Hard Cash and Soft Skills: Experimental Evidence on Combining Scholarships and Mentoring in Argentina*

Alejandro J. Ganimian†
New York University

Felipe Barrera-Osorio‡
Harvard University

M. Loreto Biehl§
Inter-American Development Bank

María A. Cortelezzi¶
Ministry of Education of Argentina

October 17, 2019

Abstract

We evaluate a long-standing program run by one of the largest education foundations in Argentina that offers scholarships and non-academic mentoring to secondary school students. We randomly assigned 408 grade 6 students within 10 public schools in the Province of Buenos Aires to either receive the program throughout secondary school or not to receive it. After three years, the program improved students' academic behaviors (e.g., studying before an exam or catching up on missed work), but we find little evidence that these changes translated into broader improvements in students' academic mindsets (e.g., self-beliefs about performance and efficacy), perseverance (e.g., grit), or learning strategies (e.g., metacognition). The program also improved students' performance in school during the first year (e.g., grades, attendance, and passing rates), but we do not find similar gains in subsequent years. This may be due to a large share of treatment students being expelled from the program for not meeting its requirements. The program did not improve student learning or personality traits (e.g., conscientiousness). Finally, we find some heterogeneous effects for female students and students from low socio-economic status.

JEL codes: C93, I21, I22, I25

Key words: Cash transfers, scholarships, mentoring, soft skills, Argentina.

*We are grateful for the funding provided by the Inter-American Development Bank and an anonymous foundation for this project. We thank Marina Bassi, Mercedes Mateo Díaz, Costas Meghir, Hugo Ñopo, Ernesto Pais, and Emiliana Vegas for their input at different stages of this project. We also thank Rafael de Hoyos, Timothée Demont, Greg Elacqua, Isy Faingold Vigil, Ariel Fiszbein, Verónica Frisancho, Brad Hershbein, Dick Murnane, Morgan Polikoff, Jeff Puryear, Laura Trucco, Michael Walker, and seminar participants at AEFP, APPAM, Brookings, GDN, IDB, LACEA-IEN, the Ministry of Education of Argentina, UNU-WIDER, and UTDT for comments. All views expressed are those of the authors and not of any of the institutions with which they are affiliated.

†Assistant Professor of Applied Psychology and Economics, New York University Steinhardt School of Culture, Education, and Human Development. E-mail: alejandro.ganimian@nyu.edu.

‡Associate Professor of Education and Economics, Harvard Graduate School of Education. E-mail: felipe_barrera-osorio@gse.harvard.edu.

§Education Lead Specialist, Inter-American Development Bank. E-mail: loretob@iadb.org.

¶National Director of Assessment of Education Quality and Equity, Ministry of Education of Argentina. E-mail: macortelezzi@gmail.com.

1 Introduction

In recent decades, developing countries have made impressive progress in expanding access to schooling. In Latin America and the Caribbean, this expansion focused on secondary school. By the early 1990s, most of the region had already achieved near-universal access to primary school, with an average enrollment rate at that level of 94%, but many youths still lacked access to secondary school, with an average enrollment rate at that level of 55%. Therefore, the lion's share of new enrollees in the period that followed were secondary school age youths. By the late 2000s, the average primary enrollment rate in the region had increased slightly to 98%, but the average secondary enrollment rate had jumped to 69% (Bassi et al. 2013).

Many of the youths who enrolled in secondary school in Latin America and the Caribbean over the past decades were among the first in their families to reach this level of education. These "first-generation" students differ from their predecessors along two important dimensions. First, the costs of schooling are more salient to them: the costs of "complements" to schooling (e.g., uniforms or textbooks) and the "opportunity cost" of schooling (i.e. the income forgone from not working) account for a larger share of their household income (Banerjee et al. 2013). Second, they are less prepared for secondary school because they cannot rely on the experiences of their parents with this level of education (see Alfonso et al. 2011; Bassi et al. 2012).

In spite of this massive influx of first-generation students, we still know very little about how to support them to ensure that they graduate from secondary school and acquire basic skills. One option is to offer these students a combination of financial support, to cover the costs and raise the immediate benefits of attending school, and mentoring, to provide them with scaffolding to develop the socio-emotional skills they need to succeed in secondary education. Yet, while there is ample evidence that financial support (e.g., scholarships and cash transfers) increase enrollment and attainment in developing countries (Ganimian and Murnane 2016), there are very few studies of mentoring in these settings (see, for example, Huan et al. 2014), and to our knowledge, no causal studies on their potential complementarity.

In this paper, we present experimental evidence on a program that offers scholarships and non-academic mentoring to secondary school students from low-income families in Argentina. The Scholarship and Mentoring Program (SMP) was created by the largest domestic education foundation and it is one of the longest-running and largest initiatives of its kind in the country: it was founded in 1997 and it currently serves more than 2,500 students across 16 provinces.[1] Every year, the program provides each student 10 monthly scholarships payments of USD 40 and 10 monthly individual or group-based, non-academic mentoring sessions by a trained professional. We randomly assigned 408 grade 6 students within 10 public schools in the Province of Buenos Aires to either receive the program for all five years of secondary school

[1]The foundation has asked us not to disclose its name or the true name of its program.

or not to receive it.[2] We evaluated the combined impact of scholarships and mentoring from grade 7 to grade 9.

We report four main sets of results based on this experiment. First, we find that the program had a positive impact on students' academic behaviors (e.g., their propensity to report that they started studying early before an exam or that they caught up on missed schoolwork).[3] These were precisely the behaviors that the mentors sought to teach program beneficiaries. We find moderate-to-large effects (from .15 to .31 standard deviations) on nearly all of them, and some evidence of smaller impacts on female students.

Second, we find little evidence that the program led to broader improvements in students' academic mindsets (e.g., beliefs about their own performance and efficacy), perseverance (e.g., grit), or learning strategies (e.g., metacognition). The program did not directly target these outcomes, but they might be expected to result from the exercise of academic behaviors. We do not find effects on any of these outcomes, but we cannot rule out small-to-moderate effects, and we find larger impacts on these outcomes for students from low socio-economic status.

Third, the program improved students' performance in school (e.g., attendance and grades) during the first year of the study, precisely the outcomes that the program aimed to influence. Yet, we do not find similar gains in subsequent years. This may be due to the foundation expelling a large share of students from the program for not meeting its requirements.

Lastly, we find no evidence that the program improved students' personality traits or learning (as measured by standardized tests of math and language). Importantly, the program did not seek to influence these outcomes, but we wanted to understand whether they would result from the expected improvements in socio-emotional skills and school performance, respectively. Interestingly, we find larger impacts on achievement for students from low-income families.

Our study contributes to the impact evaluation literature on the effects of scholarships and cash transfers in developing countries. This type of financial support has consistently increased enrollment and attainment, but with few exceptions, it has not improved learning outcomes (Barrera-Osorio et al. 2018; Barrera-Osorio and Filmer 2016; Kremer et al. 2009). Our results suggest that combining scholarships with mentoring helps students perform better in school, but it may not be enough to increase how much these students learn during secondary school. Yet, we believe that further research is needed on the effectiveness of combining scholarships with more intensive types of non-academic mentoring or with academic mentoring.

The rest of the paper is structured as follows. Section 2 describes the context, intervention, sampling, and randomization. Section 3 presents the data. Section 4 discusses the empirical strategy. Section 5 reports the results. Section 6 discusses implications for policy and research.

[2] In 12 of Argentina's 24 provinces, including the Province of Buenos Aires, primary school runs from grades 1 to 6 and secondary school runs from grades 7 to 12 (DiNIECE 2013).

[3] We discuss how we classify the different types of outcomes in this study in section 3.

2 Experiment

2.1 Context

Schooling in Argentina is compulsory and free from age 4 until the end of secondary school. In 12 of the 24 provinces, including the Province of Buenos Aires, primary school runs from grades 1 to 6 and secondary school from grades 7 to 12 (DiNIECE 2013).[4] According to the latest official figures, the Argentine school system serves nearly 11.4 million students, including 1.8 million in pre-school, 4.8 million in primary school, 3.8 million in secondary school, and 961,048 students in post-secondary education (DIEE 2016). The school calendar starts in February and ends in December.

Education policy in Argentina is shaped by both the national and the sub-national (province) governments. According to the National Education Law of 2006, the federal government is responsible for higher education and for providing technical and financial assistance to the provinces, and the provincial governments for pre-primary, primary, and secondary education.

Argentina is an interesting setting for exploring the potential complementarities between scholarships and non-academic mentoring. It expanded access to secondary education before most of Latin America: by the early 1990s, 60% of secondary school age youths were enrolled on time in Argentina, compared to 45% in the average country in the region. Its enrollment advantage persisted: by the late 2000s, 75% of secondary school age youths were enrolled in time, compared to 59% in the average neighboring country in the region (Bassi et al. 2013). Yet, its secondary school graduation rate is below those of its upper-middle income neighbors: in 2016, it stood at 63%, compared to 65% in Brazil, 91% in Chile, and 77% in Colombia (OECD 2018). Further, many secondary school students in Argentina do not reach national standards: in 2017, 69% of grade 12 students performed at the lowest two (out of four) levels of the national student assessment in math and 38% did so in language (SEE-MEDN 2018b).

We conducted this study in the Province of Buenos Aires for two main reasons. First, it is the largest sub-national school system in the country. It has 4,374 secondary schools and 1.5 million students from grades 7 to 12 (DIEE 2016). Second, its achievement is on par with that of the rest of the country. In the 2017 national assessment, 69% of its grade 12 students scored at the two lowest levels in math and 35% did so in language (SEE-MEDN 2018a).

2.2 Sample

We invited 10 public secondary schools in the Province of Buenos Aires to participate in the study. The foundation running the program selected these schools based on three

[4] In the other 12 provinces, primary runs from grades 1 to 7 and secondary from grades 8 to 12.

criteria: (a) they had to serve students from low-income families (because they are the target beneficiaries);[5] (b) they had to have participated in the program (to be familiar with its requirements); and (c) they could not be participating in the program at the time of recruitment (to avoid having students selected by the regular admissions process and by randomization in the same school). A representative from the foundation visited each school to invite it to participate in the study.

All schools agreed to participate in the study. Each school was located in a different district (*localidad*) of the Province of Buenos Aires: Campana, Ensenada, Gregorio de Laferrere, Guernica, José C. Paz, Merlo, Quilmes, Santos Lugares, Virrey del Pino, and Zárate.

At each school, we recruited grade 6 students as follows. If a school had two grade 6 sections, we invited the parents of students in both sections to attend an information session. If a school had more than two sections, we invited the parents of two randomly selected sections. A representative from the foundation explained the requirements of the program and study and then asked them to participate in an interview to collect baseline information. Typically, the foundation uses this interview to select students into the program using a set of criteria. Yet, in this case, it was simply used as an opportunity for data collection.[6] This process yielded 408 interested students across all 10 schools.

The schools in our sample are similar to all other public secondary schools in the province, especially in urban areas (Table A.1 in Appendix A). The two main differences are that in-sample schools enroll more students and have lower repetition rates than out-of-sample schools. Yet, if we restrict our comparison to grade 7, the initial target grade for the intervention, the only statistically significant difference between these schools is in their enrollment.

2.3 Randomization

We randomly assigned all 408 students in our sample to: (a) a "treatment" group, which was offered the program for all five years of secondary school; or (b) a "control" group, which was not offered the program.[7] We stratified the randomization by school to increase statistical power. This procedure resulted in 204 treatment and 204 control students across all 10 schools.

Control and treatment students were comparable at baseline. We compared both experimental groups on variables from the student survey and household surveys and found a few differences: treatment students were less likely to have dropped out of school, and more likely to have

[5]According to the latest official figures, nearly 30% of students in secondary "common" (i.e., non-special, non-adult) education are enrolled in private schools (DIEE 2016). The scholarship and mentoring program evaluated in this paper targets public schools because they typically serve students from lower-income families.

[6]We asked the staffers from the foundation to indicate whether they would have admitted each student through the regular admissions process to explore the existence of heterogeneous effects along this dimension.

