CRC
Centro di Ricerche
sulla Cooperazione e sul Nonprofit

CSCC
Centro di ricerca in Scienze
Cognitive e della Comunicazione

UNIVERSITÀ CATTOLICA del Sacro Cuore

CRC - CENTRO RICERCHE SULLA COOPERAZIONE E SUL NONPROFIT

Service Learning, Well-being and School Performance: Causal Evidence from Italian High-School Students
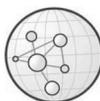
Gianluca Argentin, Gian Paolo Barbetta, Mario A. Maggioni,
Domenico Rossignoli, Luca Stella

VP VITA E PENSIERO

CRC
Centro di Ricerche
sulla Cooperazione e sul Nonprofit

CSCC
Centro di ricerca in Scienze
Cognitive e della Comunicazione

Università Cattolica del Sacro Cuore

CRC - CENTRO RICERCHE SULLA COOPERAZIONE E SUL NONPROFIT

Service Learning, Well-being and School Performance: Causal Evidence from Italian High-School Students

Gianluca Argentin, Gian Paolo Barbetta, Mario A. Maggioni, Domenico Rossignoli, Luca Stella

VP VITA E PENSIERO

Gianluca Argentin, Università degli Studi di Milano, Bicocca
Gian Paolo Barbetta, Università Cattolica del Sacro Cuore
Mario A. Maggioni, Università Cattolica del Sacro Cuore
Domenico Rossignoli, Università Cattolica del Sacro Cuore
Luca Stella, Freie Universität, Berlin, Germany

✉   coop.ricerche@unicatt.it

www.vitaepensiero.it

INDICE

Abstract

Service learning (SL) is a relatively common pedagogical method in the US, where it has been widely adopted in schools, colleges, and universities. The method requires students to take part in the activities of organizations that serve community needs. While the literature argues that SL activities could generate beneficial effects for students' cognitive abilities, self-esteem and motivation, satisfaction with schools, attitudes towards institutions, and civic engagement, empirical evaluation of these effects is scarce and frequently far from rigorous.

This paper investigates the effects of being engaged in "SL-like" activities on the school performance of 9th and 10th grade students at high risk of school failure and drop-out in Italy. We contribute to the empirical literature on SL in three ways. First, we run the first pilot randomized controlled trial (RCT) to simultaneously evaluate the effect of a SL program on both cognitive and non-cognitive skills, with the latter measured through questionnaires and incentivized tasks. Second, this is the first attempt to evaluate the impact of SL as a remedial intervention on the specific target of low-achieving students at risk of dropping out from school. Finally, this is the first attempt to rigorously investigate SL activities in Italy, and one of the first in Europe, as most RCTs involving SL have focused on the US.

Our experimental results show that—consistently with the literature developed in the US—participation in "SL-like" activities leads to a general improvement in non-cognitive skills of students at risk of dropping out. On the other hand, contrary to what the literature argues, the intervention does not improve cognitive skills, since participation in the program even increases the risk of school failure. These results suggest designing and implementing SL interventions in schools with great care to avoid unintended negative consequences.

4

1. Introduction

This paper investigates the effects of being engaged in "service learning (SL)-like" activities on the school performance of 9th and 10th grade students at high risk of school failure and drop-out.

We contribute to the empirical literature on SL in three important ways. First, to the best of our knowledge, this is the first randomized controlled trial (RCT) to simultaneously evaluate the effect of a SL program on both cognitive and non-cognitive skills, with non-cognitive skills measured not only through questionnaires but also through incentivized tasks. Second, this is the first attempt to evaluate the impact of SL as a remedial intervention on the specific target of low-achieving students at risk of dropping out from school, as a sub-population that has not been adequately investigated in previous studies. Given that dropping out of high school represents a strong predictor of future unemployment, this target is particularly crucial for countries like Italy with a relevant level of youth unemployment. Finally, this is the first attempt to rigorously investigate SL activities in Italy, and one of the first in Europe, as most RCTs involving SL have focused on the US.

SL is traditionally defined as a "credit-bearing educational experience in which students participate in an organized service activity that meets identified community needs and reflect on the service activity in such a way as to gain further understanding of course contents, a broader appreciation of the discipline and an enhanced sense

of civic responsibility" (Bringle and Hatcher, 1996, p. 222). In other words, SL typically requires students to help an organization that serves community needs or its clientele in the form of unpaid program-related labor. This pedagogical method is quite common in the US and Latin American countries where it has been widely adopted in K-12 schools, colleges, and—most of all—universities. On the contrary, the SL approach is relatively new in Europe, where only a few countries have implemented it and even a smaller subset of them has tested its efficacy.

The literature—mostly based on qualitative studies—suggests that SL activities could generate beneficial effects for undergraduate students' cognitive abilities (Novak et al., 2007; Warren, 2012), as well as other outcomes, such as self-esteem and motivation, satisfaction with schools and teachers (Henderson and Berler, 1995), attitudes towards institutions (Henderson et al., 1986), and civic engagement (Celio et al. 2011). Nonetheless, an evaluation of the beneficial effects of SL is far from rigorous: most research is based on correlational studies, and meta-analyses only include quasi-experimental studies, given the very small number of RCTs measuring the effect of SL programs (Warren, 2012). To sum things up, "the quality of service-learning research has been criticized on a number of grounds (…). Perhaps, the most troublesome problem is that of self-selection" (Aronson et al., 2005, pp. 142-143). In fact, "(s)tudies of service learning are usually conducted with existing programs and rarely use

random assignment for both financial and educational reasons" (Hecht, 2003, p. 95). In fact, scholars conducting meta-analyses recommend that researchers should "include a comparison group when examining service learning outcomes so that they can be sure the results obtained are attributable to the pedagogy and not to some other factor" (Warren, 2012, p. 59). For instance, the meta-analysis of Celio et al. (2011) identified 62 studies investigating the effects of SL, but only nineteen of them were based on randomized designs.

We attempt to fill these gaps in the literature by using an RCT to rigorously evaluate the effect of the "SL-like" project named "Non solo a scuola" (i.e., "Not only at school"), which involved several high schools located in the Northern Italian province of Monza and Brianza and local grassroots organizations. While SL activities normally involve whole classes, the "Non solo a scuola" project focuses on the specific sub-population of 9th and 10th graders at high risk of school failure and drop-out. This is a particularly relevant issue as the Italian drop-out rate is among the highest in Europe (Eurostat, 2018), which continues to be worrisome despite its decline in recent years. Given that the heterogenous effects of SL programs have rarely been investigated in existing literature (Filges at al. 2021), the specific focus of the project on high-risk students represents another original contribution of our paper.

The protocol of the project involved students at high risk of dropping out of school in the activities of voluntary or third-sector organiza-

tions operating at the local level. Students were engaged for about 60 hours, during school time. In the opinion of the local non-profit organizations promoting the SL project, carrying out activities for the community and in favor of disadvantaged people should increase students' motivation and self-esteem, enhancing and making them more conscious of the skills they have. Accordingly, the project should also improve the students' attitudes towards study and school, thereby contributing to improving their school performance. In addition, interacting during the SL project with an adult—who is neither a teacher nor a parent, but an educator trained in the project's aim—should help the student to reinforce her commitment towards school and study.

To evaluate the impact of the project, approximately 140 students from six high schools who were named by their teachers as possible beneficiaries of the intervention were randomly assigned to the treatment or control group. The experiment was designed to estimate the impact of SL on students' self-esteem, well-being at school, prosocial behavior, and school performance, as measured by failure in the 2017/18 school year. This last outcome is crucial in the Italian education system, since failing a grade implies repeating in the following year and it frequently represents the first step of a process leading to drop-out. More precisely, to assess the overall effects of the intervention, we consider two sets of outcome variables: i) a wide set of non-cognitive skills self-reported by students through a self-

filled questionnaire or measured through an incentivized task taken from the behavioral games literature, administered at both baseline and the end of the school year; and ii) school performance as measured by passing or failing the grade. To the best of our knowledge, this is the first time that incentivized tasks have been used to evaluate the effects of "SL-like" activities.

We estimate an intention-to-treat (ITT) effect and we adopt an instrumental variables (IV) approach to address the issue of non-compliance with randomization (given that some students assigned to the intervention decided not to take part in it). Our experimental results show that—consistent with the literature developed in the US—participation in "SL-like" activities leads to a general improvement in non-cognitive skills of students at risk of dropping out (although in some cases the effect is substantively relevant but not statistically significant, due to the small sample size). On the other hand, contrary to what the literature shows, participation in the program increases the risk of failing a school year.

We argue that the adverse effect that we detect for the school outcome may depend on the planning and implementation of the intervention, which required students to carry out SL activities during school hours, while their peers attended classes. This approach prevented the students assigned to the intervention arm of the RCT from fully participating in the class activities experienced by their classmates, therefore jeopardizing their cognitive performance. Therefore,

leaving class during school hours—while positively influencing non-cognitive skills—is very likely to have further disconnected low-achieving students from their classmates, teachers, and the cognitive tasks that they should carry out to pass their grade. This possible explanation for the unexpected detrimental effect of SL is supported by the qualitative evidence emerging from the focus groups involving teachers when the results of the intervention were disclosed to schools. On those occasions, it also emerged that teachers were not interested in rewarding students' SL experience, particularly when their performance in curricular disciplines did not improve. These results support the literature suggestion that SL activities should be fully integrated in the classroom experience (Novak et al., 2007).

On the policy side, our results suggest that SL interventions improve students' non-cognitive skills, as a necessary intermediate step towards enhancing their academic results. Nonetheless, poorly-implemented interventions (i.e., interventions not carefully aligned with academic curricula) could produce detrimental effects on the cognitive side. Therefore, a full understanding of the possible trade-offs generated by poorly-planned SL interventions as well as their implementation needs is crucial. This is particularly relevant given that establishing partnerships between schools and civil society organizations is a recent educational trend in Italy as well as several other countries.

The remainder of this paper continues as follows. Section 2 discusses the extant literature, before Section 3 outlines the background and presents the treatment. Section 4 presents the RCT design, our evaluation sample and the outcome variables. Section 5 illustrates the results, and finally Section 6 concludes.

2. Related Literature

This paper contributes to the extensive literature in psychology, sociology and economics analyzing the effects of SL programs on student outcomes. Previous studies suggest that participation in SL can affect a large array of student outcomes in many areas, including attitudes towards oneself, attitudes towards school and learning, civic engagement, social skills, and academic achievement (see, for instance, Maples et al, 2020; Hébert and Hauf, 2015; Billig, 2009; Conway et al, 2009). These studies have been mostly qualitative and observational, meaning that control groups were often not present and students' access to the program was rarely randomized (as shown by Warren, 2012; Celio et al., 2011 and Novak et al., 2007). In addition, previous literature has almost exclusively focused on the US.

In pre-post studies, SL students showed increased self-esteem and self-concept, more highly internalized moral standards, more positive attitudes towards school and education, improved satisfaction with classes and teachers, greater interest in, commitment to, and sensitiv-

ity towards their communities and their needs, and stronger beliefs that one can make a difference in the world (Billig, Root, & Jesse, 2005; Henderson and Berler, 1995; Henderson, Marburger and Ooms, 1986). SL programs also seem to have been successful in improving social skills related to communication, leadership, and problem-solving (McNatt, 2019; Lester et al., 2005; Papamarcos, 2005). Furthermore, several scholars have explored the relationship between SL and academic achievement, with most studies suggesting that SL leads to higher academic achievement (e.g., Billig, 2009; Giles and Eyler, 1994; Harwood and Radoff, 2009; Markus, Howard and King, 1993).

The meta-analyses by Novak et al. (2007) and Warren (2012) review the body of research measuring the impact of SL at the undergraduate level on cognitive outcomes, including enhanced academic understanding of the subject matter, the ability to apply the knowledge and skills learned in one setting to another, and the ability to reframe complex social issues. The authors conclude that students participating in SL activities exhibit better learning outcomes compared to those not taking part in it.

On the other hand, a smaller body of the literature has challenged the beneficial effects of SL on academic achievement (Poon et al., 2011; Moely et al., 2002). For instance, Poon et al. (2011) uncover that "students have an increased level of sense of social responsibility as well as ethical and moral behavior after the participation in SL pro-

jects. Nevertheless, no significant difference is found for practical learning outcomes between the pre-test and post-test" (p. 185). Meanwhile, comparing SL and non-SL students, Moely et al. (2002) find that SL students report a slight decrease in learning about the academic field over the course of the semester, although it was not as large as the decrease shown by students not participating in SL.

At the same time, as emphasized by the meta-analysis of Celio et al. (2011), success in accomplishing the desired learning outcomes critically depends on adopting recommended practices, namely the integration of SL in the classroom experience, incorporating youth voice, involving community partners and providing opportunities for reflecting on the experience.

Most relevantly, while there is a substantial consensus on the positive correlation between SL and student learning outcomes, less is known about the existence of a causal relationship, and the resulting empirical evidence remains rather inconclusive. The meta-analyses by Celio et al. (2011) and Warren (2012) show that only a small number of research studies adopt sound methodological tools.

To the best of our knowledge, only a few contributions have attempted to experimentally test the relative impact of engaging in a SL project on student outcomes. McNatt (2019) uses a longitudinal experiment and finds that engaging in SL projects improves subsequent presentation performance. Adopting an RCT, Leung et al. (2012) find that SL activities increased medical and nursing students'

knowledge of aging, their understanding of mental health needs in old age and reduced their negative attitudes towards older adults.

The goal of this paper is twofold: on the one hand, we aim to fill the methodological gap in the literature by using an RCT design; and on the other hand, we evaluate the effect of SL-like activities on the unusual and specific target of low-achieving students in their first two years of high school (i.e., 9th and 10th grades). This is particularly important given that many countries struggle with early school leaving and only a few interventions have proven successful in fighting this problem.

Finally, our work adds to the growing literature comprising evaluations of interventions aimed at reducing inequality in educational achievement and opportunities. Carlana et al. (2020) consider a program that targets high-achieving immigrant students with the aim of reducing the immigrant-native educational gap. Other interventions have targeted low-achieving students and provided a combination of information on school options and mentorship on soft skills. Some of these programs have been shown to be successful in reducing grade retention and high-school drop-out rates (e.g., Goux et al., 2017; Martins, 2010; Algan et al., 2020), while others had zero or negative effects (Rodriguez-Planas, 2012). The program studied in this paper can be seen as complementary to the latter set of interventions, as it targets the most fragile students, with the aim of improving their non-cognitive skills and school achievement.

