both statistically significant and robust to controls.

Figure A.1 displays the distribution of the planning efficacy index for the full sample and by gender. The overall distribution appears fairly symmetrical around its mean (4.43 on a 1–7 scale), with a slight skew toward higher values, indicating that many students report relatively strong planning habits. The gender breakdown reveals that female students tend to score higher on planning efficacy than male students, consistent with the broader literature on gender differences in self-regulatory behavior.

In contrast, we find no meaningful relationship between planning efficacy and a standard measure of monetary present bias.[1] This supports the view that time preferences are domain-specific (Augenblick et al. 2015), and reinforces the relevance of our chosen measure in the context of academic effort. Together, these findings validate planning efficacy as a conceptually and empirically grounded indicator of (non-)present biasedness in the domain of interest.

[1] We derive a binary marker for present biasedness, defined as having a higher discount rate with respect to immediate trade-offs than those taking place in the future. This was measured in the lab session of the survey held in January 2016, during the Spring term of the academic year 2015/2016. We elicited the "immediate" interest rate required to persuade a participant to wait four months rather than receive a payment immediately, and the "future" interest rate required to persuade a participant to wait five months rather than receive a payment in one month. Those requiring a higher immediate than future interest rate are present-biased. Their preferences are time-inconsistent in that, arriving at one month in the future, they would require a higher interest rate than they will receive, in order to be persuaded to wait. During the session, the choice is between £15 sooner, and up to £17.50 later, both paid with certainty. By this measure, 18% are present-biased.

A.2.2 Endowment Effect as Loss Aversion

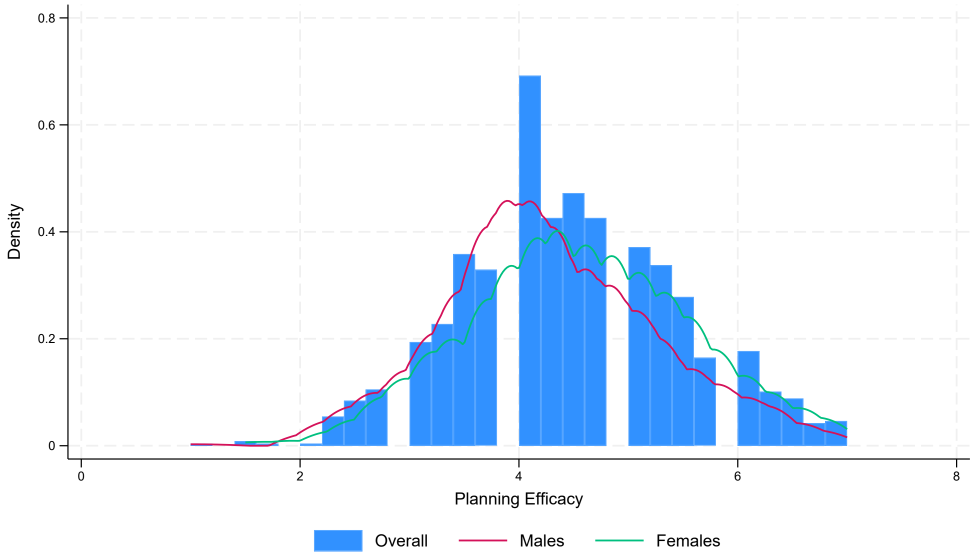
Loss aversion—defined as the tendency to weigh losses more heavily than gains—is a central feature in models of goal-setting behavior (Koch and Nafziger, 2011). According to the theory, individuals who are loss averse should be especially responsive to goal-setting interventions, since failing to meet a self-imposed goal creates a psychologically salient loss.

We measure loss aversion using a version of the classic mug experiment (Kahneman et al., 1990), in which the endowment effect — the difference between willingness to accept (WTA) and willingness to pay (WTP) — is used as a proxy. Figure A.2 shows the distribution of this measure for the overall sample and by gender. In line with expectations, over 85% of students display a strictly positive endowment effect, classifying them as loss averse. The distributions show no meaningful difference between male and female students.

The magnitude of the endowment effect in our experiment aligns well with prior studies. The mug used had a market value of £3.50, and we observe a mean absolute endowment effect of £1.80. This is broadly consistent with Colucci et al. (2024), who report a $2.65 average gap in a similar design. Using the transformed endowment effect scale introduced by Jefferson and Taplin (2011), we find a mean of 0.5 and a standard deviation of 0.5 in our sample. The average WTA/WTP ratio in our data is 2.70 (median: 1.71), closely matching estimates from Gächter et al. (2022), who report a ratio of 2.12 in a similar context.