[7]We discuss the requirements that students had to meet to remain in the program in the next section.

a car, natural gas, a refrigerator, and a cell phone at home, and to have parents who are homeowners. When we compute the standardized mean differences (which do not depend on sample size), we find that all differences are below .25 standard deviations (Imbens and Wooldridge 2009). Further, when we run joint tests using seemingly unrelated regressions, we cannot reject the null hypothesis of no difference across these groups ($\chi^2 = .055$, $p = .815$) (see Table 1). Finally, as we show in section 5, our impact estimates are nearly identical when we account for the first principal component from a principal component analysis of household assets.

2.4 Intervention

The Scholarship and Mentoring Program (SMP) provides students in public secondary schools with scholarships and non-academic mentoring. The program was developed by the largest education foundation in Argentina and it is one of the longest-running and largest initiatives of its kind in the country: in 2015, it reached 2,544 students across 16 of 24 provinces.

Students who are admitted into the program may receive it for all five years of secondary school if they comply with three requirements: (a) they remain enrolled in a program-affiliated school (because mentors cannot realistically follow students every time they move); (b) they do not repeat a grade (to encourage students to work hard at school); and (c) they are not suspended from school (to encourage students to behave well at school).

The program costs the foundation USD 733 per student per year (Table A.2).[8] Over half of the costs are due to the scholarships (i.e., the cash and costs of distributing it). The other half is spent on mentoring, administration, supervision, training, and identifying/selecting students.

2.4.1 Scholarship

Every year, the program provides each student with about USD 400 through the scholarship. The funds are disbursed on 10 monthly installments of USD 40 (from March to December), which is on par with the amount of cash transfers in Latin America (see Fiszbein et al. 2009). The foundation deposits the money in a bank account in the name of the students' parents. The funds may be withdrawn at any time and used for any purpose.

[8] We collected the costs of the program using the approach outlined in Dhaliwal et al. (2013).

2.4.2 Mentoring

Every year, the program provides each student with 10 non-academic mentoring sessions. Each session lasts 30-45 minutes and is typically held at school, before or after the school day. Sessions may be "individual" (i.e., between the mentor and one student) or "group-based" (i.e., between the mentor and multiple students). Each mentor decides the breakdown of individual and group meetings for each student. Mentors may also invite parents to join these meetings.

The mentors in this program differ from those in similar initiatives. First, they are employees of the foundation, not volunteers. Second, they completed tertiary or university education. Third, they receive pre- and in-service training and support.[9] Finally, they use an online platform to share resources and ideas.

Mentors in this program are granted considerable autonomy. First, they can determine the content of each session based on the needs of each student; it is not standardized.[10] Second, they can suspend or expel students from the program. The criteria for expulsion (outlined above) are set by the foundation, but there are no clear criteria for suspension, and in practice, mentors vary widely in the extent to which they use expulsions and suspensions.

Initially, the program was implemented as intended. In 2014, the average treatment student received eight scholarships and eight mentoring sessions, of which he/she attended eight. Most mentoring sessions were individual (seven of nine intended session) rather than group based (two of nine intended sections).[11] On average, parents were invited to six mentoring sessions, of which they attended five. Further, nearly all students had the same mentor throughout the school year (see Table A.3).

The exposure of the average treatment student to the program, however, decreased during the evaluation due to three main reasons. First, two students never joined the program. Second, many students were suspended at least once from the program: 26% of the total in 2014, 34% of students still in the program in 2015, and 31% of those remaining in the program in 2016. The average treatment student received 1.4 suspensions in 2014, 2 in 2015, and 2 in 2016.[12] Third, many students was expelled from the program during (6 students in 2014, 10 in 2015,

[9]Mentors undergo an induction process when they join and they receive an in-person and an online training session each year. They are also assigned to a coach who has mentoring experience and supervises their work.

[10]However, the foundation recommends that mentoring sessions start with an "icebreaker" for the mentor to earn the student's trust, proceed with a "diagnosis" in which the student discusses his/her strengths and weaknesses, and finish with an "action plan" in which the mentor and student set goals for the next session.

[11]The foundation did not collect information on the number of students in each group mentoring session.

[12]Suspensions varied widely across mentors. On any given year, the average number of suspensions per mentor ranged between 4.8 and 7.1 and the standard deviations from 4.4 to 5.3. By the third year of the evaluation, some mentors had not used any suspensions while others had used as many as 32.

and 3 in 2016) or at the end of the year (5 students in 2014, 26 in 2015, and 34 in 2016).[13] By the end of 2016, only 120 of the 204 treatment students (59%) remained in the program.[14]

3 Data

We designed our data collection based on the framework proposed by Farrington et al. (2012), which distinguishes between different types of socio-emotional skills and, drawing on prior theoretical and empirical work, considers how they might affect schoolwork during adolescence. Specifically, we collected data on five groups of outcomes: (a) "academic behaviors" (i.e., those commonly associated with being a "good student", such as going to class or doing homework); (b) "academic mindsets" (i.e., psycho-social attitudes students hold about themselves related to academics), "academic perseverance" (i.e., students' tendency to complete school assignments promptly and thoroughly), and "learning strategies" (i.e., tactics students use while thinking, remembering, and learning); (c) "school performance" (i.e., students' success in school); (d) "student achievement" (i.e., the extent to which students acquire new knowledge and skills); and (e) "personality traits" (i.e., psychological characteristics that are stable over time).

We expected the program to improve students' academic behaviors, given that this seemed to be the main focus of the mentoring sessions. Yet, we also wanted to understand whether, by regularly exercising these behaviors, students would develop broader academic mindsets, perseverance, and learning strategies. We thought it was unlikely that the program would influence students' personality traits, but we measured them to verify that this was the case. We expected the program to improve students' performance in school because that was its objective, but we wanted to know whether better school performance resulted in more learning. Table 2 offers an overview of all rounds of data collection and school participation rates.

3.1 Student and household surveys

We administered student and household surveys in 2014, before random assignment. The first survey asked students about their demographic and educational background and the second one asked parents or guardians about students' household conditions and assets. We use both surveys to check balance across experimental groups (see section 2.3).

[13] Expulsions also varied across mentors. On any given year, the average number of expulsions per mentor ranged from 1.1 to 4 and the standard deviations from 1.1 to 3.16. By the last year of the study, some mentors had not expelled any students while others had expelled up to 13 students.

[14] Note, however, that we continued to track all study participants for data collection purposes, regardless of whether they remained enrolled in the program.

3.2 Program participation

We collected data on students' participation in the program during all three years of the study. These data included the number of scholarship payments and mentoring sessions that each student in the treatment group was offered and received, several indicators on the mentoring sessions (e.g., whether they were individual or group-based), and whether each student was suspended or expelled. We use these data to confirm that the intervention was implemented as intended and to estimate the effect of receiving each scholarship and mentoring session (see section 4).

3.3 Academic behaviors

We collected data on students' academic behaviors in 2015 and 2016 using a survey we created. It asked students to recall the last time that they encountered a challenging situation at school (e.g., they did not understand something) and specify the steps they had taken to address it (e.g., asked the teacher to explain it again, consulted a book, asked a friend, or sought a tutor). It included 10 challenging situations (e.g., not understanding something during class, receiving homework, failing a homework assignment, having to study for a test, failing a test, failing a subject, missing a schoolday), each with a different number of potential solutions.[15] Appendix B discusses how students' responses to these situations were scored.

3.4 Academic mindsets, perseverance, and learning strategies

We collected data on students' academic mindsets, perseverance, and learning strategies during all three years of the study using a combination of self-reports and performance assessments. We selected the instruments with support from a local expert, who reviewed potential measures and identified those that had been administered and validated in Argentina (see Pais 2014).

To measure academic mindsets, we used a survey of students' self-beliefs about academics and the items from the Learning and Study Strategies Inventory (LASSI) that focus on motivation. To measure perseverance, we used the Grit scale, the Domain-Specific Impulsivity Scale for Children (DSIS-C), an assessment of self-control (CARAS) and one of planning skills (LABS). For learning strategies, we used the items in LASSI that focus on organization and planning.[16] Appendix B describes how each of these instruments were scored.

[15]The survey can be accessed at: `https://bit.ly/2Q58aqr`.

[16]The instruments can be accessed at: `https://bit.ly/2W5whL0` and `https://bit.ly/2JBLCMU`.

3.5 School performance

We collected data on students' performance in school during all three years of the study. These data included students' grades in math and language, indicator variables for whether each student failed the grade, transferred schools, or dropped out of school,[17] and their number of absences and "pending" subjects (i.e., subjects that students carried over to the next year).[18]

3.6 Student achievement

We administered assessments of math and reading in 2015 and 2016. They were designed by psychometricians at the *Centro de Medición de la Universidad Católica de Chile* (MIDE-UC). They assessed what students ought to know and be able to do according to the *Núcleos de Aprendizaje Prioritarios* (NAPs), the contents from the national curriculum prioritized by the federal government, as well as publicly-released items from the national student assessment.[19] We scored the assessments using a two-parameter logistic Item Response Theory model to account for differences in difficulty and discrimination across items (Yen and Fitzpatrick 2006). Appendix B describes the design and scoring of the assessments.

3.7 Personality traits

We administered the Big Five Inventory, a well-known survey of personality traits, in 2016 (see John et al. 2008; John and Srivastava 1999). The version that we administered measures students' extraversion, agreeableness, conscientiousness, neuroticism, and openness by asking them to indicate whether they match a series of descriptions (e.g., "I am outgoing, sociable") using a Likert scale ranging from 1 ("strongly disagree") to 5 ("strongly agree").[20]

[17]The national ministry of education requires schools to track two indicators: the number of students who switch schools and ask their original school for an authorization to switch (this is known as *salidos con pase*) and the number who switch schools without such authorization (*salidos sin pase*). The former is intended to measure transfers (this is what we report as students who "transferred schools") and the latter to measure dropouts (which we report as students who "dropped out of school"). In practice, however, a student may ask for a transfer and ultimately drop out or he/she may not ask for a transfer and still enroll in a different school. Given that neither indicator provides an accurate measure of dropout rates, we include both in our analysis.

[18]In Argentina, when students fail a subject, they need to take an exam to pass it in December. If they fail this exam, they need to take another exam in March. They can fail up to two subjects in March. If they fail more subjects, they can take these exams once again right before the school year begins. If they still fail more than two of these subjects by then, they are supposed to repeat the grade.

[19]The assessments can be accessed at: `https://bit.ly/2E7gFwD` and `https://bit.ly/2vTqk5i`.

[20]The survey can be accessed at: `https://bit.ly/2E5FvNe`.

3.8 Attrition

The vast majority of students in the sample participated in nearly all rounds of data collection. If we compute attrition in each round outlined in Table 2 with respect to the baseline sample, all rounds except for the last two (rounds 7 and 8) have attrition rates below 14% (Table A.4). In rounds 7 and 8 rates were 16% and 26%, mainly due to non-response from expelled students. Control students are slightly less likely to participate in all rounds of data collection than their treatment counterparts, but the magnitudes of the differences in participation rates are small (below 6 percentage points across all rounds) and statistically significant only at baseline.

Further, control and treatment groups remain comparable across all rounds of data collection. If we use the variables from the baseline student survey to check balance across experimental groups on each round of data collection, we find only small and mostly marginally statistically significant differences in the proportion of students who had previously dropped out of school (below 5 percentage points in all rounds) that favor the treatment group (Table A.5).