3. Background

3.1. Service Learning: A Definition with a Focus on Italy

The SL pedagogical approach has been widely adopted in the US and Latin American countries over recent decades. According to Scales and Roehlkepartain (2004) about 22% of primary schools, 31% of middle schools, and 44% of high schools in the US actively engage students in SL activities. Furthermore, Furco (2010) indicates SL as "one of the fastest growing educational initiatives in contemporary primary, secondary and post-secondary education" (p. 228). SL programs have been implemented in Argentina (Ierullo, 2016), Columbia (Perold and Tapia, 2008) and Singapore (Chua, 2010).

The wide success of this approach is one of the reasons explaining the difficulties that researchers encounter in defining its precise content. In fact, the goal of SL ranges from fostering students' participation and civicness (such as increasing their involvement in the life of their community, or their participation in school activities), to improving their attitudes towards oneself as well as school and learning (such as self-esteem and positive relations with teachers and peers), and developing academic achievements (such as increasing critical thinking and problem-solving skills or improving cognitive skills and attainments in courses). Moreover, the practical implementation of SL programs in schools ranges from interventions involving an entire institute to those including one or more grades or reserved for a single class. Student participation can be voluntary or mandatory, and

the frequency and duration of the activities can strongly vary. Furthermore, a distinction between co-curricular SL and academic SL has become quite common (Howard, 2003). In the first model, what students learn from experience is considered outside of the school domain, while in the second model the learning experience in the community is complementary to that in school. What results are "highly idiosyncratic, situational experiences for which there is minimal predictability of how each service-learning experience will unfold. Indeed, no two service-learning activities are alike" (Furco, 2003, p. 26), while even less is known about the intrinsic quality of SL experiences (Warren, 2012).

These differences notwithstanding, SL is generally considered as "a form of the broader model of experiential education, with community service as the fulcrum" (Howard, 2003, p. 17). It is exactly this focus on community efforts Moely et al. (2002)—and the related commitment to the welfare of society—that differentiates SL from other forms of experiential education, such as internships or simulations. Moreover, a few necessary features allow distinguishing SL from other learning experience: "First, there is a service (…) that responds to a need that originates in the community (…); second, students' academic learning is strengthened (…); and third, students' commitment to civic participation, active democratic citizenship, and/or social responsibility is advanced" (Howard, 2003, p. 18).

In Italy, although SL remains in its infancy, the approach is sparking interest and some schools—along with grassroot non-profit organizations based in their communities—are experimenting with "SL-like" interventions. Following this growing attention, the Italian Ministry of Education has recently produced a white paper entitled "Una via italiana per il service learning" (An Italian way to service learning) (MIUR, 2018), which describes the SL approach, providing some general guidelines, and offering practical suggestions to put this pedagogical approach into practice. The guidelines and practical suggestions are the result of a first exploratory round of SL projects promoted by the ministry, involving about 65 schools in the three Italian regions of Lombardy (the same region in which our intervention took place), Tuscany and Calabria. These projects included all kinds of schools from primary to upper secondary level, promoting a wide range of service experiences (from helping elderly people to promoting the local cultural heritage) and involving different disciplinary contents. Within schools, projects were generally aimed at classes or grades and—as far as we are aware—they did not have a remedial aim. In this respect, the "Non solo a scuola" program is unique. Moreover, as is usual in the Italian case (Abbiati et al., 2022), these projects have not been subject to a rigorous and systematic evaluation, meaning that only anecdotal and qualitative information is available.

3.2. The Treatment: "Non solo a scuola"

The "Non solo scuola" program shows the typical characteristics of the SL approach. In fact, the program stems from a close collaboration between six high schools and some grassroot non-profit organizations located in the North Italian province of Monza and Brianza (in the Lombardy region). Moreover, the students attending the program are offered to serve the needs of their communities by volunteering with local non-profit organizations in out-of-school activities during school hours. Activities are as diverse as assisting disabled persons attending hippotherapy, mentoring immigrants or serving customers in fair trade shops. Students are assisted by trained tutors in their choice between the different possible alternatives and while developing their activities they are monitored by educators, aware that the activities are aimed at developing students' skills, with particular attention to non-cognitive ones.

A peculiar characteristic that makes this SL project particularly interesting is represented by the target population. In fact, the program aims at motivating low-achieving students at high risk of leaving school, thus reducing their risk of actual drop-out. The program focuses on students attending the first and second years of the Italian high-school system (i.e., 9th and 10th grade) who are identified by their teachers as being at risk of dropping out of school. Involving only a small number of students in each class, the program is not fully integrated in the didactical activities.

The main aim of the program is to develop the non-cognitive skills of the target students, improving their self-esteem, motivation, pro-activity and pro-sociality. The theory behind the intervention assumes that the development of non-cognitive skills can play a positive role in the target population, helping students to increase their effort and commitment in increasing cognitive skills, thus reducing their risk of failing a grade and dropping out of school. More precisely, the program can affect students' behavior through two distinct channels, the first of which is mentoring: during service activities, students are supported by experienced tutors who are informed about their school difficulties, as well as the volunteers and the staff of the local organizations. The second channel is empowerment: students perform real-life tasks, taking responsibility and becoming increasingly aware that they can successfully accomplish these tasks.

4. RCT Design

4.1. Research Design and Sample

Figure 1 shows the Consort-like diagram of the RCT. At the beginning of the 2016/17 school year, the teachers of the six high schools involved were asked to propose 250 of their students to take part in the project. Most unfortunately for the power of our analysis, the schools only proposed 169 of their low-performing students—attending 9th and 10th grades and at risk of dropping out of school—

as possible candidates for the "Non solo a scuola" program (from 12 up to 52 students per school).

Figure 1 - Consort-Like Diagram of the RCT
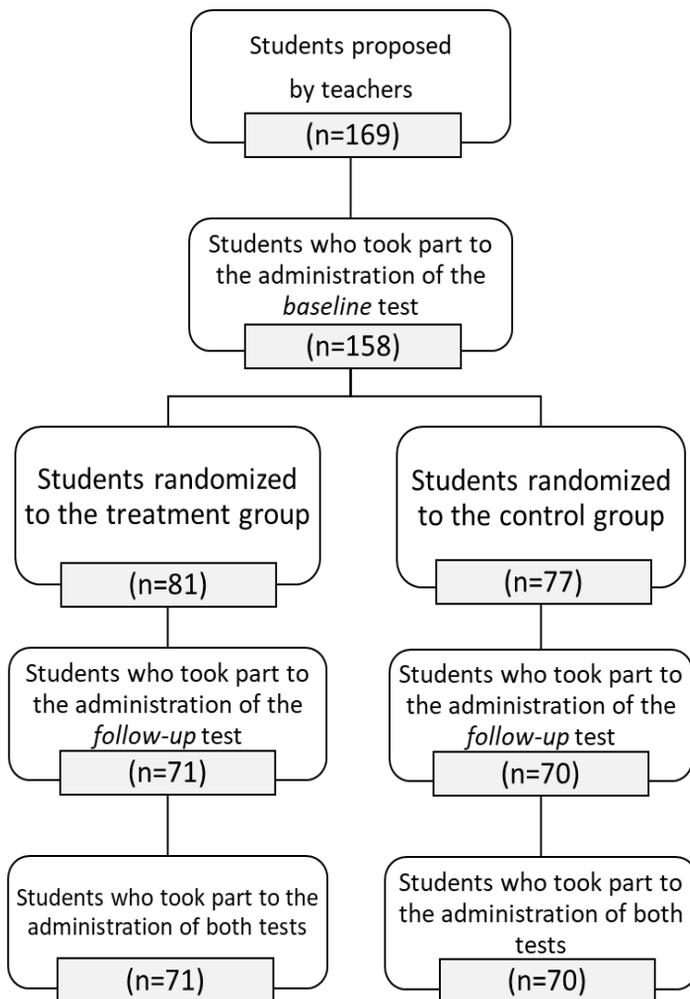
All of the proposed students were invited to participate in the base-line test administration sessions, which took place in each school during school time, and which aimed at gaining statistical power for our estimates through pre-intervention measures of each outcome. During these sessions, the staff of the Università Cattolica administered the blind questionnaire and the incentivized tasks necessary to measure two of the three sets of outcome variables described in detail in Section 4.2 (attitudes and behavior). Students were not aware that the questionnaire was aimed at estimating the effectiveness of the SL activities; in fact, it was presented as academic research.

A few students (eleven) refused (or were not authorized by their parents) to participate in the test administration sessions and were therefore excluded from the project. Consequently, 158 students were randomly assigned to either the treatment or control group (randomized students' characteristics are displayed in the following, see Table 1).

Randomization was stratified by school (and eleven different randomization blocks within schools, for institutions located in different buildings) to control for unobserved school characteristics and allow each school to experience the treatment. Overall, 81 students were randomized into the treatment group and invited to participate in SL activities over the 2016/2017 school year, while 77 students were randomized into the control group and followed their traditional activities.

Between March and May 2017, the students randomly assigned to the treatment group were involved in the different program activities described in the previous section. They were required to meet several times[1] for about three hours. Twelve students (14.8%) never showed up. The 69 students who showed up at least once attended their meetings for about 80% of the scheduled time. The activities ended before June 2017, hence before the end of the school year.

In June 2017, at the end of the intervention, the students assigned to both groups were asked to participate in the follow-up test administration sessions, where they were re-administered the questionnaire and incentivized tasks to take new measures of the outcome variables. As shown in Figure 1, some students could not be reached for administering the follow-up tests, although balancing between treated and controls was preserved (see Section 4.4). Questionnaire administrators were unaware of the treatment or control condition of each student, and once again students were not aware of the relationship between the questionnaire's administration and the SL project.

In October 2017, we collected administrative information on randomized students from each school to measure the third set of outcomes (pass or fail) (see Section 4.2).

---

[1] Different activities implied different numbers of meetings. On average, students were required to meet eight times, with a minimum of four and a maximum of twenty times.

4.2. Outcome Variables

We assessed the effectiveness of the SL learning program on three sets of outcomes: psychological, behavioral and school performance outcomes. The investigation of three sets of outcomes provides a nuanced and detailed picture of the program effect and allows developing useful insights into the mechanisms underlining the possible impact of the intervention. Furthermore, most of the literature agrees in considering the importance of these outcomes (Celio et al., 2011).

Specifically, psychological outcomes (henceforth attitudes) are measured through a set of adapted versions of previously-validated psychological scales that gauge: 1) self-esteem; 2) well-being at school; and 3) declared pro-sociality. Details about how these concepts were measured are provided in Appendix B of the paper.

While psychological scales are widely used in the education literature, they are limited by the self-describing nature of the instruments. For this reason, we decided to measure several pro-social attitudes that may have been affected by the program through a set of incentivized tasks that are widely used within the Behavioral and Experimental Economics fields. These are usually referred to in the literature as "games", thus determining some behavioral outcomes (henceforth behaviors). The use of incentivized tasks allows us to make inference on the "actual" behavior of the respondents when their choices have real consequences in monetary terms. In particular, our analysis includes five "games" to elicit three pro-social attitudes:

pure generosity/altruism, through a "Dictator Game"; inequality aversion and sense of justice, through an "Ultimatum Game"; reciprocity, through the role of "respondent" in a Trust Game (which we label "Gratitude Game"); sincerity, through the "dice-rolling task" (which we label "Sincerity Game"); and risk aversion, through the "Balloon Analogous Risk Task" (BART). Further details about how we implemented these "games" are provided in Appendix B of the paper.

Finally, we included in our analysis a crucial measure of students' performance at school (henceforth performance), namely their final result (pass to the next grade or fail) decided by their teachers and provided by the school administrations.

4.3. Estimation Technique

We estimate the effect of SL by measuring the difference in the average level of the outcome variables for the students assigned to the treatment and control groups with an ordinary least squares (OLS) regression model. Namely, we test whether being assigned to the treatment group affects the outcome variable of interest. Therefore, we estimate the following linear regression model:

$$Y_{ij} = \alpha + \beta T_{ij} + \gamma X_{ij} + \varepsilon_{ij} \qquad (1)$$

where the index *ij* denotes student *i* attending school *j*. Y_{ij} represents a set of follow-up outcomes, described in Section 4.2 and grouped in three main categories of psychological, behavioral and cognitive outcomes. T_{ij} is a binary variable taking the value of 1 for student i attending school j randomized to the treatment arm of the experiment, while X_{ij} represents the baseline measure of the outcome under investigation. All of our estimates further include fixed effects for the different randomization blocks. Finally, ε_{ij} denotes an idiosyncratic error term.

The key parameter of interest is β, which indicates the change in the outcomes of the individuals assigned to the treatment after attending the SL program, relative to the control group. Therefore, β identifies the causal effect of the intervention on the outcome variable. This is known as the reduced-form estimate, or the ITT effect of the intervention (Angrist and Pisckhe, 2009). Throughout the analysis, standard errors are clustered at the classroom level.

The ITT estimates could be confounded by the fact that while all students randomized to the treatment group were assigned to the SL activities, not all of them participated in the intervention with the same intensity, given that students could not be forced to take part in the activities. In this context, the ITT estimate of the effect of SL underestimates the value of receiving the treatment. Approximately, 76% of the students took part in at least half of the scheduled hours and meetings, and about 79% participated in the intervention for at least

ten hours. This behavior determines non-compliance with the proto-col, which is common of RCTs in the social and educational fields. We create a measure of compliance with the protocol, which is defined as attending SL activities at least 80% of the scheduled hours and meetings, as well as participating in the intervention for at least 20 hours, resulting in a compliance rate of around 40%.

To provide a further test of the effect of the program, Model (1) is then estimated using two-stage least squares (2SLS) regression, and the first-stage regression is given by:

$$C_{ij} = \rho + \sigma T_{ij} + \phi X_{ij} + v_{ij} \qquad (2)$$

where C_{ij} defines our compliance variable detailed above and is instrumented with the random assignment variable T_{ij}. The variable X_{ij}, the parameters ρ, σ, ϕ, and the error term v_{ij} are defined in the same way as in Equation (1). The second-stage model is written as:

$$Y_{ij} = \alpha + \beta C_{ij} + \gamma X_{ij} + \varepsilon_{ij} \qquad (3)$$

where the outcome variable is predicted by C_{ij} and the set of covariates included in the first-stage model. In this second-stage model, the coefficient β of variable C_{ij} represents the local average treatment effect (LATE) estimate and indicates the impact of complying with the treatment on the outcome variable.