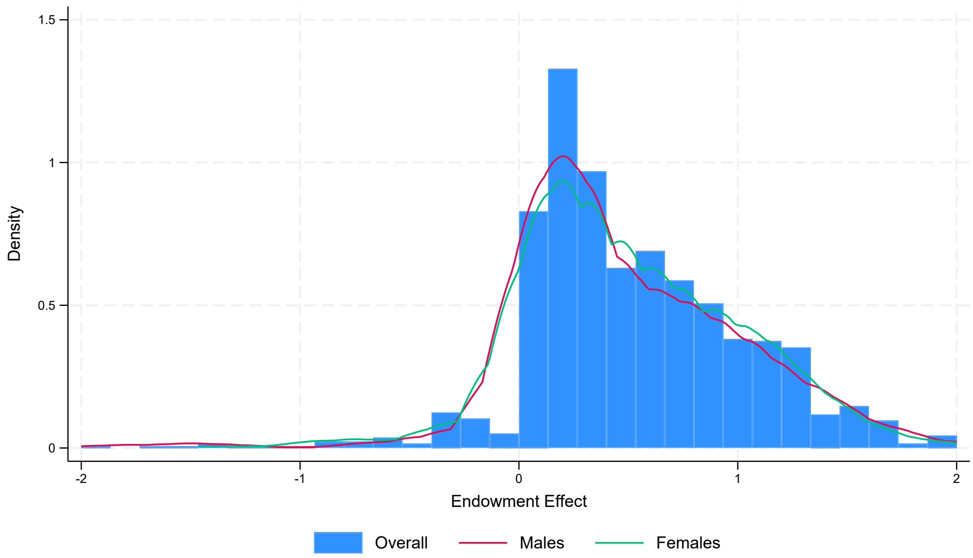
To validate this measure, we repeat the analysis conducted for planning efficacy. Specifically, we test whether our loss aversion measure correlates with pre-treatment outcomes such as prior attendance, academic achievement, and forward-looking behaviors (e.g., early lab sign-up). As shown in Table A.4, we find no significant relationship between loss aversion and any of these outcomes. While this result may suggest that the effect of loss aversion operates primarily through its interaction with the intervention, rather than observable baseline behaviors, it also reinforces the domain-specific nature of behavioral traits and the importance of contextual measurement.

Figure A.1: Planning Efficacy Distribution



Notes: The following graph showcases the distribution of planning efficacy, as measured in the online wave of the survey held in November 2016, during the Autumn term of the academic year 2016/2017. We show a histogram of the distribution for the entire sample, as well as the density for males (in red) and females (in green). Planning efficacy is measured by averaging responses to four Likert-scale items (1 = "Strongly Disagree", 7 = "Strongly Agree"). The items are: "I usually do my work assignment the day before it is due"; "I usually keep track of my work assignment on a schedule or planner"; "I do not need to plan ahead to get good marks"; "I often underestimate the time that will be required to finish a project". Responses are recoded where necessary so that higher values reflect greater planning ability. We measure the planning efficacy of all students participating in the November 2016 online session of the study, 920 in total (we exclude from the distribution those attending during the 4th week of the goal-setting experiment).

Figure A.2: Endowment Effect Distribution



Notes: The following graph showcases the distribution of the endowment effect, as measured in the lab experiment held in January 2017, during the Spring term of the academic year 2016/2017. We show a histogram of the distribution for the entire sample, as well as the density for males (in red) and females (in green). The endowment effect is measured by having the students find a mug on their desk upon entering the lab. At the beginning of the session, they are either asked how much they would pay to buy the mug or, after explaining the mug is now their property, for which price they would sell it. At the end of the session, the roles would be reversed, to measure the difference in the monetary value placed on the mug when buying it compared to when selling it. The index is then transformed following Jefferson and Taplin, 2011 so that it is bounded between -2 and 2. We measure the endowment effect for all students participating in the goal-setting experiment at the lab, in January 2017, during the Spring term of the academic year 2016/2017, 1,022 in total (we exclude from the distribution those attending during the 4th week of the goal-setting experiment).

Table A.1: Placebo Check: Attendance and Academic Performance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
			Attendance to:				Academic Performance:		
Variables	Events		Lectures		Classes		GPA	SPM	SGG
Treatment	0.695	0.766	1.034	1.352	0.538	0.340	0.525	-0.399	2.556
	(0.945)	(0.968)	(1.091)	(1.144)	(1.260)	(1.269)	(0.652)	(0.719)	(1.627)
Mean of the Outcome	61.24	58.70	67.75	65.14	53.53	51.41	61.81	98.09	62.60
Individuals	1,013	1,013	991	991	1,012	1,012	1,015	1,019	1,022
Observations	8,996	6,984	8,706	6,754	8,451	6,590	1,015	1,019	1,022
Stratifying Variables	X	X	X	X	X	X	X	X	X
Baseline	X	X	X	X	X	X			
Week Fixed Effects	X	X	X	X	X	X			
Period of Interest	Post W16	Post W19	Post W16	Post W19	Post W16	Post W19	Full Year		

Notes: We estimate the placebo-treatment effect, which is the impact of the treatment in the year before the goal-setting intervention took place. The dependent variables we use - all expressed on a scale from 0 to 100 - are: the percentage of academic events (columns (1) and (2)), lectures (columns (3) and (4)), or classes (columns (5) and (6)) attended by the students, Grade Point Average in the first academic year (7), the Share of Passed Modules (SPM) (8), the Share of Good Grades (SGG) (9). Attendance is collected through the swipe card system in place at the university, while academic performance is taken from students' administrative records. Academic events include both lectures and classes. We estimate the placebo-treatment effect on attendance using Equation 1.1 and the placebo-treatment effect on academic performance using Equation 1.2. We include controls for week fixed effects, the stratifying variables used for the randomization process (sex, age, socioeconomic status, tariff quintile, and department), and the dependent variable's baseline - which is the percentage of events/lectures/classes attended during the previous term. The baseline is not available for those variables related to students' academic performance. Standard errors are clustered at the individual level. ***: p <0.01, **: p <0.05, *: p <0.1.