In light of these results, it is unlikely that attrition from the sample has violated the equality of expectation assumption across experimental groups (Murnane and Willett 2011). Nevertheless, to address this possibility, we estimate inverse-probability weighted (IPW) treatment effects and Lee (2009) bounds for each set of outcomes in our study.[21]

4 Empirical strategy

4.1 Intent-to-treat (ITT) effect

We estimate the effect of the *offer* of the program (i.e., the intent-to-treat or ITT effect) by fitting the following model:

$$Y_{is} = \alpha_{r(s)} + \gamma X_{is} + \beta T_{is} + \epsilon_{is} \quad (1)$$

[21] We estimate the IPW effects as follows. First, we fit a probit regression with an indicator variable for participation in a given data collection round (which equals 1 if the student participated and 0 otherwise) as the outcome and the baseline covariates from the student survey in Table A.5 as the predictors. Then, we weight the ITT estimate for each student by the inverse of his/her probability of participation, thus giving greater preponderance to students with lower probability of participation. This approach works best when researchers have collected baseline data on the variables that predict attrition from the sample (see Hernán and Robins 2006; Robins et al. 1994). Lee bounds estimate an interval for the true effect of an intervention on an outcome of interest by restricting the share of observations with the observed outcome to be equal across experimental groups. This approach works best when the distribution of the outcome in the group that suffers more from attrition offers useful guidance on how to trim the distribution of the outcome in the other group (see Lee 2009).

where Y_{igs} is the outcome of interest for student i in school s, $r(s)$ is the randomization stratum of school s and $\alpha_{r(s)}$ is a stratum (i.e., school) fixed effect, X_{is} is the first principal component from a principal component analysis of indicator variables for household assets collected prior to the intervention,[22] and T_{is} is an indicator variable for assignment to the treatment group. The parameter of interest is β, which captures the causal effect of the intervention. Our main estimates use heteroskedasticity-robust standard errors that are clustered at the school level (the unit of randomization). One potential drawback of this approach is that it may overstate the precision of our estimates because the number of clusters in our study is small (Cameron and Miller 2015). We also test the sensitivity of our estimates to the inclusion of X_{is}.

We also fit variations of this model that interact the treatment dummy with indicators for female students, those who had previously repeated a grade, those from low-income families, and those who would have been admitted through the program's regular selection process. We wanted to understand the heterogeneous effects of the program on the first three groups of students because prior research had indicated that they struggle to succeed in school (Cortelezzi et al. 2013, 2012). We also wanted to measure the effects on students who would have been selected by the program to understand whether the admissions criteria actually selected students who would benefit more from the scholarships and mentoring sessions.

4.2 Treatment-on-the-treated (TOT) effect

We also estimate the effect of *receiving* the program (i.e., the treatment-on-the-treated or TOT effect) by fitting the following two-stage least squares instrumental variables model:[23]

$$\begin{aligned} A_{is} &= \phi_{r(s)} + \zeta X_{is} + \mu T_{is} + \eta_{is} \\ Y_{is} &= \psi_{r(s)} + \omega X_{is} + \nu \hat{A}_{is} + \varepsilon_{is} \end{aligned} \tag{2}$$

where A_{is} is the number of months in which a student received both a scholarship and a mentoring session (which is zero for all control students), $\hat{A}_{is}$ is the predicted value of A_{is} from the first stage regression, and everything else is defined as in equation (1). The coefficient ν indicates the relationship between each month of the combined intervention and the outcome. As above, we adjust the standard errors to account for within-school correlations across students in outcomes and test the sensitivity of our estimates to the inclusion of X_{is}.

We also estimate the dose-response relationship between the number of scholarships or the number of mentoring sessions received and the outcomes of interest. The results from this

[22] These include indicator variables for whether a student had a car, natural gas, running water, a bathroom, a solid floor, a refrigerator, a computer, Internet access, and cell phone at home.

[23] We can interpret this estimate as the TOT effect (rather than as the local average treatment effect) because the SMP is only offered by one foundation and their data verify no control students had access to this program.

estimation should not be interpreted as the TOT effect of each component of the program. On any given month, a student may receive a scholarship, a mentoring session, or both.[24] Therefore, an estimation of the dose-response relationship between the number of scholarships received and the outcome of interest instrumented by the random assignment does not meet the exclusion restriction because the instrument also affects the outcome through the number of mentoring sessions received and vice versa (see Angrist et al. 1996).

5 Results

5.1 Average effects

5.1.1 Academic behaviors

The program improved students' academic behaviors. As Table 3 shows, treatment students were more likely to report adopting the behaviors listed in the survey than their control peers. Specifically, treatment students are more likely to report adopting these behaviors when they were assigned homework, failed a homework assignment, had to study for a test, failed a test, failed a term of a subject, failed a subject, planned to be absent, or were absent to school. These effects range from .15 to .28 standard deviations and their magnitude and statistical significance remain virtually unchanged when we account for baseline covariates (Table A.6).

These results are robust to several checks. First, treatment students outperformed control students in an index combining all behaviors, which suggests that the differences that we observe are not due to multiple hypothesis testing. Second, most effects are similar in magnitude and statistical significance across both years in which these outcomes were measured, which indicates that the pooled effects are not driven by any single year (Table A.7). Third, our inverse-probability weighted estimates and Lee (2009) bounds are consistent with our main results, suggesting that they are not explained by attrition (Table A.8-A.9). The former were statistically significant for nearly all measures in both years and the latter produce consistently positive and statistically significant upper bounds.

There are two patterns in the distribution of students' responses worth highlighting. First, the differences between experimental groups in some behaviors are driven by a large share of control students who did not engage in any behaviors included in the survey (Table A.10). This suggests that many students targeted by this program are unaware of academic behaviors

[24] For example, a student may be suspended from the scholarships for a month (e.g., for not complying with the requirements of the program), but he/she might still meet with his/her mentor. Similarly, a student may receive a scholarship on a given month, but miss the mentoring session.

that may help them succeed in school, and that mentoring exposes them to these behaviors.[25] Second, the number of behaviors in which students reported engaging decreased from 2015 to 2016, which suggests that while it is possible that differences across experimental groups are explained by social desirability bias (i.e., treatment students reporting engaging in academic behaviors because they were expected to do so by their mentors), it is unlikely (Figures B.1 and B.2).

Exposure to the program is associated with a higher propensity to adopt academic behaviors. Each month that a student received a scholarship and a mentoring session, he/she improved on average about .05 standard deviations on an index of all behaviors in our survey (Table A.11). In fact, each month has a positive and statistically significant effect on nearly all behaviors. Importantly, the magnitude of these monthly effects is consistent with the overall ITT effects.

5.1.2 Academic mindsets, perseverance, and learning strategies

There is little evidence that the program improved academic mindsets, perseverance, or learning strategies. As Table 4 shows, we find only a marginally statistically significant effect on the questions of the LASSI measuring motivation (of .12 standard deviations), but it is driven by a one-time difference in 2015, which does not emerge in 2014 or 2016 (Table A.12).

This pattern of null average effects remains unchanged when we use alternative specifications, such as accounting for baseline covariates (Table A.13) or weighing each student's estimate by the inverse probability of him/her participating in each data collection round (Table A.14). Yet, we cannot rule out small-to-moderate effects on these outcomes in any specification. Further, we find positive and statistically significant upper Lee (2009) bounds for grit, self-control (as measured by DSIS-C), and motivation in 2014 and 2015 (Table A.15), suggesting that attrition may lead us to under-estimate the effect of the program on those years.

The results from the TOT estimation are consistent with the ITT effects. We find that each month of exposure to the program only has a positive and statistically significant effect on the questions of the LASSI measuring motivation (of about .02 standard deviations, Table A.16).

5.1.3 School performance

The impact of the program on students' performance in school varies considerably by year. When we estimate the pooled impact of the program across all years of the study, we find only a marginally statistically significant reduction in absenteeism (by 2.6 days per year, Table 5).

[25] This pattern can also be observed graphically by year in Figures B.1 and B.2.

In fact, this effect is no longer statistically significant when we include covariates (Table A.17). Yet, this specification masks important differences in the effect of the program between years. In 2014, it raised language grades by .2 standard deviations and reduced the number of subjects that students carried over from one year to the next (by about half a subject), the number of absences (by 7 days), and the share of students who failed the grade (by 6 percentage points). In 2015, the sign of most effects stayed the same, but their magnitude decreased, and they were not statistically significant. In 2016, the sign of some effects changed and we find a *positive* and statistically significant effect on students' propensity to fail a grade (Table A.18).

The nonsignificant results for 2015 and 2016 are not explained by attrition from the study. Our inverse-probability weighted estimates for those years also yield null effects (Table A.19) and the Lee (2009) upper bounds are never positive and statistically significant (Table A.20).

It seems more likely that the fadeout in effects is due to the large share of students who were expelled from the program during the evaluation (see Table A.3, discussed in section 2.4) (i.e., they might bring down the average treatment outcomes in the last two years of the study). We find strong empirical support for this possibility when we conduct two additional analyses. First, we compare treatment students who remain enrolled in the program to control students. As expected, the former outperforms the latter on nearly all years and indicators (Table A.21). Then, we compare treatment students expelled from the program to their control counterparts. Also as expected, the former fare far worse on nearly all indicators and years (Table A.22). These results should be interpreted with caution as staying in the program is endogenous, but they suggest that the program continues to be effective for those who remain enrolled, whereas those who are expelled from the program fall behind even the control group.

This explanation is consistent with the TOT effects. Each month of exposure to the program is associated with a reduction in the number of pending subjects (by .05 subjects), the number of absences (by .37 absences), and propensity to drop out of school (by .04 pp., see Table A.23). These results provide further evidence of the effect of the program on enrolled students.

5.1.4 Student achievement

There is no evidence that the program improved student achievement. As Table 6 shows, the program had a negative but statistically insignificant effect on math and reading test scores (and on an average of both subjects) and we can rule out effects above .15 standard deviations. This is also the case when we account for baseline covariates in our estimation (Table A.24). In fact, the negative effect on math is marginally statistically significant in 2016 (Table A.25). These results are consistent with our discussion in section 5.1.3 on the potential consequences of a large share of students being expelled from the program during the course of the study, as well as with the null and precisely estimated TOT effects that we observe (Table A.26).

Accounting for attrition does not change this pattern. We obtain very similar results when we estimate inverse-probability weighted effects (Table A.27) and Lee (2009) bounds (Table A.28).

5.1.5 Personality traits

We find no evidence that the program improved personality traits either. As Table 7 shows, the program had a negative effect on extraversion and neuroticism and a positive effect on agreeableness and openness, but none was statistically significant. It also had a negative and marginally statistically significant effect on conscientiousness (of .25 standard deviations), and its magnitude and statistical significance are similar when we include covariates (Table A.29). This is partly due to a large share of control students scoring high in this facet (Figure B.7). We obtain very similar results when account for attrition from the sample (Tables A.30-A.31). The ITT effects are also consistent in sign and magnitude with the TOT effects (Table A.32).

5.2 Heterogeneous effects

The program had heterogeneous effects by students' sex only on academic behaviors. Specifically, it had smaller effects on academic behaviors for female students (by .24 standard deviations on the family index, see Table A.33). This pattern may be due to boys starting from lower levels of academic behavior than girls, as the coefficients for female students suggest. We do not find evidence of heterogeneity on these behaviors for any other sub-group of students.

The program also had heterogeneous effects by students' family income on several outcomes. It had a larger effect on academic mindsets, perseverance, and learning strategies (by .75 standard deviations on the family index, see Table A.34) and achievement (by .46 standard deviations on the family index, see Table A.35) for students from low socio-economic status. We do not find evidence of heterogeneity on these outcomes for any other sub-group.

Finally, we did not find any heterogeneous effects for any measure of school performance (Table A.36) or personality traits (Table A.37) for any group of students.

6 Discussion

This paper presents the results from a three-year randomized evaluation of a long-standing and high-profile program that provides a scholarship and a non-academic mentoring session every month to students in public secondary schools of the Province of Buenos Aires, Argentina. The program had a positive effect on academic behaviors. This is not entirely surprising, given that these were the behaviors that the mentoring sessions encouraged participants to adopt.

We find little evidence that these impacts translated into broader improvements in academic mindsets, perseverance, or learning strategies, but we cannot discard small to moderate effects.

The program improved students' performance in school in the first year of the study, but these effects are no longer discernible in the second and third years, possibly due to a large share of students being expelled from the program for not meeting its requirements. The program had no impact on student achievement or personality traits, but these outcomes were not targeted by the scholarships or mentoring sessions. Finally, we find heterogeneity in treatment effects on academic behaviors (which were smaller for female students) and on school performance and student achievement (which were larger for students from low-income families).