4.4. Attrition and Balance

To estimate the effect of SL on students' achievement, we consider the sample of 141 students who took both the baseline and follow-up test in the experiment (see Figure 1). The overall attrition rate of this sample is about 10%, which may raise concerns about the internal validity of our experiment. One may worry that study participation in the follow-up data collections systematically relates to the treatment status, which would bias our estimates. To address this concern, in Table A.1 in Appendix A we conduct a differential attrition rate test (Ghanem, Hirshleifer, and Ortiz-Becerra, 2021) to determine whether the rates of attrition are statistically different across treatment and control groups. Reassuringly, the result of a regression of the treatment dummy variable on an indicator for not being in the sample of analysis yields a non-statistically-significant coefficient of 0.031 (p-value=0.58) with robust standard errors clustered at the classroom level. We also perform a determinants of attrition test to assess whether baseline covariates and outcomes are correlated with response status. The results confirm the absence of significant differences in the patterns of response between respondents and non-respondents in the baseline covariates and most outcomes (see Tables A.2 and A.3 in the Appendix, respectively). We further consider a selective attrition test to assess whether—conditional on being a respondent or not—the mean of observable characteristics is similar

across treatment and control groups. Tables A.4 and A.5 in the Appendix largely support this similarity.

We now move on to examining the quality of the randomization process. Table 1 reports the estimates of "reverse regressions" of each of the baseline covariates on our treatment variable, and randomization blocks. We find that the effects of SL on baseline covariates are very small and not statistically significant, except for a marginally significant coefficient on the year of birth. This suggests that the treated and control groups are well balanced on the observable characteristics, thereby boosting the confidence in the internal validity of our study.

As a further sample balance check, we regress our standardized outcomes on self-esteem, well-being at school, declared pro-sociality and behaviors at the baseline on the treatment dummy, and randomization blocks included in the main estimations of the treatment effect. Table 2 below shows that aside from the baseline score for negative attitude towards school and altruistic behavior, none of the coefficients related to the treatment dummy is significantly different from zero, indicating that for all other baseline outcomes there is no evidence of significant imbalances between the treatment and control group.

Overall, Tables 1 and 2 display the baseline characteristics and outcomes of the treatment and comparison groups and show that they are balanced, as no difference between the two groups—in the out-

comes at the baseline and in the most relevant covariates—is statistically significant. This provides evidence of the statistical equivalence of the two groups before the treatment.

Table 1 - Descriptive Statistics and Balancing Tests (Baseline Covariates)

	(1) Summary	(2) Summary	(3) Balancing	(4) Balancing
	Mean	Std. dev.	Coefficient	p-value
Male	0.589	0.494	0.115	0.141
Year of birth	2001	0.950	-0.222	0.084
Born in Italy	0.848	0.360	-0.028	0.577
Father is present	0.734	0.443	0.082	0.212
Mother is present	0.911	0.285	0.002	0.959
Parents are present	0.677	0.469	0.072	0.266
Grade 9	0.386	0.488	-0.095	0.114
Pass grade in lower secondary school	0.557	0.498	0.043	0.625
Satisfactory or good grade in lower secondary school	0.443	0.498	-0.043	0.625
One hundred or more books at home	0.437	0.498	0.121	0.1715
Lost years of schooling	0.690	0.704	0.113	0.224
Parental years of schooling	9.253	1.565	0.034	0.872
Passing to next grade (without school debts)	0.285	0.453	–	–
Passing to next grade (even with school debts)	0.658	0.476	–	–

Notes - The sample consists of 158 students. Columns 3 and 4 report the coefficient of the SL program derived by reverse regressions of the pre-intervention covariates listed in each row on SL program. Note that information on passing to the next grade is available at the end of the intervention.

Table 2 - Balancing Tests – Outcomes at Baseline

	(1) Coefficient	(2) p-value
Positive self-esteem	-0.151	0.369
Negative self-esteem	-0.091	0.559
Relational self-efficacy (with peers)	0.157	0.402
Relational self-efficacy (managing conflicts)	0.241	0.132
Cognitive self-efficacy (related to school activities)	-0.068	0.604
Cognitive self-efficacy (general)	0.120	0.455
Emotional self-efficacy (managing emotions)	0.121	0.333
Emotional self-efficacy (getting support)	0.107	0.442
Attitudes towards school (negative feelings)	-0.303	0.040
Attitudes towards school (school perceived as meaningless)	-0.215	0.102
Lack of motivation in studying	-0.176	0.247
Attitudes towards teachers (positive relationships)	-0.165	0.315
Attitudes towards teachers (feeling persecuted)	0.151	0.322
Attitudes towards teachers (getting support)	-0.017	0.914
Attitudes towards teachers (allied with my teachers)	0.010	0.947
Expectations in education	-0.194	0.226
Expectations in life	-0.019	0.902
Prosociality	0.094	0.608
Internal locus of control	0.146	0.263
Stable locus of control	-0.116	0.386
Sincerity behavior	-0.024	0.773
Altruistic behavior (dictator)	0.051	0.073
Risk attitude (bart)	0.031	0.771
Adversion to inequality behavior (ultimatum)	0.050	0.152
Gratitude behavior (gratitude)	-0.006	0.823
Cheating (max choice)	0.014	0.553

Notes - The sample consists of 158 students. Columns 1 and 2 report the coefficient of the SL program derived by reverse regressions of the pre-intervention outcomes listed in each row on SL program.

5. *Results*

ITT estimates. Tables 3 and 4 present the results of the OLS estimation of the model described by Equation 1 for attitudes, behaviors and performance, respectively.

Starting with attitudes in Table 3, we find that undertaking SL-like activities has a positive—although not statistically significant—effect on most attitudes. Moreover, we find that treated students display a more pro-social attitude. Specifically, the estimated effect size of the SL program increases pro-social attitude by 0.17 standard deviations (see column 18). We also find a significant increase in the internal locus of control (see column 19), and a negative—albeit non-significant—reduction in external locus of control (see column 20). Treated students are 0.28 standard deviations more likely to present internal locus of control, and 0.12 standard deviations less likely to have external locus of control. As shown in Panel A of Table 4, the results on behaviors are mixed: while we find a decrease in sincerity behavior, we find that the treatment leads to a significant reduction in the risk attitude of about 0.2 standard deviations. We also find a positive effect on altruistic behavior, although the coefficient is not precisely estimated. Finally, turning to performance as the outcome in Panel B of Table 4, we find that SL has a negative and marginally statistically significant effect on the probability that the student passes to the next grade. The coefficient becomes smaller and o longer significant when considering a full pass grade as the outcome.

Table 3: Effects of Service Learning on Attitudes – ITT Estimates

Dep. var.:	(1) Positive self-esteem	(2) Negative self-esteem	(3) Relational self-efficacy (with peers)	(4) Relational self-efficacy (managing conflicts)	(5) Cognitive self-efficacy (related to school activities)	(6) Cognitive self-efficacy (general)	(7) Emotional self-efficacy (managing emotions)
Service learning	0.091 (0.077)	-0.097 (0.117)	0.076 (0.067)	0.014 (0.066)	0.089 (0.066)	0.017 (0.076)	0.057 (0.094)
Effect size	0.141	-0.109	0.145	0.024	0.140	0.033	0.078
Observations	141	141	141	141	141	141	141
Mean of dep. var.	0.291	0.204	0.306	0.301	0.290	0.309	0.267
Std. dev. of dep. var.	0.573	0.854	0.496	0.524	0.576	0.476	0.676

Dep. var.:	(8) Emotional self-efficacy (getting support)	(9) Attitudes towards school (negative feelings)	(10) Attitudes towards school (school perceived as meaningless)	(11) Lack of motivation in studying	(12) Attitudes towards teachers (positive relationships)	(13) Attitudes towards teachers (feeling persecuted)	(14) Attitudes towards teachers (getting support)
Service learning	0.058 (0.094)	0.091 (0.112)	0.104 (0.102)	0.035 (0.110)	-0.030 (0.081)	-0.119 (0.138)	0.185 (0.124)
Effect size	0.079	0.109	0.143	0.040	-0.051	-0.146	0.215
Observations	141	141	141	141	141	141	141
Mean of dep. var.	0.262	0.229	0.259	0.238	0.285	0.232	0.197
Std. dev. of dep. var.	0.693	0.795	0.701	0.769	0.602	0.785	0.870

Dep. var.:	(15) Attitudes towards teachers (allied with my teachers)	(16) Expectations in education	(17) Expectations in life	(18) Prosociality	(19) Internal locus of control	(20) Stable locus of control
Service learning	-0.167* (0.098)	-0.017 (0.072)	0.039 (0.069)	0.105* (0.062)	0.284** (0.129)	-0.117 (0.172)
Effect size	-0.310	-0.030	0.061	0.165	0.275	-0.121
Observations	141	141	141	141	141	141
Mean of dep. var.	0.294	0.296	0.288	0.291	0	7.56e-09
Std. dev. of dep. var.	0.558	0.549	0.586	0.576	1	1

Notes - The results are based on least squares models including randomization blocks. Standard errors in parentheses are clustered at the classroom level.
* $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 4: Effects of Service Learning on Behaviors and School Performance – ITT Estimates

Panel A: Behaviors

Dep. var:	(1) Sincerity behavior	(2) Altruistic behavior (dictator)	(3) Risk attitude (bart)	(4) Adversion to inequality behavior (ultimatum)	(5) Gratitude behavior (gratitude)	(6) Cheating (max choice)
Service learning	-0.169**	0.055	-0.208*	-0.025	0.006	0.044*
	(0.072)	(0.033)	(0.104)	(0.049)	(0.039)	(0.024)
Effect size	*-0.332*	*0.245*	*-0.217*	*-0.088*	*0.029*	*0.317*
Observations	139	140	140	141	140	139
Mean of dep. var.	-0.404	0.406	2.978	0.423	0.454	0.566
Std.Err. of dep. var.	0.585	0.220	0.827	0.259	0.220	0.164

Panel B: School Performance

Dep. var:	Passing to next grade (even with school debts)	Passing to next grade (without school debts)
Service learning	-0.134*	-0.077
	(0.072)	(0.048)
Effect size	*-0.298*	*-0.162*
Observations	158	158
Mean of dep. var.	0.658	0.285
Std. dev. of dep. var.	0.476	0.453

Notes - The results are based on least squares models including randomization blocks. Standard errors in parentheses are clustered at the classroom level.
* p<0.10, ** p<0.05, *** p<0.01.

ATT (LATE) estimates. Considering non-compliance with the treatment of the students involved in the project, for each outcome of interest we estimated the average treatment effect on the treated (ATT), which instruments compliance with the assignment to the treatment. In our case, the effect of the program refers to the students who were fully compliant with the protocol of intervention. We start with a very restrictive definition of compliance, which determines a quite low compliance rate. A student from a school assigned to the treatment is considered as fully treated if (s)he participated in at least 80% of the scheduled hours and meetings as well as attended the meetings for at least 20 hours: following this definition, around 40% of the students assigned to the treatment have been treated[2]. In a robustness analysis, we consider alternative definitions of compliance with the protocol, and demonstrate that our results are robust to these changes (see Section 6).

Tables 5 and 6 present two-stage least squares (2SLS) estimates that use the random assignment as an instrument for actually being treated, following the model described by Equations 2 and 3. The results—which mitigate the selection bias implicit in students' decision to participate in the intervention—represent the LATE of the SL.

---

[2] The low compliance rate resulting from this definition crucially depends on the third condition that we imposed for being considered as full compliant to the protocol (at least 20 hours of treatment). In fact, fifteen out of the 69 students who showed up at least once for their meetings were asked to meet for no more than eighteen hours. When the third condition is relaxed, the compliance rate increases to 65%.

With a first-stage F-statistic that ranges from 28 to 34 (reported at the bottom of Tables 5 and 6), our instrument easily passes conventional thresholds for strong instruments (Stock and Yogo, 2002). The 2SLS estimate reported in column 18 of Table 5 implies that the treated students are 0.37 standard deviations more likely to adopt a pro-social attitude. At the same time, treated students are 0.61 standard deviations more likely to have internal locus of control (see column 19 of Table 5). On the other hand, sincerity behavior is reduced and altruistic behavior and risk aversion are increased due to the intervention (see columns 1, 2 and 3, respectively, of Panel A in Table 6). Similar to what was observed in Table 4, SL reduces the likelihood of passing to the next grade (even with school debt) and has a negative but non-significant effect on a full pass grade (see Panel B in Table 6).

It is also important to highlight that the magnitude of the 2SLS coefficients is significantly larger than that of the equivalent OLS coefficients. This suggests a positive correlation between unobservable drivers of cognitive and non-cognitive skills and the compliance with the protocol. One explanation for the sizable 2SLS estimates that we find is that these refer to the effect of SL for the compliers, i.e., the LATE.

Table 5: Effects of Service Learning on Attitudes – 2SLS Estimates

Dep. var.:	(1) Positive self-esteem	(2) Negative self-esteem	(3) Relational self-efficacy (with peers)	(4) Relational self-efficacy (managing conflicts)	(5) Cognitive self-efficacy (related to school activities)	(6) Cognitive self-efficacy (general)	(7) Emotional self-efficacy (managing emotions)
Compliance with SL	0.205 (0.161)	-0.218 (0.251)	0.171 (0.141)	0.031 (0.141)	0.201 (0.140)	0.038 (0.163)	0.128 (0.193)
Effect size	0.316	-0.245	0.324	0.055	0.316	0.075	0.176
Observations	141	141	141	141	141	141	141
Mean of dep. var.	0.291	0.204	0.306	0.301	0.290	0.309	0.267
Std. dev. of dep. var.	0.573	0.854	0.496	0.524	0.576	0.476	0.676
First stage F statistic	31.47	31.63	31.08	30.51	30.91	31.40	31.36

Dep. var.:	(8) Emotional self-efficacy (getting support)	(9) Attitudes towards school (negative feelings)	(10) Attitudes towards school (school perceived as meaningless)	(11) Lack of motivation in studying	(12) Attitudes towards teachers (positive relationships)	(13) Attitudes towards teachers (feeling persecuted)	(14) Attitudes towards teachers (getting support)
Compliance with SL	0.130 (0.195)	0.201 (0.238)	0.239 (0.226)	0.079 (0.236)	-0.068 (0.177)	-0.271 (0.301)	0.417 (0.270)
Effect size	0.178	0.240	0.327	0.090	-0.116	-0.331	0.483
Observations	141	141	141	141	141	141	141
Mean of dep. var.	0.262	0.229	0.259	0.238	0.285	0.232	0.197
Std. dev. of dep. var.	0.693	0.795	0.701	0.769	0.602	0.785	0.870
First stage F statistic	30.67	34.05	30.02	33.10	30.57	31.19	31.52

Dep. var.:	(15) Attitudes towards teachers (allied with my teachers)	(16) Expectations in education	(17) Expectations in life	(18) Prosociality	(19) Internal locus of control	(20) Stable locus of control
Compliance with SL	-0.372 (0.227)	-0.039 (0.154)	0.087 (0.148)	0.236** (0.119)	0.635** (0.278)	-0.264 (0.361)
Effect size	-0.691	-0.068	0.139	0.370	0.613	-0.273
Observations	141	141	141	141	141	141
Mean of dep. var.	0.294	0.296	0.288	0.291	0	7.56e-09
Std. dev. of dep. var.	0.558	0.549	0.586	0.576	1	1
First stage F statistic	32.80	29.68	31.21	30.94	32.38	32.14

Notes - The results are based on 2SLS models including randomization blocks. Standard errors in parentheses are clustered at the classroom level.
* p<0.10, ** p<0.05, *** p<0.01.