Table A.2: Placebo Check: Academic Performance and Well-Being

Variables	(1) Well-Being: Satisfaction	(2) Interest	(3) Study	(4) Time Allocation: Sleep	(5) Work	(6) Exercise
Treatment	-0.583	0.515	-0.044	-0.139	0.002	-0.168
	(1.125)	(1.093)	(0.507)	(0.185)	(0.106)	(0.458)
Mean of the Outcome	69.28	73.01	3.00	7.94	0.50	0.63
Individuals	762	757	727	765	765	765
Observations	762	757	727	765	765	765
Stratifying Variables	X	X	X	X	X	X
Baseline	X	X	X	X	X	X

Notes: We estimate the placebo-treatment effect, which is the impact of the treatment in the year before the goal-setting intervention took place. The dependent variables we use are the following: Satisfaction with life at the University (1), and Interest in the Field of Study (2) - expressed on a scale from 0 to 100 - as well as students' allocation of studying time (3), sleeping time (4), work time (5), and exercise time (6) during a typical day of the week - expressed in hours per day. These variables are collected during the first two online sessions held in the academic year 2015/2016, in November 2015 and March 2016. We estimate the treatment effect using Equation 1.2. We include controls for the stratifying variables used for the randomization process (sex, age, socioeconomic status, tariff quintile, and department) and the dependent variable's baseline. Standard errors are clustered at the individual level. ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$.

Table A.3: Association of Planning Efficacy and Procrastination

Variables	(1) Planning Efficacy Raw Correlation	(2) OLS	(3) N
Year 2 Attendance: Autumn Term (lectures)	0.225	0.011*** (0.002)	908
First Year Marks	0.150	0.013*** (0.003)	900
Log of days: initial sign-up to initial appointment	0.102	0.131*** (0.047)	882
Log of days: initial sign-up to eventual attendance	0.064	0.084* (0.046)	883
Missed a Session	-0.036	-0.094 (0.105)	898
Monetary present bias	0.015	-0.020 (0.096)	749
Stratifying Variables		X	

Notes: We estimate the association between planning efficacy and a set of variables characterizing the students in our sample. Planning efficacy is measured in the survey session held in November 2016, during the Autumn term of the academic year 2016/2017. The characteristics we include are the following: percentage of lectures attended during the Autumn term of the academic year 2016/2017, the student's GPA during their first year at the university, a set of variables describing students' behavior in signing up to participate to the lab session where the goal-setting experiment took place, in January 2017, during the Spring term of the academic year 2016/2017 - the logarithmic transformation of the number of days taken to sign up for an initial appointment and of the number of days between the sign-up and attendance, and finally whether they missed a session they signed up for - and students' monetary present bias, measured in the lab session of January 2016, during the Spring term of the academic year 2015/2016. We present the raw correlation between the variables in column (1), and the OLS estimate of the relationship, controlling for the stratifying variables used in the randomization process (sex, age, socioeconomic status, tariff quintile, and department). Standard errors are clustered at the individual level. ***: $p<0.01$, **: $p<0.05$, *: $p<0.1$.

Table A.4: Association of Endowment Effect and Procrastination

Variables	(1) Endowment Effect Raw Correlation	(2) OLS	(3) N
Year 2 Attendance: Autumn Term (lectures)	-0.867	0.157 (1.710)	1,054
First Year Marks	1.153	1.049 (0.891)	1,069
Log of days: initial sign up to initial appointment	0.016	0.014 (0.071)	1,019
Log of days initial sign up to eventual attendance	-0.029	-0.013 (0.074)	1,019
Missed a Session	-0.013	0.004 (0.029)	1,019
Monetary present bias	0.023	0.024 (0.043)	805
Stratifying Variables		X	

Notes: We estimate the association between the endowment effect and a set of variables characterizing the students in our sample. The endowment effect is elicited in the goal-setting experiment held in January 2017, during the Spring term of the academic year 2016/2017. The characteristics we include are the following: percentage of lectures attended during the Autumn term of the academic year 2016/2017, the student's GPA during their first year at the university, a set of variables describing students' behavior in signing up to participate to the lab session where the goal-setting experiment took place, in January 2017, during the Spring term of the academic year 2016/2017 - the logarithmic transformation of the number of days taken to sign up for an initial appointment and of the number of days between the sign-up and attendance, and finally whether they missed a session they signed up for - and students' monetary present bias, measured in the lab session of January 2016, during the Spring term of the academic year 2015/2016. We present the raw correlation between the variables in column (1), and the OLS estimate of the relationship, controlling for the stratifying variables used in the randomization process (sex, age, socioeconomic status, tariff quintile, and department). Standard errors are clustered at the individual level. ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$.

Appendix for Chapter 2: Beyond Test Scores: the Rank Effect and Non-Cognitive Skills

B.1 Randomization of Participation to the Family Survey and 2001 Follow-Up

In this appendix, we present additional checks to support the validity of our identification strategy for the estimation of the rank effect on parental investment and long-term outcomes. First, we show that academic rank does not predict the probability of participating in the Family Survey, and that observable child characteristics are balanced between participants and non-participants, confirming the effectiveness of the original randomization. We then replicate this exercise for the 2001 follow-up survey, demonstrating that rank is not associated with survey participation. However, individual baseline characteristics are unbalanced across respondents and non-respondents.

In Table B.1, we test whether academic rank predicts participation in the Family Survey or the 2001 follow-up survey. Across both outcomes, we find no evidence that rank influences the probability of participation. This result is reassuring, as it suggests that our estimates of the rank effect on parental investment and long-term outcomes are not biased by differential attrition related to students' relative standing within their school-cohort group. In other words, even though participation in these surveys is selective, that selection appears orthogonal to the variation we exploit for identification.

To further assess the nature of selection into these survey subsamples, we regress a range of baseline individual characteristics on the probability of participating in either the Family Survey or the 2001 follow-up, as reported in Table B.2. These characteristics include cognitive skills (measured by the age 9 test), non-cognitive skills (externalizing and internalizing scores), and other observables such as gender, socioeconomic background, height and weight residuals from the first medical exam, birth weight, and number of siblings.[1] The results show that participation in the Family Survey is effectively randomized: the only statistically significant coefficient is for number of siblings, and even that is marginal (10% level) and economically small — 1% of a unit of standard deviation.