This evaluation offers several lessons for education policy and practice in developing nations. First, it indicates that this specific combination of scholarships and non-academic mentoring may be insufficient to achieve mid-to-long-term improvements in students' school performance. Although the program is successful in imparting academic behaviors on students, it seems to fall short of helping them develop the academic mindsets, perseverance, and learning strategies that they need to do well in school, learn more while there, and eventually graduate on time. This is an important lesson in light of recent studies that have identified brief and relatively inexpensive interventions that can produce lasting changes in these socio-emotional skills, at least in upper-middle income countries (see, for example, Alan et al. 2019; Outes et al. 2017).

There are, however, two important caveats to the main take-away from our impact evaluation. One of them is that this combination of scholarships and mentoring worked better for students from low-income families, who are in most need of support to succeed in secondary school. The program had a much larger impact for these students not only on school performance, which was its ultimate objective, but also on achievement, an outcome not directly targeted. This finding suggests that these supports can help low-income students get more out of school.

The other important caveat is that our study evaluates a specific mentoring intervention, which occurs only once a month, is not standardized, and focuses on non-academic support. Combining scholarships with other mentoring interventions may yield entirely different results. Specifically, mentoring programs that provide more intensive support and are purposefully structured to impact specific skills have produced encouraging results in developed countries (see, for example, Guryan et al. 2017; Heller et al. 2017; Kraft 2015).

Our study also provides insights on how to improve the design of non-academic mentoring, which should be of interest not only to the foundation that runs this program in Argentina, but also to government, non-profit, and/or private organizations managing similar programs. First, the fact that the program was not more effective for students who would have been admitted through its selection process suggests that it could broaden its target beneficiaries. Second, the fadeout in effects in school performance in the last two years of the evaluation,

which seem to be partly due to the expulsion of students failing to meet program requirements, suggests that the requirements do not serve their expected purpose of motivating students, and may end up pushing out precisely the type of students who are in greatest need of support. Finally, the wide variability in mentors' propensity to suspend and expel students indicates that there is a need for clearer guidelines for these sanctions to be adjudicated consistently.

References

Alan, S., T. Boneva, and S. Ertac (2019). Ever failed, try again, succeed better: Results from a randomized educational intervention on grit. *The Quarterly Journal of Economics*.

Alfonso, M., M. S. Bos, J. Duarte, and C. Rondón (2011). *Panorama general de la educación en América Latina y el Caribe*, Volume Educación para la transformación. Banco Interamericano de Desarollo (BID).

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association 91* (434), 444–455.

Arán-Filipetti, V. (2012). Estrato socioeconómico y habilidades cognitivas en niños escolarizados: Variables predictoras y mediadoras. *PSYKE 21*, 3–20.

Arán-Filipetti, V. and M. López (2013). Las funciones ejecutivas en la clínica neuropsicológica infantil. *Psicología desde el caribe 30*, 380–415.

Arán-Filipetti, V. and M. Richaud de Minzi (2011). Efectos de un programa de intervención para aumentar la reflexividad y la planificación en un ámbito escolar de alto riesgo por pobreza. *Universitas Psychologica 10*, 341–354.

Banerjee, A. V., P. Glewwe, S. Powers, and M. Wasserman (2013). Expanding access and increasing student learning in post-primary education in developing countries: A review of the evidence. *Unpublished manuscript*. Cambridge, MA: Abdul Latif Jameel Poverty Action Lab (J-PAL).

Barrera-Osorio, F., A. de Barros, and D. Filmer (2018). Long-term impacts of alternative approaches to increase schooling: Evidence from a scholarship program in Cambodia. (Policy Research Working Paper No. 8566). The World Bank. Washington, DC.

Barrera-Osorio, F. and D. Filmer (2016). Incentivizing schooling for learning: Evidence on the impact of alternative targeting approaches. *Journal of Human Resources 51* (2), 461–499.

Bassi, M., M. Busso, and J. S. Muñoz (2013). Is the glass half empty or half full? School enrollment, graduation, and dropout rates in Latin America. (IDB Working Paper No. 462). Washington, DC: Inter-American Development Bank.

Bassi, M., M. Busso, S. Urzúa, and J. Vargas (2012). Disconnected: Skills, education, and employment in Latin America. Washington, DC: Inter-American Development Bank.

Cameron, A. C. and D. L. Miller (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources 50* (2), 317–372.

Cortelezzi, M., D. Valencia, J. Malegarie, G. Fabro, and E. Pais (2013). La estrategia de acompañamiento de Futuros Egresados y el desarrollo de habilidades que favorecen experiencias escolares significativas. Fundación Cimientos: Ciudad de Buenos Aires, Argentina.

Cortelezzi, M., D. Valencia, J. Malegarie, A. Morrone, and N. Cohen (2012). Evaluación Programa Futuros Egresados: Estudio de las trayectorias escolares de alumnos de escuelas socio económicamente vulnerables. Fundación Cimientos: Ciudad de Buenos Aires, Argentina.

Dhaliwal, I., E. Duflo, R. Glennerster, and C. Tulloch (2013). Comparative cost-effectiveness analysis to inform policy in developing countries: A general framework with applications for education. *Education Policy in Developing Countries*, 285–338.

DIEE (2016). Anuario estadístico 2016. Buenos Aires, Argentina: Dirección de Investigación y Estadística Educativa (DIEE).

DiNIECE (2013). Redefiniciones normativas y desafíos de la educación secundaria en Argentina. Acuerdos federales en un sistema descentralizado. La educación en debate. Buenos Aires, Argentina: Dirección Nacional de Información y Evaluación de la Calidad Educativa (DiNIECE).

Duckworth, A. L., C. Peterson, M. D. Matthews, and D. R. Kelly (2007). Grit: perseverance and passion for long-term goals. *Journal of personality and social psychology 92*(6), 1087.

Duckworth, A. L. and P. D. Quinn (2009). Development and validation of the short Grit scale (GRIT–S). *Journal of Personality Assessment 91*, 166–174.

Farrington, C. A., M. Roderick, E. Allensworth, J. Nagaoka, T. S. Keyes, D. W. Johnson, and N. O. Beechum (2012). Teaching adolescents to become learners: The role of noncognitive factors in shaping school performance–A critical literature review. Unpublished manuscript. The Unviersity of Chicago Consortium on Chicago School Research (CCSR). Chicago, IL.

Fiszbein, A., N. R. Schady, F. H. Ferreira, M. Grosh, N. Kelleher, P. Olinto, and E. Skoufias (2009). *Conditional cash transfers: Reducing present and future poverty*. Washington, DC: The World Bank.

Ganimian, A. J. and R. J. Murnane (2016). Improving education in developing countries: Lessons from rigorous impact evaluations. *Review of Educational Research XX*(X), 1–37.

Guryan, J., S. Christenson, A. Claessens, M. Engel, I. Lai, J. Ludwig, and M. Turner (2017). The effect of mentoring on school attendance and academic outcomes: A randomized

evaluation of the Check & Connect Program. (IfPR Working Paper No. WP-16-18). Evanston, IL: Northwestern University.

Harris, D. (2005). Comparison of 1-, 2-, and 3-parameter IRT models. *Educational Measurement: Issues and Practice 8*(1), 35–41.

Heller, S. B., A. K. Shah, J. Guryan, J. Ludwig, M. S., and H. A. Pollack (2017). Thinking fast and slow? Some field experiments to reduce crime and dropout in Chicago. *The Quarterly Journal of Economics 132*, 1–54.

Hernán, M. A. and J. M. Robins (2006). Estimating causal effects from epidemiological data. *Journal of Epidemiology & Community Health 60*(7), 578–586.

Huan, W., J. Chu, P. Loyalka, X. Tao, and Q. Y. C. R. S. Shi, Yaojiang Qu (2014). Can school counseling reduce school dropout in developing countries? (REAP Working Paper No. 275). Rural Education Action Program (REAP). Stanford, CA.

Imbens, G. W. and J. M. Wooldridge (2009). Recent developments of the econometrics of program evaluation. *Journal of Economic Literature 47*(1), 5–86.

John, O. P., L. P. Naumann, and C. J. Soto (2008). Paradigm shift to the integrative Big-Five trait taxonomy: History, measurement, and conceptual issues. In John, O. P., Robins, R. W., & Pervin, A. (Eds.) *Handbook of personality: Theory and research* (3rd ed.) New York, NY: Guilford Press.

John, O. P. and S. Srivastava (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In Pervin, L. A. & John, O. P. (Eds.) *Handbook of personality: Theory and research* (Vol. 2). New York, NY: Guilford Press.

Kraft, M. A. (2015). How to make additional time matter: Integrating individualized tutorials into an extended day. *Education Finance and Policy 10*(1), 81–116.

Kremer, M. R., E. A. Miguel, and R. L. Thorton (2009). Incentives to learn. *The Review of Economics and Statistics XCI*(3), 437–456.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies 76*(3), 1071–1102.

Murnane, R. J. and J. B. Willett (2011). *Methods matter: Improving causal inference in educational and social science research.* Oxford, UK and New York, NY: Oxford University Press.

OECD (2018). Education at a Glance 2018: OECD Indicators. Paris, France: Organisation for Economic Co-operation and Deveopment.

Outes, I., A. Sánchez, and R. Vakis (2017). Cambiando la mentalidad de los estudiantes: Evaluación de impacto de !'Expande tu mente! sobre el rendimiento académico en tres regiones del Perú. *Documentos de Investigación*. Lima, Perú: Grupo de Análisis para el Desarrollo (GRADE).

Pais, E. F. (2014). Selección de instrumentos para la evaluación de habilidades socioemocionales para estudiantes de secundaria en Argentina: Fundamentación para una batería de evaluación. *Unpublished manuscript*. Buenos Aires, Argentina: Inter-American Development Bank (IDB).

Pais, E. F., M. Cortelezzi, and D. Valencia (2013). El desarrollo de habilidades socioemocionales en estudiantes secundarios a través de una estrategia de acompañamiento. Resultados de una evaluación sobre alumnos de 1er año del nivel secundario. In *V Congreso Internacional de Investigación y Práctica Profesional en Psicología*. Facultad de Psicología, Universidad de Buenos Aires. Buenos Aires, Argentina.

Robins, J. M., A. Rotnitzky, and L. P. Zhao (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American statistical Association 89*(427), 846–866.

Schmidt, V., N. Messoulam, and F. Molina (2008). Autoconcepto académico en adolescentes de escuelas medias: Presentación de un instrumento para su evaluación. *Revista Iberoamericana de Diagnóstico y Evaluación Psicológica 25*(81-106).

SEE-MEDN (2018a). Aprender 2017: Informe de resultados, Buenos Aires, 6to año secundaria. Ciudad Autónoma de Buenos Aires: Secretaría de Evaluación Educativa. Ministerio de Educación y Deportes de la Nación.

SEE-MEDN (2018b). Aprender 2017: Informe de resultados, secundaria. Ciudad Autónoma de Buenos Aires: Secretaría de Evaluación Educativa. Ministerio de Educación y Deportes de la Nación.

Thurstone, L. L. and M. Yela (2001). *CARAS. Test de percepción de diferencias (9a edición)*. Madrid, Spain: TEA Ediciones.

Tsukayama, E., A. L. Duckworth, and B. Kim (2013). Domain-specific impulsivity in school-age children. *Developmental Science 16*(879-893).

Weinstein, C. E. and D. R. Palmer (1988). *LASSI: The learning and study strategies inventory*. Miami, FL: Publishing Company.

Yen, W. M. and A. R. Fitzpatrick (2006). Item response theory. In Brennan, R. (Ed.) *Educational measurement* (4th ed.). Westport, CT: American Council on Education and Praeger Publishers.