Table 6: Effects of Service Learning on Behaviors and School Performance – 2SLS Estimates

	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Behaviors			
Dep. var.:	Sincerity behavior	Altruistic behavior (dictator)	Risk attitude (bart)	Adversion to inequality behavior (ultimatum)	Gratitude behavior (gratitude)	Cheating (max choice)
Compliance with SL	-0.377**	0.121*	-0.469**	-0.057	0.014	0.098**
	(0.164)	(0.070)	(0.234)	(0.102)	(0.082)	(0.047)
Effect size	-0.742	0.533	-0.489	-0.200	0.065	0.706
Observations	139	140	140	141	140	139
Mean of dep. var.	-0.404	0.406	2.978	0.423	0.454	0.566
Std. dev. of dep. var.	0.585	0.220	0.827	0.259	0.220	0.164
First stage F statistic	28.55	34.18	30.56	29.79	33.15	29.96

	Panel B: School Performance	
Dep. var.:	Passing to next grade (even with school debts)	Passing to next grade (without school debts)
Compliance with SL	-0.340*	-0.195
	(0.198)	(0.122)
Effect size	-0.757	-0.413
Observations	158	158
Mean of dep. var.	0.658	0.285
Std. dev. of dep. var.	0.476	0.453
First stage F statistic	29.06	29.06

Notes - The results are based on 2SLS models including randomization blocks. Standard errors in parentheses are clustered at the classroom level.
* $p<0.10$, ** $p<0.05$, *** $p<0.01$.

6. *Robustness Checks*

We perform a variety of robustness checks to test how the results change when we modify the sample or use a different specification compared to our benchmark model (see Tables 3 and 4). The results of this analysis are reported in the Appendix A.

First, a possible threat to the internal validity of the experiment arises from the risk of spillovers from the treated to the control group, thereby contaminating the RCT. This might be the case if students involved in SL activities communicate with control students in the same class. While we cannot entirely rule out the risk of contamination, we believe that contamination should not have a major impact on our study because SL activities are individual-specific since they typically vary from one student to another. Nevertheless, in Tables A.6 and A.7 we exclude the class "22C"—whose students were all proposed for the program by their teachers—from the sample. In the case of this class, we believe that targeting was not precise, and furthermore we have a higher risk of contamination between treated and control students. Reassuringly, the results are not affected by this exclusion. In additional analyses (not reported, but available upon request), we have verified that our point estimates remain very similar to the baseline specification if we drop—one at a time—classes in which at least one treated and one control student were enrolled.

Second, as displayed in Tables A.8 and A.9, the effect sizes remain mostly unchanged when we estimate Equation (1) including several

additional covariates, namely students' gender, age in months, lower secondary mark, and parental education.

Third, in Tables A.10-A.19, we employ alternative definitions of compliance with the protocol, and show that the overall results are not sensitive to the choice of operationalization of the compliance variable. Specifically, in Tables A.10 and A.11 (A.12 and A.13) a student from a school assigned to the treatment is considered as fully treated if (s)he participated in at least 80% (50%) of the scheduled hours and meetings, resulting in a compliance rate of about 60% (77%), whereas in Tables A.14 and A.15 (A.16 and A.17) we consider as fully treated those students who attended the meetings for at least 20 (10) hours, with the compliance rate being approximately 46% (79%). Finally, in Tables A.18 and A.19, the definition of fully treated refers to students who participated in at least 50% of the scheduled hours and meetings as well as attended the meetings for at least 10 hours, leading to a compliance rate of about 75%.

7. *Discussion and Conclusions*

Our analysis reports both good and bad news about the effectiveness of the investigated SL intervention with low-performing students.

The good news is that our study rigorously confirms that the SL intervention helped in improving the non-cognitive skills of the treated students, all of who were characterized by a high risk of dropping out of school. Our main result shows that SL can substantially raise dis-

advantaged student's self-esteem, pro-social attitude, internal locus of control, and risk aversion. This is a very relevant and new result given that—to our knowledge—no previous research has investigated the causal effect of SL interventions on such disadvantaged population; moreover, non-cognitive skills are often deemed as more important than cognitive ones for creating productive adults (Heckman and Rubinstein, 2001).

The bad news is that "Non solo scuola" had a negative causal impact on the school performance of students, as measured by the percentage of students who passed their grade. This is not in line with previous results concerning the effects of SL on students' achievements, which—although mixed—point in the direction of an overall positive effect (Warren, 2012). Therefore, SL effectiveness as a remedial intervention for low-achieving students should be questioned and further investigated.

Accordingly, the main question arising from this analysis is as follows: Why did the improved psychological and pro-social attitudes of students induced by the intervention not translate into better school performance—as expected—and even prove detrimental for students? The most plausible explanation deals with the specificities of the "Non solo a scuola" intervention. Indeed, students were involved in SL activities during the school time while their peers were attending classes, thereby reducing their participation in the activities of their classmates. In this respect, the "Non solo a scuola" project

should be classified as a co-curricular SL rather than an academic one (Howard, 2003).

In accordance with the established literature, we believe that co-curricular interventions—when adopted for disadvantage populations—could generate some benefits but at the same time jeopardize the overall well-being of students. Our feeling is supported by a set of qualitative interviews that we conducted in the schools involved in the project. Teachers underlined that they did not stop teaching their topics when the treated students were attending the SL program. They confirmed a feeling of general improvement in the attitudes of the treated students towards the school, their classmates, and teachers. Nonetheless, they could not witness an increase in competences linked to the specific disciplines, and they also believed that class non-attendance could have spoiled the expected second-level effect of the SL intervention.

Therefore, leaving class during school hours—while positively influencing most non-cognitive skills of treated students—is very likely to have increased the difficulty of adequately performing in school tasks for low-achieving students and may have further disconnected them from their classmates, thus preventing peer effects to positively influence their school performance. Based on our results, SL interventions aimed at disadvantaged populations should be fully integrated in the school experience.

Furthermore, the co-curricular intervention, and the consequent school leave—despite being officially approved by the school directors—may have produced the feeling among some teachers that the school results of some "difficult students" were no longer their primary responsibility. This feeling may have given teachers an implicit incentive to reduce their effort in implementing interventions aimed at helping disadvantaged students to catch up with the gap induced as a by-product of the SL interventions.

On the other hand, treated students may have reacted to these difficulties by developing attitudes and behaviors that may have reduced their ability to productively attend school. In support of this interpretation, we find that participating in the program reduces students' propensity to be sincere, meaning that behaving in accordance with the school setting (where the questionnaire was administered) was less valuable for treated students.

Given the support that SL programs obtain (recently also by the Italian Ministry of Education) and the consequent widespread adoption in schools, the results of our RCT can have relevant policy implications. Our results show that while benefiting a wide set of non-cognitive skills, SL activities—particularly when they are conceived as co-curricular and not strictly embedded into the school educational mission—may have detrimental effects on the school performance of students, at least for those at risk of dropping out of the education system. When dealing with this target population, developing aca-

demic SL is crucial. Programs should not undertake any activities during school hours, thus avoiding taking students out of their classes. Our suggestion is that SL may also be beneficial regarding school proficiency if protocols are carefully designed and aimed at avoiding any loss of cognitive skills training, while contributing to the development of non-cognitive skills. Overall, our results suggest designing and implementing SL interventions in schools with great care to avoid unintended negative consequences.

References

Abbiati, G., G. Argentin, D. Azzolini, G. Ballarino, and L. Vergolini (2022). "Experimental Research in Education: An Appraisal of the Italian Experience", Swiss Journal of Sociology, 48(1), 21–46.

Algan, Y., C. Chevallier, A. Charpentier, and E. Huillery (2021). "The Role of Mindset in Education: Evidence from a Large-Scale Field Experiment in Disadvantaged Schools" SocArXivzs 9aq, Center for Open Science.

Ariely, D., X. Garcia-Rada, K. Gödker, et al. (2019). The impact of two different economic systems on dishonesty. European Journal of Political Economy, 59, 179-195.

Aronson, K. R., N. S. Webster, R. Reason, et al. (2005). Using randomized control field trials in service-learning research. In S. Root, J. Callahan, & S. H. Billig, (Eds.), Improving service-learning practice (pp. 141–165). Greenwich, CT: Information Age Publishing.

Berg, J., J. Dickhaut, and K. McCabe (1995). Trust, reciprocity, and social history. Games and Economic Behavior, 10(1), 122-142.

Billig S., S. Root, and D. Jesse D. (2005). The impact of participation in service-learning on high school students' civic engagement. CIRCLE Working Paper 33. College Park, MD: Center for Information and Research on Civic Learning and Engagement (CIRCLE).

Billig, S. H. (2009). Does quality really matter: Testing the new K–12 service-learning standards for quality practice. In B. E. Moely,

S. H. Billig, and B. A. Holland (Eds.), Advances in service-learning research: Vol. 9. Creating our identities in service-learning and community engagement (pp. 131–158). Greenwich, CT: Information Age Publishing.

Bringle, R.G., and J. A. Hatcher (1996). Implementing service learning in higher education. The Journal of Higher Education, 67(2), 221-239.

Carlana, M., E. La Ferrara, and P. Pinotti (2020). Goals and Gaps: Educational Careers of Immigrant Children. Econometrica, forthcoming.

Celio, C.I., J. Durlak, and A. Dymnicki (2011). A Meta-analysis of the Impact of Service-Learning on Students. Journal of Experiential Education, 34(2), 164-181.

Chua, C. C. (2010). Rethinking community-service education in Singapore schools. Social Space. 94(7).

https://ink.library.smu.edu. sg/lien_research/67

Conway, J. M., E.L. Amel, and D.P. Gerwien (2009). Teaching and learning in the social context: A meta-analysis of service learning's effects on academic, personal, social, and citizenship outcomes. Teaching of Psychology, 36, 233–245.

Eurostat (2018). Education and Training MONITOR 2018 Italy. Link:

https://ec.europa.eu/education/sites/education/files/document-library-docs/et-monitor-report-2018-italy_en.pdf

Filges, T., J. Dietrichson, B. C. A. Viinholt, and N. T. Dalgaard (2021). PROTOCOL: Service learning for improving academic success in students in grade K to 12: a systematic review. Campbell Systematic Reviews, 17, e1157. https://doi.org/10.1002/cl2.1157.

Forsythe, R., J. L. Horowitz, N. E. Savin, and M. Sefton (1994). Fairness in simple bargaining experiments. Games and Economic Behavior 6(3), 347-369.

Furco, A. (2003). Issues of definition and program diversity in the study of service-learning. In S. H. Billig, and A. S. Waterman (Eds.), Studying service-learning: innovations in education research methodology (pp. 26–44). Mahwah, NJ: Lawrence Erlbaum Associates, Inc. Publishers.

Furco, A. (2010). The community as a resource for learning: An analysis of academic service-learning in primary and secondary education. In H. Dumont, D. Istance and F. Benavides (Eds.), The nature of learning: Using research to inspire practice. Paris: Educational Research and Innovation, OECD Publishing.

https://doi.org/10.1787/9789264086487-en

Ghanem, D., S. Hirshleifer, and K. Ortiz-Becerra. 2021. Testing attrition bias in field experiments. CEGA Working Paper Series No. WPS-113. Center for Effective Global Action. University of California, Berkeley. https://doi.org/10.26085/C38C76

Giles D. E., and J. Eyler (1994). The impact of a college community service laboratory on students' personal, social, and cognitive outcomes. Journal of Adolescence, 17(4), 327-339.

Goux, D., M. Gurgand, and E. Maurin (2017). Adjusting your dreams? The effect of school and peers on dropout behaviour. Economic Journal, 127(602), 1025–1046.

Güth, W., R. Schmittberger, and B. Schwarze (1982). An experimental analysis of ultimatum bargaining, Journal of Economic Behavior and Organization 3, 367-388.

Harwood, A. M., and S. A. Radoff (2009). Reciprocal benefits of mentoring: Results of a middle school–university collaboration. In B. E. Moely, S. H. Billig, & B. A. Holland (Eds.), Advances in service-learning research: Vol 9. Creating our identities in service-learning and community engagement (pp. 131–158). Greenwich, CT: Information Age.

Hébert, A. and P. Hauf (2015). Student learning through service learning: Effects on academic development, civic responsibility, interpersonal skills and practical skills. Active Learning in Higher Education, 16(1), 37–49.

Hecht, D. (2003). Issues of research design and statistical analysis. In S. H. Billig, & A. S. Waterman (Eds.), Studying service-learning: innovations in education research methodology (pp. 95–110). Mahwah, NJ: Lawrence Erlbaum Associates, Inc. Publishers.

Heckman, J. J., and Y. Rubinstein (2001). The Importance of Non-cognitive Skills: Lessons from the GED Testing Program. American Economic Review, 91(2), 145-149.

Henderson, A., C. Marburger, and T. Ooms (1986). Beyond the bake sale: An educator's guide to working with parents, Columbia, Md: National Committee for Citizens in Education.

Henderson, A. and N. Berler (1995). A new generation of evidence: The family is critical to student achievement, Washington, DC: National Committee for Citizens in Education

Howard, J. (2003). Service-Learning research: Foundational issues. In S. H. Billig, & A. S. Waterman (Eds.), Studying service-learning: innovations in education research methodology (pp. 16–25). Mahwah, NJ: Lawrence Erlbaum Associates, Inc. Publishers.