In contrast, participation in the 2001 follow-up survey is clearly non-randomized. Respondents tend to have stronger cognitive and non-cognitive skills, are more likely to be female, and come from higher-SES families. There are also differences in anthropometric measures and family size, though the latter remains of limited economic relevance. These patterns suggest that the long-term sample is positively selected, with higher-performing children more likely to respond. While this does not undermine the internal validity of our estimates — since rank itself does not predict participation — it implies that the external validity of our long-term results may be limited to a more advantaged segment of the original sample. We limit these concerns by controlling for these individual characteristics in our main specifications.

[1] Specifically: a dummy for the child being female, a high-SES indicator (based on the parent's occupation), residualized height and weight from the first medical exam, birth weight in pounds, and the number of siblings as of December 1962.

Table B.1: Rank Effect on Family Survey and 2001 Follow-Up Survey Participation

	(1)	(2)	(3)	(4)
	Participation to the:			
Outcome Variables	Family Survey		2001 Follow-Up	
Percentile Rank	0.012	0.013	0.046	0.045
	(0.044)	(0.044)	(0.056)	(0.056)
Mean of the Outcome	0.18	0.18	0.59	0.59
SD of the Outcome	0.39	0.39	0.49	0.49
Observations	9,715	9,715	9,715	9,715
School Fixed Effects	X	X	X	X
Cohort Fixed Effects	X	X	X	X
Cognitive Skills	X	X	X	X
Cognitive Skills Squared	X	X	X	X
Mean of Peer Cognitive Skills	X	X	X	X
SD of Peer Cognitive Skills	X	X	X	X
Sex	-	X	-	X
Socioeconomic Status	-	X	-	X
Number of Siblings	-	X	-	X
Month of Birth	-	X	-	X

Notes: We estimate the relationship percentile rank within the school-cohort group and the probability of participating to the Family Survey (columns (1) and (2)) and to the 2001 Follow-Up Survey (columns (3) and (4)). Percentile rank is established using our baseline measure of cognitive skills, the outcome of the Age 9 Test. Our sample consists of 10 cohorts of children who attended primary school in Aberdeen in December 1962, who were born between October 1950 and October 1955. We control for: a categorical variable taking a different value for each school in the sample; a categorical variable taking a different value for each cohort in the sample; a quadratic polynomial of child cognitive skills (based on the outcome of the Age 9 Test); the mean and standard deviation of the cognitive skills of the peers of the students (based on the outcome of the Age 9 Test); a categorical variable taking value 1 if the child is a girl, and 0 if he is a boy; a categorical variable taking value 1 if the child belongs to a family of high socioeconomic status (defined based on the father's occupation), and 0 otherwise; a categorical variable capturing the specific month of birth of the child; and the number of siblings of the child. Standard errors are clustered at the school-cohort level.

Table B.2: Balancing Exercise: Probability of Participating in the Family Survey and the 2001 Follow-Up Survey

Independent Variables	Cognitive Skills	Externalizing	Internalizing	Woman	High SES	Height	Weight	Birth Weight	Siblings
Relationship between individual characteristics and the probability of participating in the Family Survey									
Individual Characteristics	0.001	0.007	-0.003	0.010	-0.016	0.005	0.002	0.003	-0.010*
	(0.004)	(0.004)	(0.005)	(0.008)	(0.014)	(0.004)	(0.004)	(0.004)	(0.005)
Relationship between individual characteristics and the probability of participating in the 2001 Follow-Up Survey									
Individual Characteristics	0.085***	0.060***	0.007***	0.087***	-0.069***	0.038***	0.025***	0.005	-0.019***
	(0.005)	(0.006)	(0.006)	(0.010)	(0.017)	(0.005)	(0.005)	(0.005)	(0.007)
Observations	9,717	6,790	6,790	9,970	9,970	9,666	9,668	9,970	9,970

Notes: We estimate the unconditional relationship between different individual characteristics and the probability of participating in the Family Survey and the 2001 Follow-Up Survey. These characteristics are: the student cognitive skills, measured by the standardized score in the Age 9 Test; the standardized measures of externalizing and internalizing skills, estimated using common factor analysis on the 26 items of the Rutter Questionnaire for Teachers, completed in March 1964; the student probability of being a girl, the student probability of coming from an advantaged socioeconomic background (based on the father's occupation), the student height and weight at the time of their first medical exam, the student birth weight (lbs), and the student number of siblings. We include all 10 cohorts of children who attended primary school in Aberdeen in December 1962, who were born between October 1950 and October 1955. Standard errors are clustered at the school-cohort-group level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix for Chapter 3: Breaking Barriers or Reinforcing Gaps? Scientific Education and Gendered Academic Choices

C.1 Robustness Exercises

To ensure that the gender-specific estimates are not driven by specific modeling choices, I perform a set of robustness checks altering key control variables and sample restrictions. While my baseline model includes controls for municipality type (rural, urban, or metropolitan) and school size terciles, I test whether the results are robust to alternative classifications — specifically, school-cohort size quintiles and terciles of municipality population. Additionally, I examine whether excluding students who repeated a year in high school affects the findings.

C.1.1 Alternative Controls for School-Cohort Size and Municipality Population

The results of these alternative specifications are reported in columns (3) to (6) of Table C.2. Across all alternative categorizations, the estimated effects of attending a *PNI* school on STEM enrollment remain consistent with those obtained from the baseline specification shown in columns (1) and (2). Both the size and significance of the coefficients are stable, suggesting that the original findings are not sensitive to how school or municipality characteristics are defined. These robustness checks help validate that the estimated gender heterogeneity in treatment effects is not an artifact of the specific control variables included.