Table 1: Balancing checks (baseline)

Variable	All (1)	Control (2)	Treatment (3)	Difference (4)	Standardized difference (5)	N (6)
Panel A. Student survey						
Argentine	.977	.98	.975	-.005	-.034	397
	(.149)	(.141)	(.157)	(.02)		
Female	.52	.544	.495	-.049	-.098	408
	(.5)	(.499)	(.501)	(.051)		
Age	12.435	12.502	12.368	-.131	-.123	407
	(1.062)	(1.153)	(.961)	(.11)		
Attends morning shift	.578	.583	.574	-.008	-.016	408
	(.494)	(.494)	(.496)	(.045)		
Previously repeated grade(s)	.309	.322	.297	-.024	-.052	404
	(.463)	(.468)	(.458)	(.044)		
Previously dropped out of school	.05	.073	.027	-.047*	-.216	360
	(.218)	(.262)	(.163)	(.022)		
Panel B. Household survey						
Has car	.208	.162	.255	.095***	.233	408
	(.407)	(.369)	(.437)	(.025)		
Has natural gas	.294	.265	.324	.06*	.132	408
	(.456)	(.442)	(.469)	(.032)		
Has running water	.811	.789	.833	.053	.135	408
	(.392)	(.409)	(.374)	(.05)		
Has in-house bathroom	.824	.809	.838	.03	.079	408
	(.382)	(.394)	(.369)	(.045)		
Has solid floor	.985	.98	.99	.01	.083	408
	(.121)	(.139)	(.099)	(.006)		
Has fridge	.713	.667	.76	.091***	.201	408
	(.453)	(.473)	(.428)	(.027)		
Has computer	.539	.539	.539	.002	.004	408
	(.499)	(.5)	(.5)	(.027)		
Has Internet	.382	.377	.387	.012	.025	408
	(.487)	(.486)	(.488)	(.033)		
Has cell phone	.904	.877	.931	.054*	.184	408
	(.294)	(.329)	(.253)	(.029)		
Parent is homeowner	.598	.564	.632	.07*	.143	408
	(.491)	(.497)	(.483)	(.033)		
χ^2				.055		
p-value				.815		

Notes: (1) The table shows the mean and standard deviations of all students in the sample (column 1), control group (column 2), and treatment group (column 3). It also tests for differences across these two groups (column 4), presents the standardized mean difference (column 5), and shows the number of non-missing observations (column 6). (2) * significant at 10%; ** significant at 5%; *** significant at 1%. (3) Standard errors in column 4 are clustered at the school level. (4) The bottom row of each panel displays the result from a joint test using Seemingly Unrelated Regressions (SURs).

Table 2: Data collection timeline

Event	Date	Location	Participation rates		
			Total	C	T
2014					
• School year starts	Feb				
• Round 0: Student survey	May 14-26	School	100%	100%	100%
• Household survey		School	81%	83%	78%
		Phone	19%	17%	22%
• Lottery is conducted and SMP starts	Jun				
• Round 1: Surveys of academic mindsets,	Nov 10-Dec 4	School	83%	75%	85%
perseverance, and learning strategies	Dec 18-Jan 16	Home	17%	19%	14%
2015					
• SMP data for 2014	Jan	-	100%	-	100%
• School year starts	Feb				
• Round 2: School performance data for 2014	May	School	100%	100%	100%
• Round 3: Math and reading tests	Jun 22-Jul 6	School	71%	68%	73%
	Jul 13-Aug 12	Home	29%	22%	18%
• Round 4: Surveys of academic mindsets,	Oct 14-Nov 6	School	66%	60%	72%
perseverance, and learning strategies	Nov 3-Dec 1	Home	24%	28%	20%
• Survey of academic behaviors					
2016					
• SMP data for 2015	Jan	-	94%	-	94%
• School year starts	Feb				
• Round 5: School performance data for 2015	May	School	86%	84%	88%
• Round 6: Math and reading tests	May 9-21	School	64%	60%	69%
	May 30-Jun 21	Home	28%	32%	25%
• Round 7: Surveys of academic mindsets,	Sep 19-30	School	61%	60%	62%
perseverance, and learning strategies	Oct 7-29	Home	23%	21%	24%
• Survey of academic behaviors					
• Survey of personality traits					
2017					
• SMP data for 2016	Jan	-	76%	-	76%
• School year starts	Feb				
• Round 8: School performance data for 2016	May	School	76%	73%	76%

Notes: (1) We collected SMP data from the non-profit, which is why we have not specified the location. (2) SMP data are available only for treatment students because control students did not participate in the program. (3) Total participation rates refer to all 408 students in the study, control participation rates refer to the 208 students in the control group, and treatment participation rates refer to the 208 students in the treatment group. (3) All SMP and school performance data are collected on a given year with respect to the prior school year.

Table 3: ITT effects on academic behaviors (2015-2016)

	(1) Proactive school behavior (std.)	(2) Preventinve homework behavior (std.)	(3) Corrective homework behavior (std.)	(4) Preventive test behavior (std.)	(5) Corrective test behavior (std.)	(6) Corrective failing behavior (std.)	(7) Corrective flunking behavior (std.)	(8) Preventive absenteeism behavior (std.)	(9) Corrective absenteeism behavior (std.)	(10) Corrective free period behavior (std.)	(11) Family index (std.)
Treatment	0.158	0.222**	0.223**	0.193***	0.204***	0.299***	0.154**	0.251***	0.279***	0.021	0.307***
	(0.091)	(0.088)	(0.078)	(0.040)	(0.054)	(0.051)	(0.057)	(0.063)	(0.063)	(0.093)	(0.064)
Observations	707	707	707	707	707	707	707	707	707	707	707
R^2	0.079	0.104	0.069	0.086	0.080	0.062	0.042	0.078	0.076	0.064	0.050

Notes: (1) This table shows the intent-to-treat (ITT) effect of the intervention, pooled across all years in which these outcomes were measured (see Table 2). (2) For an explanation of how each outcome was measured, see Appendix B. (3) All dependent variables have been standardized with respect to the control group in the first year that these outcomes were measured. (4) The observations indicate the total number of students who participated in data collection across all years in which these outcomes were measured. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 4: ITT effects on academic mindsets, perseverance, and learning strategies (2014-2016)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Grit (std.)	Self-control (DSIS-C) (std.)	Self-beliefs (std.)	Organization and planning (LASSI) (std.)	Motivation (LASSI) (std.)	Self-control (CARAS) (std.)	Organization and planning (LABS) (std.)	Family index (std.)
Treatment	0.052	0.093	0.073	0.028	0.124*	0.002	-0.054	0.083
	(0.053)	(0.073)	(0.089)	(0.083)	(0.065)	(0.074)	(0.044)	(0.082)
Observations	1102	1102	1102	1102	1102	1094	1102	1102
R^2	0.013	0.035	0.018	0.008	0.025	0.018	0.050	0.010

Notes: (1) This table shows the intent-to-treat (ITT) effect of the intervention, pooled across all years in which these outcomes were measured (see Table 2). (2) For an explanation of how each outcome was measured, see Appendix B. (3) All dependent variables have been standardized with respect to the control group in the first year that these outcomes were measured. (4) The observations indicate the total number of students who participated in data collection across all years in which these outcomes were measured. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 5: ITT effects on school performance (2014-2016)

	(1) Language grade (std.)	(2) Math grade (std.)	(3) Number of pending subjects	(4) Number of absences	(5) Failed grade	(6) Transferred schools	(7) Dropped out of school	(8) Family index (std.)
Treatment	0.028	0.009	-0.332	-2.643*	-0.015	-0.019	-0.025	0.047
	(0.086)	(0.054)	(0.182)	(1.427)	(0.018)	(0.013)	(0.014)	(0.065)
Observations	1012	1011	980	944	1057	1057	1057	1064
R^2	0.044	0.059	0.135	0.197	0.077	0.042	0.024	0.078
Control mean	-0.000	-0.000	1.516	33.400	0.148	0.025	0.054	0.000

Notes: (1) This table shows the intent-to-treat (ITT) effect of the intervention, pooled across all years in which these outcomes were measured (see Table 2). (2) For an explanation of how each outcome was measured, see Appendix B. (3) All dependent variables have been standardized with respect to the control group in the first year that these outcomes were measured. (4) The observations indicate the total number of students who participated in data collection across all years in which these outcomes were measured. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 6: ITT effects on student achievement (2015-2016)

	(1) Math test score (std.)	(2) Reading test score (std.)	(3) Family index (std.)
Treatment	-0.055	-0.008	-0.018
	(0.093)	(0.066)	(0.070)
Observations	683	706	714
R^2	0.191	0.171	0.224

Notes: (1) This table shows the intent-to-treat (ITT) effect of the intervention, pooled across all years in which these outcomes were measured (see Table 2). (2) For an explanation of how each outcome was measured, see Appendix B. (3) All dependent variables have been standardized with respect to the control group in the first year that these outcomes were measured. (4) The observations indicate the total number of students who participated in data collection across all years in which these outcomes were measured. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 7: ITT effects on personality traits (2016)

	(1) Extraversion (std.)	(2) Agreeableness (std.)	(3) Conscientiousness (std.)	(4) Neuroticism (std.)	(5) Openness (std.)	(6) Family index (std.)
Treatment	-0.055	0.182	-0.245*	-0.104	0.186	-0.017
	(0.090)	(0.116)	(0.111)	(0.087)	(0.106)	(0.097)
Observations	341	341	341	341	341	341
R^2	0.031	0.020	0.037	0.021	0.039	0.019

Notes: (1) This table shows the intent-to-treat (ITT) effect of the intervention on the only year in which these outcomes were measured (see Table 2). (2) For an explanation of how each outcome was measured, see Appendix B. (3) All dependent variables have been standardized with respect to the control group in the year that these outcomes were measured. (4) The observations indicate the total number of students who participated in data collection. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix A Additional graphs and tables

Table A.1: Comparison between in- and out-of-sample schools on school performance (2014)

	(1) All schools	(2) Out-of-sample schools All	(3) Out-of-sample schools Urban	(4) In-sample schools	(5) Col.(4)-Col.(2)	(6) Col.(4)-Col.(3)
Panel A. Secondary school						
Number of students enrolled	317.786	316.57	344.049	691.7	375.13***	347.651***
	(287.256)	(286.224)	(290.126)	(371.646)	(90.75)	(92.009)
Percentage of students who passed the grade	72.982	73	71.939	66.857	-6.143	-5.082
	(13.156)	(13.168)	(12.926)	(6.188)	(4.391)	(4.311)
Percentage of students who failed the grade	23.764	23.745	24.856	30.321	6.576	5.465
	(12.193)	(12.2)	(11.886)	(7.088)	(4.069)	(3.965)
Percentage of students who dropped out of school	3.254	3.255	3.205	2.823	-.432	-.382
	(4.886)	(4.887)	(4.86)	(4.802)	(1.631)	(1.623)
Percentage of students who repeated the grade	14.744	14.761	15.373	9.505	-5.256*	-5.868**
	(9.537)	(9.544)	(9.425)	(5.173)	(3.02)	(2.983)
N (schools)	3086	3076	2747	10	3086	2757
Panel B. Grade 7						
Number of students enrolled	73.037	72.898	79.267	113.7	40.802**	34.433**
	(51.203)	(51.163)	(50.269)	(48.794)	(16.205)	(15.925)
Percentage of students who passed the grade	75.113	75.126	74.017	70.214	-4.913	-3.803
	(15.212)	(15.216)	(14.953)	(13.688)	(5.386)	(5.294)
Percentage of students who failed the grade	21.623	21.612	22.768	25.821	4.21	3.054
	(13.946)	(13.947)	(13.664)	(14.124)	(4.938)	(4.839)
Percentage of students who dropped out of school	3.264	3.262	3.216	3.964	.702	.748
	(6.132)	(6.129)	(6.008)	(7.306)	(2.171)	(2.129)
Percentage of students who repeated the grade	15.736	15.737	16.519	15.285	-.452	-1.234
	(12.173)	(12.18)	(12.047)	(10.294)	(3.857)	(3.815)
N (schools)	2942	2932	2613	10	2942	2623

Notes: (1) The table shows the means and standard deviations of all public primary schools in the Province of Buenos Aires (column 1), non-RCT schools (columns 2-3), and RCT schools (column 4). It also tests for differences between all non-RCT and RCT schools (column 5), and between urban, non-RCT schools and RCT schools (column 6). Panel A shows results for all secondary school students and Panel B for grade 6 students. (2) Dropout rates should be interpreted as a upper-bound estimate, as they actually refer to the percentage of students who leave their schools without asking for a pass to another school. (3) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.2: Program costs per year (2014)

	Cost per year		Cost per student		Share
Budget line	ARS	USD	ARS	USD	of total
Cash transfers	$ 4,498,893	$ 464,035	$ 3,711.95	$ 382.86	.52
Mentoring sessions	$ 2,352,918	$ 242,690	$ 1,941.35	$ 200.23	.27
Administration	$ 616,546	$ 63,593	$ 508.70	$ 52.46	.07
Supervision and monitoring	$ 557,076	$ 57,459	$ 459.63	$ 47.40	.06
Training	$ 350,455	$ 36,147	$ 289.15	$ 29.82	.04
Identifying/selecting students	$ 233,491	$ 24,083	$ 192.64	$ 19.87	.03
Total	$ 8,609,380	$ 888,008	$ 7,103.44	$ 732.67	1

Notes: (1) The table shows the costs per year in Argentine pesos (ARS, column 1) and US dollars (USD, column 2), the cost per student in ARS (column 3) and USD (column 4), and the share of the total budget that each line represents (column 5). (2) The costs were estimated using information collected on the 1,212 students participating in the program in the PBA and its surrounding provinces in 2014. (3) The costs in USD were calculated using the historical exchange rate for December 2014, when the cost data were collected.