Ierullo, M. (2016). The institutionalization of service-learning projects in Argentine schools. International Journal of Research on Service-Learning and Community Engagement, 4(1), 351–354.

Kahneman, D., J. L. Knetsch, and R. Thaler (1986). Fairness as a constraint on profit seeking: Entitlements in the market. The American economic review, 76(4), 728-741.

Lejuez, C. W., J. P. Read, C. W. Kahler et al. (2002). Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). Journal of Experimental Psychology: Applied, 8(2), 75.

Lester, S. W., C. Tomkovick, T. Wells et al. (2005). Does service-learning add value? Examining the perspectives of multiple stake-holders. The Academy of Management Learning and Education, 4(3), 278–294.

Leung, A. Y. M., S. S. C. Chan, C. W. Kwan, et al. (2012). Service learning in medical and nursing training: A randomized controlled trial. Advances in Health Sciences Education, 17(4), 529–545.

Maples, A. E., A. Williams-Wengerd, J. E. Braughton, K. L. Henry, et al. (2020). The role of service-learning experiences in promoting flourishing among college-student youth mentors. The Journal of Positive Psychology, Published online: 09 Dec 2020.

Markus, G. B., J. P. Howard, and D. C. King (1993). Integrating community service and classroom instruction enhances learning: Results from an experiment. Educational Evaluation and Policy Analysis, 15, 410–419.

Martins, P. S. (2010). Can targeted, non-cognitive skills programs improve achievement? Evidence from EPIS. CEG-IST and IZA Discussion Paper No. 5266.

McNatt D. B. (2019). Enhancing public speaking confidence, skills, and performance: An experiment of service-learning. The International Journal of Management Education, 17, 276-285.

MIUR (2018). Una via italiana per il service learning.

https://www.miur.gov.it/web/guest/-/una-via-italiana-per-il-service-learning

Moely, B. E., M. McFarland, D. Miron, et al. (2002). Changes in college students' attitudes and intentions for civic involvement as a function of service-learning experiences. Michigan Journal of Community Service Learning, 9, 18–26.

National Youth Leadership Council. (2011). K–12 service-learning standards for quality practice. Retrieved from https://www.nylc.org/page/standards

Novak, J., V. Markey, and M. Allen (2007). Evaluating Cognitive Outcomes of Service Learning in Higher Education: A Meta-Analysis. Communication Research Reports, 24(2), 149-157.

Papamarcos, S. D. (2005). Giving traction to management theory: Today's service-learning. The Academy of Management Learning and Education, 4(3), 325–335.

Perold, H., and M. N. Tapia (Eds.). (2008). Service enquiry: Civic service and volunteering in Latin America and the Caribbean (Vol. 2). Buenos Aires: Centro Latinoamericano de Aprendizaje y Servicio Solidario; Johannesburg: Volunteer and Service Enquiry Southern Africa; Washington University in St Louis: The Center for Social Development; Washington DC: Innovations in Civic Participation.

Poon, P., T. S. Chan, and L. Zhou (2011). Implementation of service-learning in business education: Issues and challenges. Journal of Teaching in International Business, 22, 185–192.

Rodriguez-Planas, N. (2012). Longer-term impacts of mentoring, educational services, and learning incentives: Evidence from a randomized trial in the United States. American Economic Journal: Applied Economics, 4(4):121–139.

Scales, P. C., and E. Roehlkepartain (2004). Community service and service-learning in US public schools, 2004: Findings from a national survey. National Youth Leadership Council, Vienna, VA.

Stock, J. H. and M. Yogo (2002). Testing for weak instruments in linear IV regression. NBER Technical Working Papers.

Warren, J. L., (2012), Does Service-Learning Increase Student Learning?: A Meta-Analysis. Michigan Journal of Community Service Learning, Spring:56-61.

Appendix A: Supplemental Tables

Table A1 - Test for Differential Attrition

Dep. var.:	(1) Attrition
Service learning	0.031
	(0.056)
Observations	158
Mean of dep. var.	0.108
Std. dev. of dep. var.	0.311

Notes - The results are based on least squares models including randomization blocks. Standard errors in parentheses are clustered at the classroom level.
* p<0.10, ** p<0.05, *** p<0.01.

Table A2 - Determinants of Attrition Test – Baseline Covariates

	(1)	(2)
	Response status	
	Coefficient	p-value
Male	-0.095	0.366
Year of birth	0.043	0.807
Born in Italy	-0.130	0.166
Father is present	0.131	0.142
Mother is present	0.057	0.400
Parents are present	0.159*	0.069
Grade 9	0.094	0.306
Pass grade in lower secondary school	-0.151	0.273
Satisfactory or good grade in lower secondary education	0.151	0.273
One hundred or more books at home	0.044	0.793
Lost years of schooling	-0.106	0.329
Parental years of schooling	0.082	0.813

Notes - Columns 1 and 2 report the coefficient of the response status derived by reverse regressions of the pre-intervention covariates listed in each row on response status.

Table A3 - Determinants of Attrition Test – Outcomes at Baseline

	(1)	(2)
	Response status	
	Coefficient	p-value
Positive self-esteem	-0.273	0.381
Negative self-esteem	0.164	0.567
Relational self-efficacy (with peers)	-0.103	0.682
Relational self-efficacy (managing conflicts)	0.153	0.479
Cognitive self-efficacy (related to school activities)	-0.428	0.101
Cognitive self-efficacy (general)	0.159	0.558
Emotional self-efficacy (managing emotions)	-0.155	0.573
Emotional self-efficacy (getting support)	-0.409	0.148
Attitudes towards school (negative feelings)	0.412	0.147
Attitudes towards school (school perceived as meaningless)	0.055	0.831
Lack of motivation in studying	0.690***	0.001
Attitudes towards teachers (positive relationships)	-0.211	0.358
Attitudes towards teachers (feeling persecuted)	0.589**	0.014
Attitudes towards teachers (getting support)	-0.207	0.233
Attitudes towards teachers (allied with my teachers)	-0.343	0.267
Expectations in education	-0.093	0.782
Expectations in life	-0.446*	0.091
Prosociality	-0.036	0.916
Internal locus of control	0.077	0.772
Stable locus of control	-0.339	0.321
Sincerity behavior	-0.409***	0.001
Altruistic behavior (dictator)	-0.016	0.702
Risk attitude (bart)	0.136	0.650
Adversion to inequality behavior (ultimatum)	0.071*	0.068
Gratitude behavior (gratitude)	-0.009	0.805
Cheating (max choice)	0.107**	0.013

Notes - Columns 1 and 2 report the coefficient of the response status derived by reverse regressions of the pre-intervention outcomes listed in each row on response status.

Table A4 - Selective Attrition Test – Baseline Covariates

	(1)	(2)	(3)	(4)
	Service learning			
	Respondents (N=141)		Attritors (N=17)	
	Coefficient	*p*-value	Coefficient	*p*-value
Male	0.139	0.102	-0.235	0.426
Year of birth	-0.270*	0.055	0.118	0.859
Born in Italy	-0.016	0.746	0.059	0.856
Father is present	0.084	0.227	-0.000	1
Mother is present	-0.017	0.704	0.176	0.501
Parents are present	0.052	0.452	0.176	0.623
Grade 9	-0.127**	0.049	0.000	0.26
Pass grade in lower secondary school	0.022	0.796	0.706**	0.024
Satisfactory or good grade in lower secondary education	-0.022	0.796	-0.706**	0.024
One hundred or more books at home	0.104	0.273	0.529**	0.015
Lost years of schooling	0.125	0.239	0.059	0.902
Parental years of schooling	0.050	0.835	-0.676	0.5

Notes - This table reports the coefficient of the response status derived by reverse regressions of the pre-intervention covariates listed in each row on SL program for the respondents (columns 1 and 2) and attritors (columns 3 and 4).

Table A5 - Selective Attrition Test – Baseline Outcomes

	(1)	(2)	(3)	(4)
	Service learning			
	Respondents (N=141)		Attritors (N=17)	
	Coefficient	p-value	Coefficient	p-value
Positive self-esteem	-0.193	0.310	0.018	0.985
Negative self-esteem	-0.055	0.735	-0.363	0.745
Relational self-efficacy (with peers)	0.118	0.532	0.027	0.973
Relational self-efficacy (managing conflicts)	0.174	0.299	0.577	0.341
Cognitive self-efficacy (related to school activities)	-0.044	0.751	-0.149	0.761
Cognitive self-efficacy (general)	0.137	0.410	-0.412	0.642
Emotional self-efficacy (managing emotions)	0.053	0.683	1.143*	0.054
Emotional self-efficacy (getting support)	0.002	0.987	0.802	0.334
Attitudes towards school (negative feelings)	-0.380**	0.011	0.476	0.471
Attitudes towards school (school perceived as meaningless)	-0.200	0.150	-0.130	0.848
Lack of motivation in studying	-0.196	0.209	-0.056	0.851
Attitudes towards teachers (positive relationships)	-0.135	0.427	-0.689	0.271
Attitudes towards teachers (feeling persecuted)	0.088	0.596	0.710	0.107
Attitudes towards teachers (getting support)	0.042	0.808	-0.668	0.193
Attitudes towards teachers (allied with my teachers)	0.137	0.356	-1.218	0.109
Expectations in education	-0.252	0.150	0.635	0.613
Expectations in life	-0.052	0.734	0.513	0.441
Prosociality	0.104	0.572	0.095	0.859
Internal locus of control	0.201	0.124	-0.691	0.331
Stable locus of control	-0.067	0.631	0.148	0.788
Sincerity behavior	0.002	0.985	-0.180	0.524
Altruistic behavior (dictator)	0.054*	0.080	0.139**	0.033
Risk attitude (bart)	0.009	0.932	-0.102	0.893
Adversion to inequality behavior (ultimatum)	0.044	0.225	0.064	0.352
Gratitude behavior (gratitude)	0.003	0.931	-0.106	0.136
Cheating (max choice)	0.002	0.923	0.083	0.357

Notes - This table reports the coefficient of the response status derived by reverse regressions of the pre-intervention outcomes listed in each row on SL program for the respondents (columns 1 and 2) and attritors (columns 3 and 4).

Table A6: Effects of Service Learning on Attitudes – ITT Estimates – Drop Class "22C"

Dep. var.:	(1) Positive self-esteem	(2) Negative self-esteem	(3) Relational self-efficacy (with peers)	(4) Relational self-efficacy (managing conflicts)	(5) Cognitive self-efficacy (related to school activities)	(6) Cognitive self-efficacy (general)	(7) Emotional self-efficacy (managing emotions)
Service learning	0.123 (0.084)	-0.122 (0.132)	0.044 (0.068)	0.012 (0.074)	0.098 (0.076)	-0.023 (0.078)	0.055 (0.108)
Effect size	0.177	-0.131	0.082	0.021	0.152	-0.045	0.073
Observations	123	123	123	123	123	123	123
Mean of dep. var.	0.289	0.189	0.304	0.312	0.230	0.286	0.234
Std. dev. of dep. var.	0.608	0.884	0.509	0.538	0.576	0.478	0.699

Dep. var.:	(8) Emotional self-efficacy (getting support)	(9) Attitudes towards school (negative feelings)	(10) Attitudes towards school (school perceived as meaningless)	(11) Lack of motivation in studying	(12) Attitudes towards teachers (positive relationships)	(13) Attitudes towards teachers (feeling persecuted)	(14) Attitudes towards teachers (getting support)
Service learning	0.083 (0.106)	0.012 (0.096)	0.032 (0.088)	0.025 (0.127)	-0.007 (0.091)	-0.228* (0.114)	0.200 (0.142)
Effect size	0.111	0.015	0.046	0.030	-0.013	-0.297	0.231
Observations	123	123	123	123	123	123	123
Mean of dep. var.	0.220	0.283	0.337	0.291	0.210	0.333	0.131
Std. dev. of dep. var.	0.713	0.784	0.674	0.726	0.567	0.747	0.872

Dep. var.:	(15) Attitudes towards teachers (allied with my teachers)	(16) Expectations in education	(17) Expectations in life	(18) Prosociality	(19) Internal locus of control	(20) Stable locus of control
Service learning	-0.107 (0.091)	-0.041 (0.079)	0.060 (0.076)	0.123* (0.070)	0.300** (0.146)	-0.098 (0.198)
Effect size	-0.200	-0.070	0.093	0.187	0.290	-0.105
Observations	123	123	123	123	123	123
Mean of dep. var.	0.292	0.338	0.240	0.299	-0.0322	0.0584
Std. dev. of dep. var.	0.533	0.560	0.992	0.599	0.994	0.972

Notes – The results are based on least squares models including randomization blocks. Standard errors in parentheses are clustered at the classroom level.
* p<0.10, ** p<0.05, *** p<0.01.