C.1.2 Including Students Who Repeated One Year of High School

In a second robustness exercise, presented in Table C.2, I re-estimate the model including students who repeated one or more years in high school. The results confirm that the impact on STEM enrollment remains effectively unchanged for both boys and girls. Interestingly, the inclusion of these students increases the precision of the estimates on dropout, which become statistically significant for both genders. This result may reflect a broader effect of the *PNI* program on academic preparation during high school, potentially equipping even less academically successful students with the skills needed to persist in higher education.

C.2 Gender Attitudes and the 1981 Referendum on Abortion

In this section, I provide a more extensive discussion to support the choice of using a historically rooted proxy for gender norms: the outcome of the 1981 abortion referendum in Italy. In that referendum, citizens were asked to vote on a conservative proposal that sought to significantly restrict access to abortion, effectively undoing many of the rights granted by a 1978 law. While the proposal was ultimately rejected at the national level, participation was high, and the results revealed substantial variation in attitudes. Support for the policy ranged widely across municipalities, from less than 4% in the more progressive to over 80% in the more conservative.

Support for restrictive abortion laws — like the one proposed in the 1981 referendum — reflects a broader endorsement of traditional gender roles, particularly those that assign women a primary role in the private sphere as caregivers and mothers. Such views are often associated with resistance to women's full participation in public life, including both the labor market and political institutions.[1]

Municipalities with higher support for the conservative proposal are likely to have harbored more traditional attitudes toward gender roles at the time — attitudes that may continue to shape local norms and behaviors today through cultural persistence and intergenerational transmission. A large body of research shows that cultural traits, including gender norms, are remarkably persistent. Becker et al. (2016) show that long-gone institutional environments continue to influence present-day civic attitudes and trust in public administration. On gender-specific dimensions, Alesina et al. (2013) find that regions with historically rigid gender divisions of labor, rooted in practices as far back as the Neolithic era, exhibit lower gender equality today. Similarly, Fernández (2007) documents the intergenerational transmission of gender norms, emphasizing their stability over time and their enduring influence on female labor market participation and educational choices. Stronger evidence comes from two recent studies that apply a similar approach to mine (Kaiser and Mata, 2025; Arni et al., 2024). Both examine the 1981 Swiss referendum, in which voters were asked to confirm a law granting equal rights to men and women in family, work, and social life. The level of support for the law is shown to predict both the likelihood of men applying to traditionally male-dominated occupations and the ratio of female-to-male-founded startups across regions more than 40 years later.

On the other hand, the literature has made extensive use of indicators based on labor market outcomes — most notably, the ratio between women's and men's labor force participation — to proxy for gender norms (Meluzzi, 2024; Jayachandran, 2015; Fogli and Veldkamp, 2011; Fernández and Fogli, 2009; Fortin, 2005). While informative, these measures can easily conflate cultural and structural economic factors, making it difficult to disentangle norms from labor market conditions or policy constraints. However, using

[1] Several studies link opposition to abortion rights with broader adherence to traditional gender roles. Deckman and McTague (2015) show that support for restrictive abortion policies correlates with conservative views about women's participation in the workforce and politics in the United States. Similarly, Norris et al. (2021) argue that abortion attitudes are a reliable indicator of value systems that resist gender equality. Most directly, Fröhlich and Ruedin (2023) provide cross-national evidence that abortion preferences are not only shaped by individual ideology, but also reflect deeply rooted cultural gender norms that affect both public opinion and women's political representation.

local preferences for the suppression of the abortion law also has potential drawbacks. Attitudes may have evolved since 1981, and referenda often reflect issue-specific beliefs, which may not perfectly align with broader views on gender roles. Moreover, the literature that exploits public votes to proxy for prevailing gender norms is still relatively underdeveloped, with only two studies set in Switzerland adopting a similar strategy to mine (Kaiser and Mata, 2025; Arni et al., 2024).

Still, in the context of my study, the referendum-based measure offers clear advantages. Since I am examining educational choices — especially the decision to pursue a STEM degree — using labor force participation rates as a proxy of gender attitudes presents a risk of conflating cultural attitudes with labor market incentives. A high female-to-male labor force participation ratio might reflect more progressive gender norms, but it might also be the result of economic necessity, job availability, or public childcare provision. As such, relying on labor market indicators to proxy cultural norms may blur the distinction between norms as constraints and economic conditions as drivers of behavior. By contrast, the referendum offers a more stable and meaningful measure of the cultural attitudes that shape educational trajectories. Although it took place roughly a decade before the students in my sample were born, it captures the normative environment in which their parents and local institutions were socialized — and which likely continued to influence expectations around gender roles during the students' childhood. Moreover, since high school choice in Italy is highly localized — students typically attend schools close to home — young people are especially likely to be exposed to the prevailing cultural norms of their municipality of residence. Referendum outcomes, being fixed in time and tied to place, provide a geographically and historically grounded proxy for those norms, avoiding the confounding influence of more mobile, contemporary labor market indicators.

In Table C.3, I provide evidence on the validity of using the 1981 referendum results as a proxy for prevailing gender norms by estimating their relationship with women's participation in local labor markets and political institutions. Specifically, I examine three municipality-level indicators as of 2011: the ratio of female to male labor force participation, obtained from the Population Census; and the share of women among elected officials in city councils and executive committees, sourced from the Ministry of the Interior's Registry of Local and Regional Administrators.[2] All three outcomes display a strong and statistically significant negative association with support for the restriction of abortion rights.