Table A.3: Program participation (2014-2016)

Variable	2014 Treatment (1)	N (2)	2015 Treatment (3)	N (4)	2016 Treatment (5)	N (6)
Scholarships received	8.52	204	7.817	191	6.516	155
	(1.686)		(3.347)		(3.454)	
Intended mentoring sessions	9.093	204	8.77	191	7.387	155
	(1.025)		(2.902)		(2.964)	
Actual mentoring sessions	7.819	204	7.487	191	6.348	155
	(1.782)		(3.291)		(3.38)	
Individual mentoring sessions	7.245	204	8.152	191	6.613	155
	(1.912)		(2.723)		(2.688)	
Group mentoring sessions	1.848	204	.618	191	.774	155
	(1.503)		(.707)		(.865)	
Sessions that had to be rescheduled	.24	204	.565	191	.477	155
	(.558)		(1.069)		(.907)	
Sessions to which parent was invited	5.858	204	7.157	191	5.123	155
	(2.295)		(2.56)		(2.642)	
Sessions to which parent attended	5.49	204	4.738	191	2.877	155
	(2.412)		(2.758)		(2.142)	
Sessions that used required materials	6.26	204	5.665	191	2.8	155
	(2.342)		(3.136)		(3.105)	
Mentors per student	1.191	204	1.099	191	1.077	155
	(.394)		(.3)		(.268)	
Student never joined	.01	204	0	191	0	155
	(.099)		(0)		(0)	
Student was suspended	.265	204	.335	191	.31	155
	(.442)		(.473)		(.464)	
No. of suspensions (if ever suspended)	1.444	54	2	64	1.958	48
	(.861)		(1.309)		(1.414)	
Student was expelled during the year	.029	204	.052	191	.019	155
	(.169)		(.223)		(.138)	
Student was expelled at the end of the year	.025	204	.136	191	.219	155
	(.155)		(.344)		(.415)	

Notes: (1) The table shows the mean and standard deviations of students in the treatment group (columns 1, 3, 5) and the number of non-missing observations (columns 2, 4, 6).

Table A.4: Attrition rates, by follow-up round of data collection

	Attrition from baseline				Attrition from previous round			
	All	Control	Treatment	Difference	All	Control	Treatment	Difference
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Round 1	.032	.054	.01	-.044**	.032	.054	.01	-.044**
	(.176)	(.226)	(.099)	(.019)	(.176)	(.226)	(.099)	(.019)
Round 2	0	0	0	-	0	0	0	-
	(0)	(0)	(0)		(0)	(0)	(0)	
Round 3	.096	.098	.093	-.003	.096	.098	.093	-.003
	(.294)	(.298)	(.291)	(.037)	(.294)	(.298)	(.291)	(.037)
Round 4	.103	.118	.088	-.03	.047	.049	.044	-.005
	(.304)	(.323)	(.284)	(.03)	(.211)	(.216)	(.206)	(.017)
Round 5	.137	.157	.118	-.037	.091	.113	.069	-.042*
	(.345)	(.365)	(.323)	(.021)	(.288)	(.317)	(.253)	(.021)
Round 6	.074	.083	.064	-.019	.032	.044	.02	-.024*
	(.261)	(.277)	(.245)	(.028)	(.176)	(.206)	(.139)	(.011)
Round 7	.164	.191	.137	-.053	.1	.123	.078	-.044
	(.371)	(.394)	(.345)	(.034)	(.301)	(.329)	(.27)	(.032)
Round 8	.255	.275	.235	-.037	.157	.157	.157	.002
	(.436)	(.447)	(.425)	(.053)	(.364)	(.365)	(.365)	(.037)
N	408	204	204	408	408	204	204	408

Notes: (1) The table shows the attrition rates for all students, and by experimental group, for each follow-up round of data collection (see Table 2). Attrition from one round equally affects all instruments administered in that round. Columns 1-4 show attrition rates with respect to baseline and columns 5-8 show attrition rates with respect to the previous data collection round. (2) * significant at 10%; ** significant at 5%; *** significant at 1%. (3) Standard errors in column 4 are clustered at the school level.

Table A.5: Balancing checks, by follow-up round of data collection

	Control	Difference with treatment							
	Round 1	Round 1	Round 2	Round 3	Round 4	Round 5	Round 6	Round 7	Round 8
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Argentine	.98	.001	-.005	0	.001	.013	0	.001	.002
	(.141)	(.016)	(.02)	(.017)	(.017)	(.017)	(.017)	(.018)	(.014)
Female	.544	-.064	-.049	-.068	-.094	-.059	-.072	-.074	-.049
	(.499)	(.053)	(.051)	(.057)	(.054)	(.06)	(.058)	(.056)	(.08)
Age	12.502	-.112	-.131	-.096	-.09	-.008	-.098	-.061	.058
	(1.153)	(.112)	(.11)	(.12)	(.116)	(.128)	(.115)	(.107)	(.092)
Attends morning shift	.583	-.014	-.008	-.014	-.02	-.047	-.024	.003	-.027
	(.494)	(.044)	(.045)	(.049)	(.052)	(.056)	(.05)	(.059)	(.061)
Previously repeated grade(s)	.322	-.027	-.024	-.038	-.032	.007	-.028	-.032	-.013
	(.468)	(.045)	(.044)	(.044)	(.042)	(.048)	(.042)	(.046)	(.039)
Previously dropped out of school	.073	-.034	-.047*	-.042*	-.032*	-.002	-.037*	-.04**	.004
	(.262)	(.022)	(.022)	(.021)	(.015)	(.026)	(.02)	(.014)	(.013)

Notes: (1) The table shows the attrition rates for all students, and by experimental group, for each follow-up round of data collection (see Table 2). Columns 1-4 show attrition rates with respect to baseline and columns 5-8 show attrition rates with respect to the previous data collection round. (2) * significant at 10%; ** significant at 5%; *** significant at 1%. (3) Standard errors in column 4 are clustered at the school level.

Table A.6: ITT effects on academic behaviors with covariates (2015-2016)

	(1) Proactive school behavior (std.)	(2) Preventinve homework behavior (std.)	(3) Corrective homework behavior (std.)	(4) Preventive test behavior (std.)	(5) Corrective test behavior (std.)	(6) Corrective failing behavior (std.)	(7) Corrective flunking behavior (std.)	(8) Preventive absenteeism behavior (std.)	(9) Corrective absenteeism behavior (std.)	(10) Corrective free period behavior (std.)	(11) Family index (std.)
Treatment	0.133	0.206**	0.213**	0.186***	0.188***	0.279***	0.143**	0.254***	0.273***	0.041	0.292***
	(0.087)	(0.076)	(0.078)	(0.039)	(0.056)	(0.052)	(0.056)	(0.066)	(0.064)	(0.099)	(0.063)
SES index	0.075	0.049	0.029	0.021	0.048	0.063*	0.032	-0.011	0.017	-0.060	0.044
	(0.042)	(0.050)	(0.038)	(0.035)	(0.029)	(0.028)	(0.026)	(0.040)	(0.039)	(0.034)	(0.047)
Observations	707	707	707	707	707	707	707	707	707	707	707
R^2	0.090	0.108	0.071	0.087	0.085	0.070	0.044	0.078	0.077	0.071	0.053

Notes: (1) This table shows the intent-to-treat (ITT) effect of the intervention, accounting for students' socio-economic status at baseline (see section 4), pooled across all years in which these outcomes were measured (see Table 2). (2) For an explanation of how each outcome was measured, see Appendix B. (3) All dependent variables have been standardized with respect to the control group in the first year that these outcomes were measured. (4) The observations indicate the total number of students who participated in data collection across all years in which these outcomes were measured. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.7: ITT effects on academic behaviors, by year (2015-2016)

	(1) Proactive school behavior (std.)	(2) Preventinve homework behavior (std.)	(3) Corrective homework behavior (std.)	(4) Preventive test behavior (std.)	(5) Corrective test behavior (std.)	(6) Corrective failing behavior (std.)	(7) Corrective flunking behavior (std.)	(8) Preventive absenteeism behavior (std.)	(9) Corrective absenteeism behavior (std.)	(10) Corrective free period behavior (std.)	(11) Family index (std.)
Panel A: 2015											
Treatment	0.115	0.235*	0.205**	0.209**	0.216**	0.263**	0.120	0.176	0.255**	0.039	0.239**
	(0.161)	(0.125)	(0.081)	(0.070)	(0.081)	(0.084)	(0.087)	(0.096)	(0.089)	(0.110)	(0.083)
Observations	366	366	366	366	366	366	366	366	366	366	366
R^2	0.051	0.048	0.034	0.041	0.041	0.058	0.042	0.040	0.041	0.060	0.049
Control mean	0.000	-0.000	0.000	-0.000	0.000	0.000	0.000	0.000	0.000	-0.000	-0.000
Panel B: 2016											
Treatment	0.208*	0.216**	0.253**	0.186***	0.198**	0.348***	0.200***	0.339***	0.313***	0.005	0.392***
	(0.100)	(0.087)	(0.102)	(0.048)	(0.068)	(0.046)	(0.047)	(0.061)	(0.062)	(0.121)	(0.082)
Observations	341	341	341	341	341	341	341	341	341	341	341
R^2	0.132	0.090	0.095	0.075	0.060	0.065	0.032	0.091	0.093	0.036	0.082
Control mean	-0.217	-0.468	-0.347	-0.413	-0.370	-0.292	-0.286	-0.418	-0.342	-0.277	0.000

Notes: (1) This table shows the intent-to-treat (ITT) effect of the intervention for each year in which these outcomes were measured (see Table 2). (2) For an explanation of how each outcome was measured, see Appendix B. (3) All dependent variables have been standardized with respect to the control group in the first year that these outcomes were measured. (4) The observations indicate the total number of students who participated in data collection each year in which these outcomes were measured. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.8: ITT effects on academic behaviors with inverse probability weighting, by year (2015-2016)