Table A7: Effects of Service Learning on Behaviors and School Performance – ITT Estimates – Drop Class "22C"

Dep. var.:	(1)	(2)	(3)	(4)	(5)	(6)
	Sincerity behavior	Altruistic behavior (dictator)	Risk attitude (bart)	Adversion to inequality behavior (ultimatum)	Gratitude behavior (gratitude)	Cheating (max choice)
			Panel A: Behaviors			
Service learning	-0.174**	0.072**	-0.189	0.003	0.023	0.048*
	(0.082)	(0.034)	(0.118)	(0.045)	(0.042)	(0.027)
Effect size	-0.344	0.328	-0.215	0.012	0.106	0.340
Observations	122	122	122	123	122	122
Mean of dep. var.	-0.401	0.391	2.921	0.415	0.430	0.567
Std. dev. of dep. var.	0.572	0.220	0.766	0.260	0.221	0.168

Dep. var.:	Passing to next grade (even with school debts)	Passing to next grade (without school debts)
		Panel B: School Performance
Service learning	-0.140*	-0.074
	(0.083)	(0.055)
Effect size	-0.300	-0.177
Observations	137	137
Mean of dep. var.	0.613	0.182
Std. dev. of dep. var.	0.489	0.388

Notes - The results are based on least squares models including randomization blocks. Standard errors in parentheses are clustered at the classroom level.
* $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A8: Effects of Service Learning on Attitudes – ITT Estimates – Adding Controls

Dep. var.:	(1) Positive self-esteem	(2) Negative self-esteem	(3) Relational self-efficacy (with peers)	(4) Relational self-efficacy (managing conflicts)	(5) Cognitive self-efficacy (related to school activities)	(6) Cognitive self-efficacy (general)	(7) Emotional self-efficacy (managing emotions)
Service learning	0.049 (0.079)	-0.054 (0.139)	0.048 (0.071)	-0.021 (0.066)	0.064 (0.066)	0.009 (0.078)	0.118 (0.102)
Effect size	0.076	-0.061	0.091	-0.038	0.100	0.017	0.162
Observations	141	141	141	141	141	141	141
Mean of dep. var.	0.291	0.204	0.306	0.301	0.290	0.309	0.267
Std. dev. of dep. var.	0.573	0.854	0.496	0.524	0.576	0.476	0.676

Dep. var.:	(8) Emotional self-efficacy (getting support)	(9) Attitudes towards school (negative feelings)	(10) Attitudes towards school (school perceived as meaningless)	(11) Lack of motivation in studying	(12) Attitudes towards teachers (positive relationships)	(13) Attitudes towards teachers (feeling persecuted)	(14) Attitudes towards teachers (getting support)
Service learning	-0.001 (0.094)	0.152 (0.134)	0.167 (0.125)	0.060 (0.125)	-0.101 (0.077)	-0.042 (0.170)	0.112 (0.120)
Effect size	-0.001	0.182	0.229	0.069	-0.173	-0.051	0.130
Observations	141	141	141	141	141	141	141
Mean of dep. var.	0.262	0.229	0.259	0.238	0.285	0.232	0.197
Std. dev. of dep. var.	0.693	0.795	0.701	0.769	0.602	0.785	0.870

Dep. var.:	(15) Attitudes towards teachers (allied with my teachers)	(16) Expectations in education	(17) Expectations in life	(18) Prosociality	(19) Internal locus of control	(20) Stable locus of control
Service learning	-0.223** (0.100)	-0.115 (0.090)	0.040 (0.072)	0.120* (0.067)	0.253* (0.145)	-0.014 (0.171)
Effect size	-0.414	-0.201	0.063	0.188	0.244	-0.014
Observations	141	141	141	141	141	141
Mean of dep. var.	0.294	0.296	0.288	0.291	0	7.55e-09
Std. dev. of dep. var.	0.558	0.549	0.586	0.576	1	1

Notes – The results are based on least squares models including randomization blocks, as well as controls for gender, age (in months), lower secondary mark and parental education. Standard errors in parentheses are clustered at the classroom level.

* $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A9: Effects of Service Learning on Behaviors and School Performance – ITT Estimates – Adding Controls

Dep. var:	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Behaviors			
	Sincerity behavior	Altruistic behavior (dictator)	Risk attitude (bart)	Adversion to inequality behavior (ultimatum)	Gratitude behavior (gratitude)	Cheating (max choice)
Service learning	-0.204**	0.049	-0.237*	-0.016	-0.013	0.052*
	(0.077)	(0.035)	(0.118)	(0.049)	(0.037)	(0.029)
Effect size	-0.401	0.215	-0.247	-0.057	-0.057	0.379
Observations	139	140	140	141	140	139
Mean of dep. var.	-0.404	0.406	2.978	0.423	0.454	0.566
Std. dev. of dep. var.	0.585	0.220	0.827	0.259	0.220	0.164

Dep. var:		
		Panel B: School Performance
	Passing to next grade (even with school debts)	Passing to next grade (without school debts)
Service learning	-0.132	-0.048
	(0.080)	(0.048)
Effect size	-0.294	-0.102
Observations	158	158
Mean of dep. var.	0.658	0.285
Std. dev. of dep. var.	0.476	0.453

Notes - The results are based on least squares models including randomization blocks, as well as controls for gender, age (in months), lower secondary mark and parental education. Standard errors in parentheses are clustered at the classroom level.
* p<0.10, ** p<0.05, *** p<0.01.

Table A10: Effects of Service Learning on Attitudes – 2SLS Estimates – Alternative Definition of Compliance with the Protocol

Dep. var.:	(1) Positive self-esteem	(2) Negative self-esteem	(3) Relational self-efficacy (with peers)	(4) Relational self-efficacy (managing conflicts)	(5) Cognitive self-efficacy (related to school activities)	(6) Cognitive self-efficacy (general)	(7) Emotional self-efficacy (managing emotions)
Compliance with SL	0.135 (0.112)	-0.142 (0.161)	0.112 (0.093)	0.020 (0.093)	0.131 (0.095)	0.024 (0.106)	0.084 (0.132)
Effect size	0.208	-0.160	0.212	0.036	0.206	0.048	0.115
Observations	141	141	141	141	141	141	141
Mean of dep. var.	0.291	0.204	0.306	0.301	0.290	0.309	0.267
Std. dev. of dep. var.	0.573	0.854	0.496	0.524	0.576	0.476	0.676
First stage F statistic	107	117.6	118.7	114.5	111.4	113.4	116.2

Dep. var.:	(8) Emotional self-efficacy (getting support)	(9) Attitudes towards school (negative feelings)	(10) Attitudes towards school (school perceived as meaningless)	(11) Lack of motivation in studying	(12) Attitudes towards teachers (positive relationships)	(13) Attitudes towards teachers (feeling persecuted)	(14) Attitudes towards teachers (getting support)
Compliance with SL	0.085 (0.132)	0.132 (0.152)	0.154 (0.144)	0.052 (0.154)	-0.044 (0.112)	-0.176 (0.198)	0.272 (0.182)
Effect size	0.116	0.158	0.211	0.059	-0.076	-0.215	0.316
Observations	141	141	141	141	141	141	141
Mean of dep. var.	0.262	0.229	0.259	0.238	0.285	0.232	0.197
Std. dev. of dep. var.	0.693	0.795	0.701	0.769	0.602	0.785	0.870
First stage F statistic	116.1	130.9	110	115.6	110.7	113.1	124.9

Dep. var.:	(15) Attitudes towards teachers (allied with my teachers)	(16) Expectations in education	(17) Expectations in life	(18) Prosociality	(19) Internal locus of control	(20) Stable locus of control
Compliance with SL	-0.245* (0.136)	-0.025 (0.101)	0.057 (0.097)	0.154* (0.083)	0.420** (0.191)	-0.173 (0.243)
Effect size	-0.454	-0.044	0.091	0.242	0.405	-0.179
Observations	141	141	141	141	141	141
Mean of dep. var.	0.294	0.296	0.288	0.291	0	7.56e-09
Std. dev. of dep. var.	0.558	0.549	0.586	0.576	1	1
First stage F statistic	122.5	113.7	114.5	115.2	118.1	108.6

Notes – The results are based on 2SLS models including randomization blocks. A student from a school is considered as actually treated if s(he) participated in at least 80% of the scheduled hours and meetings. Standard errors in parentheses are clustered at the classroom level.
* $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A11: Effects of Service Learning on Behaviors and School Performance – 2SLS Estimates – Alternative Definition of Compliance with the Protocol

Dep. var.:	(1)	(2)	(3)	(4)	(5)	(6)
	Sincerity behavior	Altruistic behavior (dictator)	Risk attitude (bart)	Aversion to inequality behavior (ultimatum)	Gratitude behavior (gratitude)	Cheating (max choice)
			Panel A: Behaviors			
Compliance with SL	-0.251**	0.081*	-0.306**	-0.037	0.010	0.065*
	(0.108)	(0.047)	(0.147)	(0.068)	(0.054)	(0.036)
Effect size	-0.494	0.357	-0.319	-0.131	0.043	0.470
Observations	139	140	140	141	140	139
Mean of dep. var.	-0.404	0.406	2.978	0.423	0.454	0.566
Std. dev. of dep. var.	0.585	0.220	0.827	0.259	0.220	0.164
First stage F statistic	116.1	128.7	118.5	107.9	114.6	116.5

Dep. var.:	Passing to next grade (even with school debts)	Passing to next grade (without school debts)
	Panel B: School Performance	
Compliance with SL	-0.223*	-0.128
	(0.121)	(0.081)
Effect size	-0.497	-0.271
Observations	158	158
Mean of dep. var.	0.658	0.285
Std. dev. of dep. var.	0.476	0.453
First stage F statistic	84.49	84.49

Notes – The results are based on 2SLS models including randomization blocks. A student from a school is considered as actually treated if s(he) participated in at least 80% of the scheduled hours and meetings. Standard errors in parentheses are clustered at the classroom level.
* $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A12: Effects of Service Learning on Attitudes – 2SLS Estimates – Alternative Definition of Compliance with the Protocol

Dep. var.:	(1) Positive self-esteem	(2) Negative self-esteem	(3) Relational self-efficacy (with peers)	(4) Relational self-efficacy (managing conflicts)	(5) Cognitive self-efficacy (related to school activities)	(6) Cognitive self-efficacy (general)	(7) Emotional self-efficacy (managing emotions)
Compliance with SL	0.110 (0.091)	-0.117 (0.133)	0.092 (0.076)	0.017 (0.076)	0.107 (0.078)	0.020 (0.087)	0.069 (0.108)
Effect size	0.170	-0.131	0.175	0.029	0.169	0.040	0.094
Observations	141	141	141	141	141	141	141
Mean of dep. var.	0.291	0.204	0.306	0.301	0.290	0.309	0.267
Std. dev. of dep. var.	0.573	0.854	0.496	0.524	0.576	0.476	0.676
First stage F statistic	206.8	212.3	213	216.7	208	210.4	215

Dep. var.:	(8) Emotional self-efficacy (getting support)	(9) Attitudes towards school (negative feelings)	(10) Attitudes towards school (school perceived as meaningless)	(11) Lack of motivation in studying	(12) Attitudes towards teachers (positive relationships)	(13) Attitudes towards teachers (feeling persecuted)	(14) Attitudes towards teachers (getting support)
Compliance with SL	0.070 (0.108)	0.107 (0.123)	0.124 (0.114)	0.043 (0.126)	-0.036 (0.092)	-0.145 (0.164)	0.223 (0.147)
Effect size	0.095	0.128	0.170	0.049	-0.062	-0.177	0.258
Observations	141	141	141	141	141	141	141
Mean of dep. var.	0.262	0.229	0.259	0.238	0.285	0.232	0.197
Std. dev. of dep. var.	0.693	0.795	0.701	0.769	0.602	0.785	0.870
First stage F statistic	213	265.9	249.5	225.9	193.7	193	250.9

Dep. var.:	(15) Attitudes towards teachers (allied with my teachers)	(16) Expectations in education	(17) Expectations in life	(18) Prosociality	(19) Internal locus of control	(20) Stable locus of control
Compliance with SL	-0.201* (0.110)	-0.021 (0.082)	0.047 (0.079)	0.126* (0.069)	0.341** (0.150)	-0.141 (0.199)
Effect size	-0.373	-0.036	0.074	0.198	0.329	-0.147
Observations	141	141	141	141	141	141
Mean of dep. var.	0.294	0.296	0.288	0.291	0	7.55e-09
Std. dev. of dep. var.	0.558	0.549	0.586	0.576	1	1
First stage F statistic	220.5	219.9	205.7	213.3	227.1	200.5

Notes – The results are based on 2SLS models including randomization blocks. A student from a school is considered as actually treated if s(he) participated in at least 50% of the scheduled hours and meetings. Standard errors in parentheses are clustered at the classroom level.
* p<0.10, ** p<0.05, *** p<0.01.

Table A13: Effects of Service Learning on Behaviors and School Performance – 2SLS Estimates – Alternative Definition of Compliance with the Protocol

Dep. var:	(1) Sincerity behavior	(2) Altruistic behavior (dictator)	(3) Risk attitude (bart)	(4) Adversion to inequality behavior (ultimatum)	(5) Gratitude behavior (gratitude)	(6) Cheating (max choice)
			Panel A: Behaviors			
Compliance with SL	-0.205** (0.085)	0.066* (0.037)	-0.250** (0.120)	-0.030 (0.055)	0.008 (0.044)	0.053* (0.028)
Effect size	-0.403	0.290	-0.261	-0.107	0.035	0.384
Observations	139	140	140	141	140	139
Mean of dep. var.	-0.404	0.406	2.978	0.423	0.454	0.566
Std. dev. of dep. var.	0.585	0.220	0.827	0.259	0.220	0.164
First stage F statistic	213.8	293.6	221	206.6	204.1	213.4

Dep. var:	Passing to next grade (even with school debts)	Passing to next grade (without school debts)
	Panel B: School Performance	
Compliance with SL	-0.175* (0.093)	-0.100 (0.062)
Effect size	-0.390	-0.213
Observations	158	158
Mean of dep. var.	0.658	0.285
Std. dev. of dep. var.	0.476	0.453
First stage F statistic	194.7	194.7

Notes – The results are based on 2SLS models including randomization blocks. A student from a school is considered as actually treated if s(he) participated in at least 50% of the scheduled hours and meetings. Standard errors in parentheses are clustered at the classroom level.
* p<0.10, ** p<0.05, *** p<0.01.