The size of the effects is economically meaningful. A one-unit increase in support for the conservative proposal — which ranges from 0 to 1, with a mean of 0.34 and standard deviation of 0.112 — is associated with a 9.6 percentage point decline in the labor force participation ratio (13% of the mean, over one standard deviation), and a 3.3 to 5.9 percentage point reduction in the share of women in city councils. The strongest relationship emerges with women's representation in executive committees, where the association reaches 9.3 percentage points (over half a standard deviation). While the coefficient for labor force participation remains stable when adding controls for province fixed effects and municipality type (rural, urban, metropolitan), the estimates for women in political office become stronger, likely reflecting the fact that gender representation in politics is more sensitive to institutional and structural characteristics across municipal contexts. These results lend empirical support to interpreting the referendum outcome as a valid and historically grounded proxy for deeply rooted gender norms. They echo earlier findings in the literature documenting the intergenerational persistence of cultural attitudes, including those shaping women's roles in labor markets and political life (Alesina et al., 2013; Fernández, 2007).

C.3 Breaking Down Results for Gender Norms

In this section, I provide a more elaborate breakdown of the variation behind the coefficients estimated using 2SLS, for girls in the 4th Quartile of Conservative Gender Norms - in other words, living in those municipalities that had the highest share of votes to restrict access to abortion during the 1981 referendum.

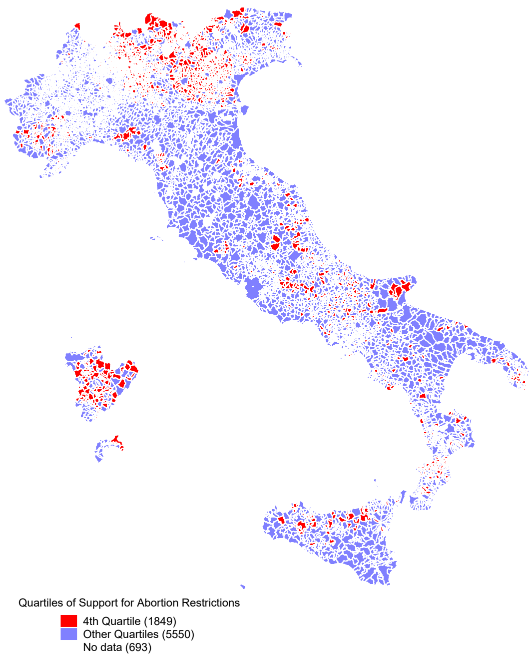
[2] In Italian municipalities, the *consiglio comunale* (city council) is the elected legislative body responsible for approving budgets, regulations, and local policies. The *giunta comunale* (executive committee), by contrast, is appointed by the mayor and is responsible for implementing council decisions and managing daily administrative functions.

Figure C.1 maps the municipalities in the top quartile of conservative gender norms across the national territory. A clear concentration emerges in the North-Eastern regions, particularly in *Veneto*, *Trentino Alto-Adige*, and the eastern part of *Lombardia*. However, their overall distribution does not align with a traditional North–South divide. Beyond simple geographical location, I assess in Table C.4 how much each region contributes to the total number of girls living in these municipalities. As shown in column (2), more than 60% are located in *Lombardia* and *Veneto*, where their percentages are 3.5 and 4 times higher than their respective shares in the overall sample, reported in column (1). While *Trentino Alto-Adige* contributes less in absolute terms, it still displays a disproportionately high concentration relative to its national weight. Southern regions such as *Campania*, *Puglia*, *Calabria*, and *Sicilia* account for most of the remaining one-third of the subsample.

Because my identification strategy essentially relies on comparisons between treated and control students within cohort-by-province groups, I examine the data at that same level to better understand the source of the estimated effects. Due to the sample restriction to girls living in municipalities with the most conservative gender norms, these cohort-by-province cells are sometimes small or highly imbalanced. Roughly 5% of the cells contain fewer than 10 girls, and in about 20% of them, the number of students in one cell is less than 20% of the other. To assess whether these imbalances might be driving the results, I re-estimate the model under two increasingly restrictive sample criteria: the first requires each cell to include at least 10 girls and a treatment-to-control ratio no lower than 0.25; the second tightens the ratio threshold to 0.35. These checks allow me to test the robustness of the results to exclude highly imbalanced comparisons.

I present the results of these robustness checks in Table C.5. The most immediate consequence of the restrictions is a substantial drop in sample size, as many of the more imbalanced cohort-by-province cells are excluded. While this reduction naturally affects both the size of the estimates, the magnitude of the effects remains relatively stable across specifications. In both restricted samples, the estimated coefficients are approximately 20% of the baseline probability of choosing a STEM degree. Their persistence under stricter sample criteria shows that they are not driven by a handful of highly imbalanced comparisons.

Figure C.1: Location of the Municipalities with the Most Conservative Gender Norms



Notes: The graph highlights the municipalities belonging to the top quartile of conservative gender norms (in red) and those belonging to all other quartiles (in blue), separately. The quartiles of conservative gender norms are calculated using the percentage of votes received by the policy issuing restrictions on abortion during the 1981 referendum. The number of observations per category is reported in parenthesis. I exclude all municipalities in the provinces of *Aosta*, *Alto-Adige*, *Carbonia-Iglesias*, *Medio Campidano*, *Ogliastra*, and *Olbia-Tempio*. The sample size is 7,634 municipalities. That does not include the 692 for which there are no data (189 of those located in the provinces I exclude).