	(1) Proactive school behavior (std.)	(2) Preventinve homework behavior (std.)	(3) Corrective homework behavior (std.)	(4) Preventive test behavior (std.)	(5) Corrective test behavior (std.)	(6) Corrective failing behavior (std.)	(7) Corrective flunking behavior (std.)	(8) Preventive absenteeism behavior (std.)	(9) Corrective absenteeism behavior (std.)	(10) Corrective free period behavior (std.)	(11) Family index (std.)
Panel A: 2015											
Treatment	0.131	0.251*	0.209**	0.217**	0.224**	0.269***	0.132	0.187*	0.266**	0.039	0.251**
	(0.165)	(0.126)	(0.084)	(0.068)	(0.078)	(0.082)	(0.084)	(0.100)	(0.089)	(0.105)	(0.083)
Observations	366	366	366	366	366	366	366	366	366	366	366
R^2	0.055	0.050	0.034	0.042	0.044	0.059	0.043	0.041	0.042	0.060	0.051
Control mean	0.000	-0.000	0.000	-0.000	0.000	0.000	0.000	0.000	0.000	-0.000	-0.000
Panel B: 2016											
Treatment	0.211*	0.222**	0.249**	0.188***	0.203**	0.354***	0.208***	0.350***	0.322***	0.020	0.403***
	(0.101)	(0.084)	(0.107)	(0.050)	(0.072)	(0.050)	(0.049)	(0.058)	(0.058)	(0.120)	(0.082)
Observations	341	341	341	341	341	341	341	341	341	341	341
R^2	0.131	0.089	0.097	0.076	0.060	0.067	0.034	0.096	0.096	0.036	0.084
Control mean	-0.217	-0.468	-0.347	-0.413	-0.370	-0.292	-0.286	-0.418	-0.342	-0.277	0.000

Notes: (1) This table shows the intent-to-treat (ITT) effect of the intervention for each year in which these outcomes were measured (see Table 2), weighted by the inverse probability of each student participating in a data collection round. (2) The dependent variables take the value of 1 if a student has adopted at least one of the behaviors included in the survey and 0 otherwise. (3) All control means can be interpreted as proportions and all impact estimates as marginal effects. (4) The observations indicate the total number of students who participated in data collection each year in which these outcomes were measured. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.9: Lee bounds estimates of ITT effects on academic behaviors, by year (2015-2016)

	2015				2016			
	Lower (1)	Upper (2)	95% CI (3)	N (4)	Lower (5)	Upper (6)	95% CI (7)	N (8)
Proactive school behavior (std.)	-.103	.429***	[-.353, .629]	408	-.041	.423***	[-.273, .637]	408
	(.152)	(.121)			(.141)	(.13)		
Preventive homework behavior (std.)	.066	.287**	[-.142, .516]	408	.116	.252*	[-.126, .487]	408
	(.126)	(.138)			(.143)	(.139)		
Corrective homework behavior (std.)	.018	.247*	[-.195, .465]	408	.099	.306**	[-.138, .528]	408
	(.129)	(.132)			(.143)	(.134)		
Preventive test behavior (std.)	.112	.263**	[-.101, .484]	408	.035	.237**	[-.176, .424]	408
	(.127)	(.133)			(.128)	(.113)		
Corrective test behavior (std.)	.109	.265**	[-.092, .48]	408	.043	.26**	[-.156, .457]	408
	(.121)	(.129)			(.121)	(.12)		
Corrective failing behavior (std.)	.167	.309**	[-.045, .533]	408	.218*	.377***	[.014, .559]	408
	(.127)	(.134)			(.123)	(.109)		
Corrective flunking behavior (std.)	-.038	.164	[-.255, .381]	408	.05	.249**	[-.162, .425]	408
	(.131)	(.131)			(.129)	(.107)		
Preventive absenteeism behavior (std.)	.179	.399***	[-.019, .598]	408	.279**	.569***	[.052, .768]	408
	(.12)	(.121)			(.138)	(.121)		
Corrective absenteeism behavior (std.)	.215*	.304**	[.004, .529]	408	.227*	.377***	[.003, .58]	408
	(.122)	(.131)			(.134)	(.121)		
Corrective free period behavior (std.)	-.014	.284**	[-.242, .504]	408	-.107	.3**	[-.342, .51]	408
	(.139)	(.133)			(.143)	(.128)		

Notes: (1) The table shows the Lee bound estimates of the intent-to-treat (ITT) effect of the intervention for each year in which these outcomes were measured (see Table 2). It shows the lower (columns 1 and 5) and upper (columns 2 and 6) bounds, the 95% confidence interval (columns 3 and 7), and the number of observations (columns 4 and 8). (2) For an explanation of how each outcome was measured, see Appendix B. (3) All dependent variables have been standardized with respect to the control group in the first year that these outcomes were measured. (4) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.10: ITT effects on *any* academic behaviors, by year (2015-2016)

	(1) Proactive school behavior (std.)	(2) Preventinve homework behavior (std.)	(3) Corrective homework behavior (std.)	(4) Preventive test behavior (std.)	(5) Corrective test behavior (std.)	(6) Corrective failing behavior (std.)	(7) Corrective flunking behavior (std.)	(8) Preventive absenteeism behavior (std.)	(9) Corrective absenteeism behavior (std.)	(10) Corrective free period behavior (std.)
Panel A: 2015										
Treatment	0.015	0.045	0.045	0.056	0.107**	0.117*	0.070*	0.052*	0.080**	-0.022
	(0.053)	(0.030)	(0.026)	(0.032)	(0.045)	(0.053)	(0.038)	(0.025)	(0.031)	(0.036)
Observations	366	366	366	366	366	366	366	366	366	366
R^2	0.069	0.075	0.048	0.059	0.060	0.045	0.036	0.060	0.067	0.024
Control mean	0.817	0.883	0.850	0.867	0.811	0.694	0.583	0.844	0.822	0.889
Panel B: 2016										
Treatment	0.021	0.026	0.044	0.032	0.070*	0.073**	0.058*	0.062*	0.045	-0.038
	(0.027)	(0.031)	(0.041)	(0.022)	(0.035)	(0.025)	(0.028)	(0.028)	(0.034)	(0.042)
Observations	341	341	341	341	341	341	341	341	341	341
R^2	0.122	0.122	0.139	0.121	0.091	0.040	0.033	0.101	0.122	0.061
Control mean	0.855	0.861	0.812	0.848	0.782	0.715	0.612	0.818	0.818	0.855

Notes: (1) This table shows the intent-to-treat (ITT) effect of the intervention for each year in which these outcomes were measured (see Table 2). (2) The dependent variables take the value of 1 if a student has adopted at least one of the behaviors included in the survey and 0 otherwise. (3) All control means can be interpreted as proportions and all impact estimates as marginal effects. (4) The observations indicate the total number of students who participated in data collection each year in which these outcomes were measured. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.11: TOT effects on academic behaviors (2015-2016)

	(1) Months/year with scholarship	(2) Months/year with mentoring	(3) Months/year with both								
Panel A: First stage											
Treatment	6.786***	6.577***	6.321***								
	(0.386)	(0.348)	(0.363)								
Observations	707	707	707								
R^2	0.644	0.639	0.626								
	(1) Proactive school behavior (std.)	(2) Preventinve homework behavior (std.)	(3) Corrective homework behavior (std.)	(4) Preventive test behavior (std.)	(5) Corrective test behavior (std.)	(6) Corrective failing behavior (std.)	(7) Corrective flunking behavior (std.)	(8) Preventive absenteeism behavior (std.)	(9) Corrective absenteeism behavior (std.)	(10) Corrective free period behavior (std.)	(11) Family index (std.)
Panel B: Second stage											
Scholarships	0.023*	0.033***	0.033***	0.028***	0.030***	0.044***	0.023***	0.037***	0.041***	0.003	0.045***
	(0.012)	(0.012)	(0.010)	(0.006)	(0.008)	(0.007)	(0.008)	(0.008)	(0.008)	(0.013)	(0.009)
Observations	707	707	707	707	707	707	707	707	707	707	707
R^2	0.098	0.133	0.096	0.106	0.099	0.079	0.048	0.110	0.109	0.063	0.083
Mentoring	0.024*	0.034***	0.034***	0.029***	0.031***	0.045***	0.023***	0.038***	0.042***	0.003	0.047***
	(0.013)	(0.012)	(0.010)	(0.006)	(0.008)	(0.007)	(0.008)	(0.009)	(0.008)	(0.013)	(0.009)
Observations	707	707	707	707	707	707	707	707	707	707	707
R^2	0.098	0.132	0.096	0.106	0.101	0.081	0.050	0.109	0.111	0.063	0.083
Combined	0.025*	0.035***	0.035***	0.031***	0.032***	0.047***	0.024***	0.040***	0.044***	0.003	0.049***
	(0.013)	(0.013)	(0.011)	(0.006)	(0.008)	(0.008)	(0.009)	(0.009)	(0.008)	(0.014)	(0.009)
Observations	707	707	707	707	707	707	707	707	707	707	707
R^2	0.097	0.133	0.095	0.105	0.099	0.077	0.049	0.110	0.110	0.063	0.082

Notes: (1) The table shows the treatment-on-the-treated (TOT) effect of the intervention, pooled across all years in which these outcomes were measured (see Table 2). Panel A shows the results from the first stage regression of the number of months per year with a scholarship (column 1), a mentoring session (column 2), or both (column 3). Panel B shows the dose-response relationship of scholarships and mentoring, and the TOT effect of both. (2) For an explanation of how each outcome was measured, see Appendix B. (3) All dependent variables in Panel B have been standardized with respect to the control group in the first year that these outcomes were measured. (4) The observations indicate the total number of students who participated in data collection across all years in which these outcomes were measured. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.12: ITT effects on academic mindsets, perseverance, and learning strategies, by year (2015-2016)

	(1) Grit (std.)	(2) Self-control (DSIS-C) (std.)	(3) Self-beliefs (std.)	(4) Organization and planning (LASSI) (std.)	(5) Motivation (LASSI) (std.)	(6) Self-control (CARAS) (std.)	(7) Organization and planning (LABS) (std.)	(8) Family index (std.)
Panel A: 2014								
Treatment	0.079	0.099	0.042	0.027	0.153	-0.009	-0.014	0.097
	(0.073)	(0.099)	(0.102)	(0.101)	(0.128)	(0.082)	(0.066)	(0.103)
Observations	395	395	395	395	395	394	395	395
R^2	0.032	0.039	0.030	0.033	0.034	0.027	0.057	0.024
Control mean	-0.000	-0.000	-0.000	0.000	-0.000	0.000	0.000	0.000
Panel B: 2015								
Treatment	0.115	0.133	0.114	0.030	0.173**	0.044	-0.114	0.132
	(0.088)	(0.076)	(0.085)	(0.079)	(0.074)	(0.082)	(0.109)	(0.084)
Observations	366	366	366	366	366	360	366	366
R^2	0.034	0.052	0.029	0.008	0.063	0.034	0.057	0.013
Control mean	0.000	0.000	0.000	0.000	-0.000	0.000	-0.000	0.000
Panel C: 2016								
Treatment	-0.053	0.040	0.061	0.024	0.034	-0.036	-0.024	0.012
	(0.093)	(0.082)	(0.104)	(0.112)	(0.056)	(0.100)	(0.055)	(0.092)
Observations	341	341	341	341	341	340	341	341
R^2	0.011	0.051	0.015	0.026	0.030	0.041	0.067	0.024
Control mean	-0.000	0.000	0.000	0.000	-0.000	0.000	0.000	-0.000

Notes: (1) This table shows the intent-to-treat (ITT) effect of the intervention for each year in which these outcomes were measured (see Table 2). (2) For an explanation of how each outcome was measured, see Appendix B. (3) All dependent variables have been standardized with respect to the control group in the first year that these outcomes were measured. (4) The observations indicate the total number of students who participated in data collection each year in which these outcomes were measured. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.13: ITT effects on academic mindsets, perseverance, and learning strategies with covariates (2014-2016)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Grit (std.)	Self-control (DSIS-C) (std.)	Self-beliefs (std.)	Organization and planning (LASSI) (std.)	Motivation (LASSI) (std.)	Self-control (CARAS) (std.)	Organization and planning (LABS) (std.)	Family index (std.)
Treatment	0.062	0.100	0.073	0.041	0.122*	0.025	-0.066	0.093
	(0.056)	(0.073)	(0.084)	(0.087)	(0.065)	(0.075)	(0.043)	(0.083)
SES index	-0.032	-0.022	-0.000	-0.043	0.006	-0.072**	0.037	-0.032
	(0.036)	(0.038)	(0.046)	(0.031)	(0.041)	(0.031)	(0.026)	(0.044)
Observations	1102	1102	1102	1102	1102	1094	1102	1102
R^2	0.015	0.036	0.018	0.012	0.025	0.028	0.052	0.012