Table A14: Effects of Service Learning on Attitudes – 2SLS Estimates – Alternative Definition of Compliance with the Protocol

Dep. var:	(1) Positive self-esteem	(2) Negative self-esteem	(3) Relational self-efficacy (with peers)	(4) Relational self-efficacy (managing conflicts)	(5) Cognitive self-efficacy (related to school activities)	(6) Cognitive self-efficacy (general)	(7) Emotional self-efficacy (managing emotions)
Compliance with SL	0.180 (0.140)	-0.192 (0.221)	0.151 (0.126)	0.027 (0.124)	0.176 (0.122)	0.033 (0.143)	0.113 (0.169)
Effect size	0.277	-0.215	0.286	0.048	0.276	0.065	0.154
Observations	141	141	141	141	141	141	141
Mean of dep. var.	0.291	0.204	0.306	0.301	0.290	0.309	0.267
Std. dev. of dep. var.	0.573	0.854	0.496	0.524	0.576	0.476	0.676
First stage F statistic	34.67	34.53	34.11	33.02	34.38	34.06	34.14

Dep. var:	(8) Emotional self-efficacy (getting support)	(9) Attitudes towards school (negative feelings)	(10) Attitudes towards school (school perceived as meaningless)	(11) Lack of motivation in studying	(12) Attitudes towards teachers (positive relationships)	(13) Attitudes towards teachers (feeling persecuted)	(14) Attitudes towards teachers (getting support)
Compliance with SL	0.114 (0.171)	0.176 (0.210)	0.207 (0.199)	0.069 (0.207)	-0.059 (0.155)	-0.238 (0.261)	0.366 (0.232)
Effect size	0.156	0.210	0.284	0.079	-0.102	-0.291	0.424
Observations	141	141	141	141	141	141	141
Mean of dep. var.	0.262	0.229	0.259	0.238	0.285	0.232	0.197
Std. dev. of dep. var.	0.693	0.795	0.701	0.769	0.602	0.785	0.870
First stage F statistic	33.86	37.45	32.87	36.12	33.90	34.28	34.35

Dep. var:	(15) Attitudes towards teachers (allied with my teachers)	(16) Expectations in education	(17) Expectations in life	(18) Prosociality	(19) Internal locus of control	(20) Stable locus of control
Compliance with SL	-0.327 (0.199)	-0.034 (0.135)	0.077 (0.129)	0.206* (0.106)	0.556** (0.236)	-0.231 (0.315)
Effect size	-0.607	-0.060	0.121	0.324	0.537	-0.239
Observations	141	141	141	141	141	141
Mean of dep. var.	0.294	0.296	0.288	0.291	0	7.5e-09
Std. dev. of dep. var.	0.558	0.549	0.586	0.576	1	1
First stage F statistic	35.63	31.87	34.18	33.77	35.44	36

Notes – The results are based on 2SLS models including randomization blocks. A student from a school is considered as actually treated if s(he) attended the meetings for at least 20 hours. Standard errors in parentheses are clustered at the classroom level.
* p<0.10, ** p<0.05, *** p<0.01.

64

Table A15: Effects of Service Learning on Behaviors and School Performance – 2SLS Estimates – Alternative Definition of Compliance with the Protocol

	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Behaviors			
Dep. var.:	Sincerity behavior	Altruistic behavior (dictator)	Risk attitude (bart)	Adversion to inequality behavior (ultimatum)	Gratitude behavior (gratitude)	Cheating (max choice)
Compliance with SL	-0.331**	0.106*	-0.411**	-0.050	0.013	0.086**
	(0.136)	(0.061)	(0.205)	(0.090)	(0.072)	(0.041)
Effect size	-0.650	0.466	-0.429	-0.174	0.057	0.620
Observations	139	140	140	141	140	139
Mean of dep. var.	-0.404	0.406	2.978	0.423	0.454	0.566
Std. dev. of dep. var.	0.585	0.220	0.827	0.259	0.220	0.164
First stage F statistic	32.52	38.24	33.33	33.23	35.91	33.36

	Panel B: School Performance	
Dep. var.:	Passing to next grade (even with school debts)	Passing to next grade (without school debts)
Compliance with SL	-0.292*	-0.167
	(0.170)	(0.103)
Effect size	-0.650	-0.354
Observations	158	158
Mean of dep. var.	0.658	0.285
Std. dev. of dep. var.	0.476	0.453
First stage F statistic	32.63	32.63

Notes - The results are based on 2SLS models including randomization blocks. A student from a school is considered as actually treated if s(he) attended the meetings for at least 20 hours. Standard errors in parentheses are clustered at the classroom level.
* p<0.10, ** p<0.05, *** p<0.01.

Table A16: Effects of Service Learning on Attitudes – 2SLS Estimates – Alternative Definition of Compliance with the Protocol

Dep. var.:	(1) Positive self-esteem	(2) Negative self-esteem	(3) Relational self-efficacy (with peers)	(4) Relational self-efficacy (managing conflicts)	(5) Cognitive self-efficacy (related to school activities)	(6) Cognitive self-efficacy (general)	(7) Emotional self-efficacy (managing emotions)
Compliance with SL	0.108 (0.088)	-0.115 (0.129)	0.090 (0.074)	0.016 (0.075)	0.105 (0.075)	0.020 (0.085)	0.068 (0.106)
Effect size	0.167	-0.129	0.171	0.029	0.166	0.039	0.093
Observations	141	141	141	141	141	141	141
Mean of dep. var.	0.291	0.204	0.306	0.301	0.290	0.309	0.267
Std. dev. of dep. var.	0.573	0.854	0.496	0.524	0.576	0.476	0.676
First stage F statistic	253.8	266.1	238.4	263.9	249.9	256.8	261.6

Dep. var.:	(8) Emotional self-efficacy (getting support)	(9) Attitudes towards school (negative feelings)	(10) Attitudes towards school (school perceived as meaningless)	(11) Lack of motivation in studying	(12) Attitudes towards teachers (positive relationships)	(13) Attitudes towards teachers (feeling persecuted)	(14) Attitudes towards teachers (getting support)
Compliance with SL	0.068 (0.105)	0.106 (0.121)	0.122 (0.112)	0.042 (0.123)	-0.035 (0.091)	-0.142 (0.160)	0.219 (0.143)
Effect size	0.093	0.126	0.168	0.048	-0.061	-0.174	0.254
Observations	141	141	141	141	141	141	141
Mean of dep. var.	0.262	0.229	0.259	0.238	0.285	0.232	0.197
Std. dev. of dep. var.	0.693	0.795	0.701	0.769	0.602	0.785	0.870
First stage F statistic	260.3	320.6	286.6	280.7	235.3	232.6	272.9

Dep. var.:	(15) Attitudes towards teachers (allied with my teachers)	(16) Expectations in education	(17) Expectations in life	(18) Prosociality	(19) Internal locus of control	(20) Stable locus of control
Compliance with SL	-0.197* (0.108)	-0.020 (0.081)	0.046 (0.078)	0.124* (0.068)	0.336** (0.148)	-0.139 (0.195)
Effect size	-0.366	-0.036	0.073	0.195	0.325	-0.144
Observations	141	141	141	141	141	141
Mean of dep. var.	0.294	0.296	0.288	0.291	0	7.56e-09
Std. dev. of dep. var.	0.558	0.549	0.586	0.576	1	1
First stage F statistic	271.3	263.2	255.6	259.7	263	247

Notes - The results are based on 2SLS models including randomization blocks. A student from a school is considered as actually treated if s(he) attended the meetings for at least 10 hours. Standard errors in parentheses are clustered at the classroom level.
* p<0.10, ** p<0.05, *** p<0.01.

Table A17: Effects of Service Learning on Behaviors and School Performance – 2SLS Estimates – Alternative Definition of Compliance with the Protocol

Dep. var.:	(1) Sincerity behavior	(2) Altruistic behavior (dictator)	(3) Risk attitude (bart)	(4) Adversion to inequality behavior (ultimatum)	(5) Gratitude behavior (gratitude)	(6) Cheating (max choice)
			Panel A: Behaviors			
Compliance with SL	-0.201**	0.065*	-0.246**	-0.030	0.008	0.052*
	(0.084)	(0.037)	(0.118)	(0.054)	(0.043)	(0.028)
Effect size	-0.395	0.285	-0.256	-0.105	0.035	0.377
Observations	139	140	140	141	140	139
Mean of dep. var.	-0.404	0.406	2.978	0.423	0.454	0.566
Std. dev. of dep. var.	0.585	0.220	0.827	0.259	0.220	0.164
First stage F statistic	259.1	351.8	273.5	259.8	248.5	260.6

Dep. var.:	Passing to next grade (even with school debts)	Passing to next grade (without school debts)
	Panel B: School Performance	
Compliance with SL	-0.169*	-0.097
	(0.090)	(0.059)
Effect size	-0.378	-0.206
Observations	158	158
Mean of dep. var.	0.658	0.285
Std. dev. of dep. var.	0.476	0.453
First stage F statistic	286.8	286.8

Notes - The results are based on 2SLS models including randomization blocks. A student from a school is considered as actually treated if s(he) attended the meetings for at least 10 hours. Standard errors in parentheses are clustered at the classroom level.

* $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A18: Effects of Service Learning on Attitudes – 2SLS Estimates – Alternative Definition of Compliance with the Protocol

Dep. var:	(1) Positive self-esteem	(2) Negative self-esteem	(3) Relational self-efficacy (with peers)	(4) Relational self-efficacy (managing conflicts)	(5) Cognitive self-efficacy (related to school activities)	(6) Cognitive self-efficacy (general)	(7) Emotional self-efficacy (managing emotions)
Compliance with SL	0.113 (0.093)	-0.119 (0.135)	0.093 (0.077)	0.017 (0.077)	0.109 (0.078)	0.020 (0.088)	0.070 (0.110)
Effect size	0.173	-0.133	0.178	0.030	0.172	0.040	0.096
Observations	141	141	141	141	141	141	141
Mean of dep. var	0.291	0.204	0.306	0.301	0.290	0.309	0.267
Std. dev. of dep. var	0.573	0.854	0.496	0.524	0.576	0.476	0.676
First stage F statistic	192.6	200.2	199.2	204.6	194	196.8	199.2

Dep. var:	(8) Emotional self-efficacy (getting support)	(9) Attitudes towards school (negative feelings)	(10) Attitudes towards school (school perceived as meaningless)	(11) Lack of motivation in studying	(12) Attitudes towards teachers (positive relationships)	(13) Attitudes towards teachers (feeling persecuted)	(14) Attitudes towards teachers (getting support)
Compliance with SL	0.071 (0.109)	0.109 (0.124)	0.127 (0.116)	0.043 (0.128)	-0.037 (0.094)	-0.148 (0.167)	0.227 (0.148)
Effect size	0.097	0.131	0.173	0.049	-0.063	-0.180	0.263
Observations	141	141	141	141	141	141	141
Mean of dep. var	0.262	0.229	0.259	0.238	0.285	0.232	0.197
Std. dev. of dep. var	0.693	0.795	0.701	0.769	0.602	0.785	0.870
First stage F statistic	199.5	243.6	230.2	210	179.5	178.8	235.4

Dep. var:	(15) Attitudes towards teachers (allied with my teachers)	(16) Expectations in education	(17) Expectations in life	(18) Prosociality	(19) Internal locus of control	(20) Stable locus of control
Compliance with SL	-0.205* (0.112)	-0.021 (0.084)	0.048 (0.081)	0.129* (0.071)	0.348** (0.154)	-0.144 (0.203)
Effect size	-0.380	-0.037	0.075	0.202	0.336	-0.149
Observations	141	141	141	141	141	141
Mean of dep. var.	0.294	0.296	0.288	0.291	0	7.56e-09
Std. dev. of dep. var.	0.558	0.549	0.586	0.576	1	1
First stage F statistic	205.3	200.4	194.8	200.6	203.8	187.2

Notes – The results are based on 2SLS models including randomization blocks. A student from a school is considered as actually treated if s(he) participated in at least 50% of the scheduled hours and meetings as well as attended the meetings for at least 10 hours. Standard errors in parentheses are clustered at the classroom level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A19: Effects of Service Learning on Behaviors and School Performance – 2SLS Estimates – Alternative Definition of Compliance with the Protocol

	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Behaviors			
Dep. var.:	Sincerity behavior	Altruistic behavior (dictator)	Risk attitude (bart)	Adversion to inequality behavior (ultimatum)	Gratitude behavior (gratitude)	Cheating (max choice)
Compliance with SL	-0.209**	0.067*	-0.255**	-0.031	0.008	0.054*
	(0.087)	(0.038)	(0.122)	(0.056)	(0.045)	(0.029)
Effect size	-0.411	0.296	-0.266	-0.109	0.036	0.392
Observations	139	140	140	141	140	139
Mean of dep. var.	-0.404	0.406	2.978	0.423	0.454	0.566
Std. dev. of dep. var.	0.585	0.220	0.827	0.259	0.220	0.164
First stage F statistic	200	266.1	206	193.7	191.4	198.9

Dep. var.:	Passing to next grade (even with school debts)	Passing to next grade (without school debts)
	Panel B: School Performance	
Compliance (c1)	-0.178*	-0.102
	(0.095)	(0.063)
Effect size	-0.397	-0.216
Observations	158	158
Mean of dep. var.	0.658	0.285
Std. dev. of dep. var.	0.476	0.453
First stage F statistic	187.5	187.5

Notes – The results are based on 2SLS models including randomization blocks. A student from a school is considered as actually treated if s(he) participated in at least 50% of the scheduled hours and meetings as well as attended the meetings for at least 10 hours. Standard errors in parentheses are clustered at the classroom level.
* $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Appendix B: Outcome measures

1. Attitudes

Students' attitudes were measured before and after the intervention through several psychological scales administered in both questionnaires.

In Table B1, for each administered scale we report the number of items included in the questionnaire. The scales were retrieved from questionnaires that had already been validated by the scientific literature but deeply reviewed before administration, to fit the target population of this study. More precisely, pre-existing scales were not only translated into Italian but also adapted to the Italian context (including with ad-hoc items) and frequently reduced in length to ensure an overall reasonable length and questionnaire administration. In addition, all items were reported to a common agreement response scale, based on six alternatives: "not at all", "very little", "little", "enough", "a lot", and "totally". The scale aimed at measuring expectations in education was slightly different, despite maintaining the six-alternative answer, because it asked students to express their forecasts about specific events (i.e. obtaining a tertiary degree). In this case, the response options were as follows: "certainly not", "probably not", "maybe not", "maybe yes", "probably yes", and "certainly yes".

Moving from the original items, using principal component analysis, for each scale we identified the subset of items leading to a solution satisfying the following four criteria: a) maximizing the amount of explained variance; b) not generating ad-hoc factors, based on single items or the strong correlation between only two items; c) displaying items only strongly correlated with one factor; and d) showing the stronger correlations between each item and its main factor.

In the right columns of Table B1, in addition to the number of items originally administered in the questionnaire, for each scale we report the number of items selected through the principal component analyses and how they distributed among sub-dimensions, the amount of explained variance, Cronbach's Alpha, and a measure of the unidimensionality of each index.

Each index was obtained as a predicted standardized score (mean 0, standard deviation 1), from the final principal components analysis of the related scale.

The only scale for which we followed an entirely different analytical strategy is the one related to locus of control. In this case, we relied on ten items and the response scale was entirely different from the previous ones. Here, five options were provided for ten hypothetical positive or negative events, where students may have reached a certain goal or not. Respondents were asked to explain why each event took place, choosing among: "I was helped", "I was lucky", "It was easy", "I was able", and "I put effort into it" (or the opposite options, for negative events). Fo-

cusing on the choice of specific answer options, we identified two indexes: the first one measuring how often students imputed their success/failure to internal or external factors, and the second one indicating how often students imputed their success/failure to stable or unstable factors.