Table C.1: Robustness Exercise: Alternative Controls for School-Cohort Size and Municipality Population

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome Variable	STEM					
Subsample	Boys	Girls	Boys	Girls	Boys	Girls
PNI	0.021*	0.003	0.021*	0.003	0.020*	-0.003
	(0.012)	(0.010)	(0.012)	(0.010)	(0.022)	(0.011)
First-Stage	0.561***	0.578***	0.562***	0.578***	0.498***	0.525***
	(0.039)	(0.038)	(0.038)	(0.037)	(0.036)	(0.035)
F-statistic	225.62	179.11	214.09	154.49	216.64	173.43
Mean of the Outcome	0.26	0.23	0.26	0.23	0.26	0.23
SD of the Outcome	0.44	0.42	0.44	0.42	0.44	0.42
Observations	201,055	197,058	201,055	197,058	201,055	197,058
Cohort-Province Fixed Effects	X	X	X	X	X	X
Tercile of School-Cohort Size	X	X	-	-	X	X
Quintile of School-Cohort Size	-	-	X	X	-	-
Type of Municipality	X	X	X	X	-	-
Tercile of Municipality Population	-	-	-	-	X	X

Notes: I estimate the relationship between the probability of choosing a STEM degree ("STEM") and attending an academic track scientific high school (*Liceo Scientifico*) offering the *PNI* program, for boys and girls separately. I instrument *PNI* assignment using the share of available *PNI*, which is the ratio between the number of *Liceo Scientifico* schools offering the *PNI* option and the number of *Liceo Scientifico* schools within a 60-minute commute from the municipality of residence of the student. I always control for cohort-by-province fixed effects. In columns (1) and (2), I estimate Equation 3.1, which includes a categorical variable representing the tercile number of students in the school-cohort group, and a categorical variable indicating whether a municipality is considered rural, urban, or metropolitan. In columns (3) and (4), I instead include a categorical variable representing the quintile number of students in the school-cohort group, and a categorical variable indicating whether a municipality is considered rural, urban, or metropolitan. In columns (5) and (6), I instead include a categorical variable representing the tercile number of students in the school-cohort group, and a categorical variable representing the tercile of municipality population. I show the coefficient of the first-stage regression and the relative F-statistic, modifying Equation 3.2 accordingly. Only students who attended *Liceo Scientifico*, who enrolled in higher education between the academic years 2010/2011 and 2014/2015, did not repeat one or more years, and enrolled immediately in higher education are included. I also exclude all students and schools from the provinces of *Aosta*, *Alto-Adige*, *Carbonia-Iglesias*, *Medio Campidano*, *Ogliastra*, and *Olbia-Tempio*. Standard errors are clustered at the school level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table C.2: Robustness Exercise: Including Students Who Repeat One Year in High School

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome Variable	STEM		Dropout		Performance	
Subsample	Boys	Girls	Boys	Girls	Boys	Girls
PNI	0.021*	0.008	-0.031**	-0.019*	0.025	0.007
	(0.011)	(0.010)	(0.014)	(0.011)	(0.021)	(0.016)
First-Stage	0.556***	0.576***	0.556***	0.576***	0.551***	0.573***
	(0.039)	(0.038)	(0.039)	(0.038)	(0.039)	(0.038)
F-statistic	249.55	191.83	249.55	191.83	229.96	180.95
Mean of the Outcome	0.25	0.22	0.50	0.43	39.85	42.52
SD of the Outcome	0.43	0.42	0.50	0.50	18.25	17.46
Observations	228,590	213,194	228,590	213,194	199,577	195,517
Cohort-Province Fixed Effects	X	X	X	X	X	X
Tercile of School-Cohort Size	X	X	X	X	X	X
Type of Municipality	X	X	X	X	X	X

Notes: I estimate the relationship between the outcomes of interest and attending an academic track scientific high school (*Liceo Scientifico*) offering the *PNI* program using Equation 3.1, for boys and girls separately. The outcomes are: probability of choosing a STEM degree ("STEM"), the probability of dropping out from higher education ("Dropout") and students' performance during their first year of university ("Performance", which is proxied by the number of credits, ECTS, acquired during the first academic year. ECTS are granted upon exam completion, and students can acquire a maximum of 60 during each academic year). I use a 2SLS estimator in all columns. I instrument *PNI* assignment using the share of available *PNI*, which is the ratio between the number of *Liceo Scientifico* schools offering the *PNI* option and the number of *Liceo Scientifico* schools within a 60-minute commute from the municipality of residence of the student. I show the coefficient of the first-stage Equation 3.2 regression and the relative F-statistic. I control for cohort-by-province fixed effects, a categorical variable representing the tercile number of students in the school-cohort group, and a categorical variable indicating whether a municipality is considered rural, urban, or metropolitan. Only students who attended *Liceo Scientifico* and who enrolled in higher education between the academic years 2010/2011 and 2014/2015 immediately after finishing high school are included. I also exclude all students and schools from the provinces of *Aosta*, *Alto-Adige*, *Carbonia-Iglesias*, *Medio Campidano*, *Ogliastra*, and *Olbia-Tempio*. Standard errors are clustered at the school level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table C.3: Validation of Support for Abortion Restrictions as a Measure of Gender Norms

	(1)	(2)	(3)	(4)	(5)	(6)
	Women-to-Men		Share of Women Elected in:			
Outcome Variable	Labor Force Participation		City Councils		City Executive Committees	
Support for Abortion Restrictions	-0.096***	-0.084***	-0.033**	-0.060***	-0.078***	-0.093***
	(0.010)	(0.010)	(0.015)	(0.015)	(0.029)	(0.029)
Mean of the Outcome	0.67	0.67	0.15	0.15	0.17	0.17
SD of the Outcome	0.09	0.09	0.11	0.11	0.19	0.19
Observations	7,018	7,018	6,902	6,902	6,904	6,904
Province Fixed Effects	X	X	X	X	X	X
Type of Municipality		X		X		X