Table B1 – Scales in the questionnaire

Table B1 – Scales in the questionnaire and indexes obtained for the analyses

Scale's content	Extracted components	Administered items	Items in the final indexes	Variance explained (%)	Cronbach's alfa
Cognitive self-efficacy	Related to school activities	11	10	5	60	0.86
	General			4		0.67
Relational self-efficacy	With peers	11	8	5	61	0.83
	Managing conflicts			3		0.66
Emotional self-efficacy	Managing emotions	9	7	4	58	0.74
	Getting support			3		0.61
Self-esteem	Negative	10	7	3	67	0.86
	Positive			4		0.75
Attitudes towards school	Negative feelings	16	13	8	57	0.89
	School perceived as meaningless			5		0.77
Lack of motivation in studying	Lack of motivation in studying	7	3	3	56	0.61
Attitudes towards teachers	Positive relationships	27	21	8	66	0.92
	Feeling persecuted			6		0.87
	Getting support			4		0.82
	Allied with my teachers			3		0.73
Expectations in education	Expectations in education	4	3	3	61	0.68
Expectations in life	Expectations in life	12	8	8	62	0.91
Prosociality	Prosociality	17	8	8	55	0.88
Locus of control	Internal locus of control	10	10	10	-	-
	Stable locus of control	10		10	-	-

2. *Behaviors*

To complement the measure introduced in the previous section, we also administered students a set of incentivized tasks/behavioral games, which are described in the next few paragraphs.

Dictator

In the Dictator Game, originally developed by Kahneman et al. (1986) and then designed in the current version by Forsythe et al. (1994), a subject who is assigned the role of the proponent is provided with an exogenous endowment (in our experiment 11 euro). She is matched to an anonymous partner, assigned to the role of respondent, who received no endowment. The proponent chooses how to split the endowment between herself and the respondent. The latter has no influence over the outcome of the game. Within the standard theoretical assumptions of self-regarding agents, the Dictator Game has a unique Nash equilibrium in which the proponent maximizes her pay-off by keeping the entire endowment, thus sending no money to the respondent. Therefore, any deviation from the equilibrium solution in the Dictator Game is interpreted as a measure of altruism and/or pure generosity. The share of the initial endowment sent to the partner is a proxy for generosity or—more generally—"other-regarding preferences" (Engel, 2011; Guala e Mittone, 2010). In the established literature, a bimodal distribution of share

sent is found (whose peaks are at 0% and 50%, see Camerer, 2003), with an average around 30% (Engel, 2011).

Our outcome variable (dictator) comprises a normalized continuous indicator between 0 and 1, with 0 indicating no amount sent to the respondent and 1 indicating that the entire endowment has been sent to the respondent.

Ultimatum

In the Ultimatum Game, originally developed by Guth et al. (1982), the subject assigned to the role of respondent interacts with an anonymous partner who has received a given and known amount of money and plays the role of the proponent (in our experiment, the proponent receives 11 euro). The proponent is free to choose how to split the amount received with the subject. Once the proponent has chosen how much money to send to the respondent, the respondent is asked to accept or refuse the proposed split. If she accepts, the split is implemented, whereas if she rejects, neither the proponent nor the respondent receive anything. Our experiment is designed in strategic (rather than interactive) mode, i.e. asking subjects to state the minimum amount that they are willing to accept from the anonymous proponent (therefore, they are stating their "strategy", rather than reacting to an actual proposal). All subjects are assigned the role of the respondent. Their choices are then summarized by the "minimum acceptable offer" (MAO), i.e. the mini-

mum amount sent by the proponent that the respondent is willing to accept.

Since the choice of the respondent has implications on the outcome for both players, she is able to "punish" an iniquitous behavior. For this purpose, she must bear the cost of inflicting punishment (equal to the refused share). Therefore, within the standard theoretical assumptions of self-regarding agents, the equilibrium strategy of the respondent is to accept any positive offer by the proponent. However, empirical evidence reports deviations from the predicted equilibrium. Therefore, the actual behavior of the respondent in an Ultimatum Game proxies the degree of inequality aversion of the subject (Guth et al., 1982). In the established literature, offered shares lower than 30% are generally rejected (Camerer, 2003), and there is thus empirical evidence of a natural tendency towards punishing even if it implies a cost, in an iniquitous and strategic behavior by the proponent.

Our outcome variable (ultimatum) comprises a normalized continuous indicator between 0 and 1, indicating in relative terms the MAO of our subjects, where 0 indicates that subjects are willing to accept a null offer from the proponent and 1 indicating that they will only accept an offer amounting to the whole endowment.

Gratitude (trustworthiness)

The Gratitude Game designed in our experiment proposes to the subject the second stage of a Trust Game. In the Trust Game, also known as the

Investment Game (Berg et al. 1995), a proponent (he) is provided with an exogenous endowment (8 euro in our experiment), and he is matched to an anonymous partner (she) who has received no endowment. His decision concerns whether and how much of his endowment to send to the anonymous partner, while the proponent is also informed that the experimenter will multiply (triple) any amount sent. In the second stage—the only one actually played by the participants in our study—the respondent is asked whether she decides to send back to the proponent part of the amount received. The returned share is a measure of an «induced altruism» and/or the «gratitude» and «reciprocity» of the subject. In our experiment, this is the role assigned to all of our subjects, who are asked to reveal their full strategy, i.e., to state how much they are willing to send back to the proponent for every hypothetical level of amount received. The final pay-off of the proponent will be equal to the initial endowment, less the amount sent to the respondent, plus the amount sent back by the respondent to the proponent, while the pay-off of the respondent will be equal to the amount received less any amount sent back to the proponent. This game has a unique sub-game perfect Nash equilibrium in which the proponent maximizes his pay-off by keeping all of the endowment and sending 0 to the partner. Therefore, sending a positive share of the initial endowment to anonymous partners signals agents' propensity to interact with unknown partners, providing a proxy for generalized trust (Camerer, 2003; Berg et al. 1995; Johnson and Mislin, 2011). By contrast, the amount sent back by the respondent

proxies a measure of trustworthiness, or even gratitude, in response to the trust granted by the proponent.

Our outcome variable (gratitude) comprises a normalized continuous indicator between 0 and 1, indicating in relative terms the average share of endowment received (after multiplication) that the subjects are willing to send back to the proponent.

Cheating

To measure cheating/sincerity/truthfulness, we include a dice-rolling task (DRT) originally developed by Fischbacher and Follmi-Heusi (2013) and subsequently modified and further developed by Ariely and Garcia-Rada (2015). In our experiment, we implement a modified version of the DRT proposed by Ariely and Garcia-Rada (2015).

The purpose of the experimental task is to measure the attitude of the subject to truthfully report a series of favorable/unfavorable events. The subject is asked to report the results of a series of die-throwing tasks (in our experiment, twenty throws). Before every throw, the subject must choose—in her mind—one side of the die, "U" (Up) or "D" (Down), and memorize this decision without revealing it. She will gain the points corresponding to the chosen side, as declared after completing the throw. The distance between the average reported score and the expected value of a series of die-throwing task (equal to 3.5) provides an average statistical measure of truthfulness for a given population. The subject in fact could cheat by strategically reporting the chosen side of

the die to maximize its value. This situation has been used to analyze the influence of different cultural and social environment on sincerity. Ariely and Garcia-Rada (2014) have applied this situation to people born and raised in DDR vs. BRD, and Cohn, Fehr and Marechal (2014) to bank managers.

We use two alternative outcome variables. **Cheating** is an indicator reporting the difference between the average value of subjects' choice and 3.5 (i.e. the expected value of the series). A value of **cheating** larger than 3.5 indicates that on average students are likely to have lied in reporting their choices (i.e. they systematically reported the higher die instead of the one actually decided 'in their mind' before the throw), or they have been very lucky. The indicator therefore ranges from -2.5 (theoretical value of the "perfectly unlucky" person who in every throw always chooses a die displaying a value of 1) to 2.5 (theoretical value of the liar—or super-lucky person—who always chooses 6 in every throw). The second indicator—**maxchoice** —simply represents the proportion of throws in which students chose the die with the highest value. Again, higher values of the indicator correspond to a higher probability that students have changed the outcome of their choices to their advantage.

Note that the statistical properties of this indicator only apply to comparisons between groups of students (i.e. treated vs control), and not to individual students.

Bart

The Balloon Analogue Risk Task (BART) is a computerized measure of risk-taking behavior. The BART models real-world risk behavior through the conceptual frame of balancing the potential for reward versus loss. In the task, the participant is presented with a balloon and offered the chance to earn money by pumping the balloon up by clicking a button. Each click causes the balloon to incrementally inflate and money to be added to a counter up until some threshold, at which point the balloon is over inflated and explodes. Thus, each pump confers greater risk, but also greater potential reward. If the participant chooses to cash out prior to the balloon exploding, then they collect the money earned for that trail, but if the balloon explodes earnings for that trial are lost. Participants are not informed about the balloons' breakpoints, whereby the absence of this information allows for testing both participants' initial responses to the task and changes in responding as they gain experience with the task contingencies. Risk taking is a related but phenomenologically distinct process from impulsivity. For more information, refer to http://www.impulsivity.org/measurement/BART.

In our experiment, subjects face twenty balloons, and we adopt two alternative indicators. The first one—**bart**—is constructed as the average number of clicks on non-burst balloons, while the second indicator—**bart_avg**—is constructed in the same way but considering all of the balloons in the series, thus including those that have been burst.

References

Ariely, D., Garcia-Rada, X., Hornuf, L., and Mann, H. (2015). The (true) legacy of two really existing economic systems. Munich Discussion Paper No. 2014-26.

Berg, J., J. Dickhaut, and K. McCabe (1995). Trust, reciprocity, and social history. Games and Economic Behavior 10(1), 122–142.

Camerer, C. (2003). Behavioral game theory: Experiments in strategic interaction. Princeton: Princeton University Press.

Engel, C. (2011). Dictator games: A meta study. Experimental Economics, 14(4), 583-610.

Fischbacher, U. and Follmi-Heusi, F. (2013). Lies in disguise-n experimental study on cheating. Journal of the European Economic Association, 11(3):525–547.

Forsythe, R., J. L. Horowitz, N. E. Savin, and M. Sefton (1994). Fairness in simple bargaining experiments. Games and Economic Behavior 6(3), 347–369.

Guala, F., and L. Mittone (2010). Paradigmatic experiments: the dictator game. The Journal of Socio-Economics 39(5), 578–584.

Güth, W., Schmittberger, R., and Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. Journal of Economic Behavior & Organization, 3(4), 367-388.

Johnson, N. D., and A. A. Mislin (2011). Trust games: A meta-analysis. Journal of Economic Psychology 32(5), 865–889.

Kahneman, D., J. L. Knetsch, and R. Thaler (1986). Fairness as a constraint on profit seeking: Entitlements in the market. The American Economic Review 76(4), 728–741.

CRC - Centro Ricerche sulla Cooperazione e sul Nonprofit

Working Papers

1) Emilio Colombo - Patrizio Tirelli, *Il mercato delle banane e il commercio equo e solidale*, giugno 2006.

2) Emilio Colombo - Patrizio Tirelli, *Il mercato del caffè e il commercio equo e solidale*, giugno 2006.

3) Gian Paolo Barbetta, *Il commercio equo e solidale in Italia*, giugno 2006.

4) Antonella Sciarrone Alibrandi (a cura di), *Quali norme per il commercio equo e solidale?*, giugno 2006.

5) Emilio Colombo - Patrizio Tirelli, *Il valore sociale delle attività del commercio equo e solidale: l'impatto sui produttori*, maggio 2007.

6) Giacomo Boesso, Fabrizio Cerbioni, Andrea Menini, Antonio Parbonetti, *Foundations' governance for strategic philanthropy*, aprile 2012.

7) Stefan Einarsson - Jasmine McGinnis - Hanna Schneider, *Exploring the talk-action gap: a qualitative investigation of foundation practices over three regime types*, aprile 2012.

8) Marco Minciullo – Matteo Pedrini, *Knowledge transfer methods between founder firms and corporate foundations: upshots on orientation to effectiveness*, maggio 2012.

9) Gian Paolo Barbetta – Luca Colombo – Gilberto Turati, *The impact of fiscal rules on the grant-making behavior of American foundations*, luglio 2012.

10) Raffaella Rametta, *Regulation and the autonomy of foundations: a comparative analysis of the Italian and U.S. Patterns*, aprile 2013.

11) Emily Jansons*, From Gaining to Giving Wealth: The Shaping of a New Generation of Philanthropic Foundations in India*, dicembre 2013.

12) Edoardo Gaffeo, *Using information markets in grantmaking. An assessment of the issues involved and an application to Italian banking foundations*, gennaio 2014.

13) Giacomo Boesso, Fabrizio Cerbioni, Andrea Menini, Antonio Parbonetti, *The Role of Board's Competences and Processes in Shaping an Effective Grant-Making Strategy*, marzo 2014.

14) Siobhan Daly, *Making Grants, But to serve what purpose? An Analysis of Trust and Foundation Support for Women in the UK*, giugno 2014.

15) Francesca Calò, Elisa Ricciuti, *The state-of-the-art of Italian foundations' frameworks and methods to measure impact*, aprile 2016.

16) Sara Moggi, Chiara Leardini, Gina Rossi, Alessandro Zardini, *Meeting local community needs. A dashboard from the stakeholder engagement experience*, aprile 2016.

17) María José Sanzo-Perez, Marta Rey-García, Luis Ignacio Álva-rez-González, *Professionalization and partnerships with busi-nesses as drivers of foundation performance*, aprile 2016.

18) Henrik Mahncke, *Public-philanthropic partnership in provision of cultural services*, aprile 2016.

19) Gian Paolo Barbetta, Paolo Canino, Stefano Cima, Flavio Ver-recchia, *Entry and Exit of Nonprofit Organizations. An investi-gation with Italian census Data*, settembre 2017

20) Antonio Fici, L*e fondazioni filantropiche nella riforma del terzo settore,* maggio 2018.

21) Benedetta De Pieri, *Philanthropic Foundations as policy entre-preneurs. The role of philanthropic foundations in shaping the Social Innovation discourse in Italy and the UK,* maggio 2018.

22) Elisa Ricciuti, Urszula Swierczynska*, Powering social inno-vations – The role of Italian foundations is facilitating social in-novation. The grantees' perpective*, luglio 2018.

Visibili on line sul sito: http://centridiricerca.unicatt.it/crc-pubblicazioni-working-paper#content