Notes: I estimate the relationship between the municipality-level support for the conservative abortion reform during the 1981 abortion referendum and three different dependent variables. One is the municipality-level ratio of women-to-men labor force participation as of 2011 (LFP), measured during the Italian National Census. The others are the municipality-level share of elected women administrators in the city council and the executive committee as of 2011, calculated using the Ministry of the Interior's Registry of Local and Regional Administrators. While the council functions as the local legislature, the executive committee is appointed by the mayor and carries out day-to-day governance. I control for province fixed effects (columns (1), (3), and (5)), and then for whether a municipality is considered rural, urban, or metropolitan (columns (2), (4), and (6)). I also exclude all municipalities from the provinces of *Aosta*, *Alto-Adige*, *Carbonia-Iglesias*, *Medio Campidano*, *Ogliastra*, and *Olbia-Tempio*. Standard errors are clustered at the municipality level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table C.4: Regional Distribution of Girls Residents in Municipalities with Conservative Gender Norms

	(1)	(2)
	Percentage of Girls by Region within:	
Region	Girls' Population	4th Quartile of CGN
Piemonte	6.25	1.89
Lombardia	12.38	33.46
Trentino Alto-Adige	0.76	3.56
Veneto	6.90	28.01
Friuli Venezia Giulia	1.97	0.47
Liguria	2.15	0.08
Emilia Romagna	5.65	0.32
Toscana	5.67	0.47
Umbria	1.46	0.10
Marche	2.71	1.45
Lazio	9.28	1.01
Abruzzo	2.98	1.31
Molise	0.79	0.89
Campania	14.34	6.48
Puglia	8.92	7.98
Basilicata	1.51	0.35
Calabria	5.15	4.57
Sicilia	9.51	5.10
Sardegna	1.62	2.50
Total (%)	100	100
Number of Girls	197,058	22,425

Notes: The table breaks down regional differences in the percentage of girls who enroll in higher education (column (1)) and those who live in a municipality belonging to the 4th quartile of the conservative gender norms ("CGN") distribution (column (2)), based on the share of votes in favor of more restrictive abortion laws in the 1981 referendum. I include girls who graduated from a *Liceo Scientifico*, who enrolled in higher education between the academic years 2010/2011 and 2014/2015, who did not fail one year or more years, and who did not immediately enter higher education after achieving their diploma. I also exclude all girls and schools from the provinces of *Aosta*, *Alto-Adige*, *Carbonia-Iglesias*, *Medio Campidano*, *Ogliastra*, and *Olbia-Tempio*.

Table C.5: Robustness Exercise: *PNI* and Local Gender Norms - Girls

	(1)	(2)	(3)	(4)	(5)	(6)
		Restriction #1			Restriction #2	
Outcome Variable	STEM	Dropout	Performance	STEM	Dropout	Performance
PNI	0.064**	0.037	0.020	0.048*	0.039	-0.039
	(0.026)	(0.030)	(0.034)	(0.028)	(0.039)	(0.037)
First-Stage	0.699***	0.699***	0.701***	0.783***	0.783***	0.787***
	(0.084)	(0.084)	(0.085)	(0.108)	(0.108)	(0.110)
F-statistic	73.12	73.12	73.16	70.92	70.92	71.97
Mean of the Outcome	0.22	0.40	46.03	0.21	0.39	45.99
SD of the Outcome	0.41	0.49	17.09	0.41	0.49	17.38
Observations	12,427	12,427	11,701	7,373	7,373	6,923
Cohort-Province Fixed Effects	X	X	X	X	X	X
Tercile of School-Cohort Size	X	X	X	X	X	X
Type of Municipality	X	X	X	X	X	X

Notes: I estimate the relationship between the outcomes of interest and attending an academic track scientific high school (*Liceo Scientifico*) offering the *PNI* program using Equation 3.1 for girls in the top quartile of the conservative gender norms distribution. Gender norms are calculated using the share of votes in favor of the conservative norm in the 1981 abortion referendum. I apply 2 different restrictions to the sample. Restriction #1 limits the sample to girls in cohort-by-province cells where both the treatment and control groups include at least 10 individuals, and where the ratio of treated to control girls (or vice versa) is no lower than 0.25. Restriction #2 limits the sample to girls in cohort-by-province cells where both the treatment and control groups include at least 10 individuals, and where the ratio of treated to control girls (or vice versa) is no lower than 0.35. The outcomes are: probability of choosing a STEM degree ("STEM"), the probability of dropping out from higher education ("Dropout") and students' performance during their first year of university ("Performance", which is proxied by the number of credits, ECTS, acquired during the first academic year. ECTS are granted upon exam completion, and students can acquire a maximum of 60 during each academic year). I use a 2SLS estimator in all columns. I instrument *PNI* assignment using the share of available *PNI*, which is the ratio between the number of *Liceo Scientifico* schools offering the *PNI* option and the number of *Liceo Scientifico* schools within a 60-minute commute from the municipality of residence of the student. I show the coefficient of the first-stage Equation 3.2 regression and the relative F-statistic. I control for cohort-by-province fixed effects, a categorical variable representing the tercile number of students in the school-cohort group, and a categorical variable indicating whether a municipality is considered rural, urban, or metropolitan. Only students who attended *Liceo Scientifico*, who enrolled in higher education between the academic years 2010/2011 and 2014/2015, did not repeat one or more years, and enrolled immediately in higher education are included. I also exclude all girls and schools from the provinces of *Aosta*, *Alto-Adige*, *Carbonia-Iglesias*, *Medio Campidano*, *Ogliastra*, and *Olbia-Tempio*. Standard errors are clustered at the school level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.