Notes: (1) This table shows the intent-to-treat (ITT) effect of the intervention, accounting for students' socio-economic status at baseline (see section 4), pooled across all years in which these outcomes were measured (see Table 2). (2) For an explanation of how each outcome was measured, see Appendix B. (3) All dependent variables have been standardized with respect to the control group in the first year that these outcomes were measured. (4) The observations indicate the total number of students who participated in data collection across all years in which these outcomes were measured. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.14: ITT effects on academic mindsets, perseverance, and learning strategies with inverse probability weighting, by year (2015-2016)

	(1) Grit (std.)	(2) Self-control (DSIS-C) (std.)	(3) Self-beliefs (std.)	(4) Organization and planning (LASSI) (std.)	(5) Motivation (LASSI) (std.)	(6) Self-control (CARAS) (std.)	(7) Organization and planning (LABS) (std.)	(8) Family index (std.)
Panel A: 2014								
Treatment	-0.002	0.015	-0.009	-0.055	0.152	-0.067	0.053	0.022
	(0.105)	(0.145)	(0.145)	(0.148)	(0.194)	(0.056)	(0.076)	(0.151)
Observations	247	247	247	247	247	247	247	247
R^2	0.036	0.017	0.034	0.022	0.026	0.036	0.016	0.026
Control mean	-0.000	-0.000	-0.000	0.000	-0.000	0.000	0.000	0.000
Panel B: 2015								
Treatment	0.128	0.133	0.131	0.044	0.179**	0.048	-0.107	0.148
	(0.096)	(0.088)	(0.094)	(0.088)	(0.075)	(0.082)	(0.118)	(0.097)
Observations	366	366	366	366	366	360	366	366
R^2	0.036	0.050	0.031	0.008	0.061	0.035	0.057	0.014
Control mean	0.000	0.000	0.000	0.000	-0.000	0.000	-0.000	0.000
Panel C: 2016								
Treatment	-0.044	0.068	0.083	0.034	0.057	-0.039	-0.028	0.033
	(0.093)	(0.086)	(0.115)	(0.120)	(0.056)	(0.101)	(0.056)	(0.101)
Observations	341	341	341	341	341	340	341	341
R^2	0.011	0.049	0.016	0.026	0.031	0.039	0.068	0.022
Control mean	-0.000	0.000	0.000	0.000	-0.000	0.000	0.000	-0.000

Notes: (1) This table shows the intent-to-treat (ITT) effect of the intervention for each year in which these outcomes were measured (see Table 2), weighted by the inverse probability of each student participating in a data collection round. (2) The dependent variables take the value of 1 if a student has adopted at least one of the behaviors included in the survey and 0 otherwise. (3) All control means can be interpreted as proportions and all impact estimates as marginal effects. (4) The observations indicate the total number of students who participated in data collection each year in which these outcomes were measured. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.15: Lee bounds estimates of ITT effects on academic mindsets, perseverance, and learning strategies, by year (2015-2016)

	2014				2015				2016			
	Lower (1)	Upper (2)	95% CI (3)	N (4)	Lower (5)	Upper (6)	95% CI (7)	N (8)	Lower (9)	Upper (10)	95% CI (11)	N (12)
Grit (std.)	-.016	.191*	[-.206, .37]	408	.069	.244**	[-.117, .428]	408	-.173	.067	[-.365, .262]	408
	(.116)	(.108)			(.113)	(.111)			(.116)	(.118)		
Self-control - DSIS-C (std.)	-.006	.199**	[-.172, .361]	408	.05	.217**	[-.127, .385]	408	-.068	.176	[-.262, .389]	408
	(.101)	(.098)			(.107)	(.102)			(.118)	(.13)		
Self-beliefs (std.)	-.095	.158	[-.272, .342]	408	.017	.258**	[-.173, .453]	408	-.017	.208	[-.234, .426]	408
	(.107)	(.111)			(.116)	(.118)			(.132)	(.132)		
Org. and planning - LASSI (std.)	-.073	.126	[-.245, .296]	408	-.021	.13	[-.177, .315]	408	-.141	.161	[-.371, .378]	408
	(.104)	(.103)			(.094)	(.112)			(.14)	(.132)		
Motivation - LASSI (std.)	.094	.222**	[-.059, .374]	408	.124	.214**	[-.054, .387]	408	-.18	.164*	[-.372, .328]	408
	(.092)	(.091)			(.105)	(.102)			(.117)	(.099)		
Self-control - CARAS (std.)	-.094	.014	[-.253, .164]	408	-.032	.068	[-.271, .231]	408	-.141	.025	[-.335, .221]	408
	(.095)	(.09)			(.139)	(.095)			(.117)	(.119)		
Org. and planning - LABS (std.)	-.134	.12	[-.293, .283]	408	-.258**	-.022	[-.445, .165]	408	-.152	.184	[-.345, .384]	408
	(.096)	(.099)			(.113)	(.113)			(.117)	(.122)		

Notes: (1) The table shows the Lee bound estimates of the intent-to-treat (ITT) effect of the intervention for each year in which these outcomes were measured (see Table 2). It shows the lower (columns 1 and 5) and upper (columns 2 and 6) bounds, the 95% confidence interval (columns 3 and 7), and the number of observations (columns 4 and 8). (2) For an explanation of how each outcome was measured, see Appendix B. (3) All dependent variables have been standardized with respect to the control group in the first year that these outcomes were measured. (4) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.16: TOT effects on academic mindsets, perseverance, and learning strategies (2014-2016)

	(1) Months/year with scholarship	(2) Months/year with mentoring	(3) Months/year with both					
Panel A: First stage								
Treatment	6.786***	6.577***	6.321***					
	(0.386)	(0.348)	(0.363)					
Observations	707	707	707					
R^2	0.644	0.639	0.626					
	(1) Grit (std.)	(2) Self-control (DSIS-C) (std.)	(3) Self-beliefs (std.)	(4) Organization and planning (LASSI) (std.)	(5) Motivation (LASSI) (std.)	(6) Self-control (CARAS) (std.)	(7) Organization and planning (LABS) (std.)	(8) Family index (std.)
Panel B: Second stage								
Scholarships	0.007	0.013	0.010	0.004	0.017*	0.000	-0.007	0.011
	(0.007)	(0.009)	(0.011)	(0.011)	(0.009)	(0.009)	(0.006)	(0.011)
Observations	1102	1102	1102	1102	1102	1094	1102	1102
R^2	0.015	0.036	0.021	0.009	0.037	0.018	0.047	0.016
Mentoring	0.007	0.013	0.010	0.004	0.018**	0.000	-0.008	0.012
	(0.007)	(0.010)	(0.012)	(0.011)	(0.009)	(0.010)	(0.006)	(0.011)
Observations	1102	1102	1102	1102	1102	1094	1102	1102
R^2	0.015	0.035	0.020	0.009	0.036	0.018	0.047	0.015
Combined	0.008	0.014	0.011	0.004	0.018*	0.000	-0.008	0.012
	(0.008)	(0.010)	(0.013)	(0.012)	(0.009)	(0.010)	(0.006)	(0.012)
Observations	1102	1102	1102	1102	1102	1094	1102	1102
R^2	0.015	0.036	0.021	0.009	0.038	0.018	0.047	0.016

Notes: (1) The table shows the treatment-on-the-treated (TOT) effect of the intervention, pooled across all years in which these outcomes were measured (see Table 2). Panel A shows the results from the first stage regression of the number of months per year with a scholarship (column 1), a mentoring session (column 2), or both (column 3). Panel B shows the dose-response relationship of scholarships and mentoring, and the TOT effect of both. (2) For an explanation of how each outcome was measured, see Appendix B. (3) All dependent variables in Panel B have been standardized with respect to the control group in the first year that these outcomes were measured. (4) The observations indicate the total number of students who participated in data collection across all years in which these outcomes were measured. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.17: ITT effects on school performance with covariates (2014-2016)

	(1) Language grade (std.)	(2) Math grade (std.)	(3) Number of pending subjects	(4) Number of absences	(5) Failed grade	(6) Transferred schools	(7) Dropped out of school	(8) Family index (std.)
Treatment	0.003	-0.007	-0.301	-1.964	-0.009	-0.017	-0.024	0.025
	(0.086)	(0.056)	(0.188)	(1.561)	(0.020)	(0.012)	(0.014)	(0.068)
SES index	0.091	0.059	-0.117	-3.050**	-0.020	-0.007	-0.004	0.078
	(0.058)	(0.049)	(0.120)	(1.348)	(0.017)	(0.005)	(0.004)	(0.054)
Observations	1012	1011	980	944	1057	1057	1057	1064
R^2	0.060	0.066	0.138	0.220	0.081	0.044	0.025	0.090
Control mean	-0.000	-0.000	1.516	33.400	0.148	0.025	0.054	0.000

Notes: (1) This table shows the intent-to-treat (ITT) effect of the intervention, accounting for students' socio-economic status at baseline (see section 4), pooled across all years in which these outcomes were measured (see Table 2). (2) For an explanation of how each outcome was measured, see Appendix B. (3) All dependent variables have been standardized with respect to the control group in the first year that these outcomes were measured. (4) The observations indicate the total number of students who participated in data collection across all years in which these outcomes were measured. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.18: ITT effects on school performance, by year (2014-2016)

	(1) Language grade (std.)	(2) Math grade (std.)	(3) Number of pending subjects	(4) Number of absences	(5) Failed grade	(6) Transferred schools	(7) Dropped out of school	(8) Family index (std.)
Panel A: 2014								
Treatment	0.195*	0.103	-0.494**	-7.098**	-0.060**	-0.010	-0.026	0.222**
	(0.099)	(0.097)	(0.191)	(2.337)	(0.022)	(0.012)	(0.021)	(0.072)
Observations	406	406	386	346	406	406	406	408
R^2	0.060	0.099	0.149	0.333	0.113	0.025	0.053	0.130
Control mean	-0.000	-0.000	1.516	33.400	0.148	0.025	0.054	0.000
Panel B: 2015								
Treatment	-0.039	-0.005	-0.307	-0.443	-0.033	-0.012	-0.013	-0.007
	(0.092)	(0.077)	(0.433)	(2.047)	(0.035)	(0.039)	(0.019)	(0.078)
Observations	331	331	325	321	348	348	348	352
R^2	0.084	0.104	0.084	0.166	0.105	0.060	0.030	0.105
Control mean	0.000	0.000	2.688	35.582	0.253	0.088	0.024	0.000
Panel C: 2016								
Treatment	-0.156	-0.108	-0.145	0.395	0.062*	-0.038	-0.038	-0.136
	(0.108)	(0.101)	(0.095)	(2.232)	(0.032)	(0.027)	(0.026)	(0.105)
Observations	275	274	269	277	303	303	303	304
R^2	0.063	0.126	0.036	0.217	0.057	0.075	0.026	0.079
Control mean	0.000	-0.000	0.912	35.607	0.162	0.108	0.074	0.000

Notes: (1) This table shows the intent-to-treat (ITT) effect of the intervention for each year in which these outcomes were measured (see Table 2). (2) For an explanation of how each outcome was measured, see Appendix B. (3) All dependent variables have been standardized with respect to the control group in the first year that these outcomes were measured. (4) The observations indicate the total number of students who participated in data collection each year in which these outcomes were measured. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.19: ITT effects on school performance with inverse probability weighting, by year (2015-2016)

	(1) Language grade (std.)	(2) Math grade (std.)	(3) Number of pending subjects	(4) Number of absences	(5) Failed grade	(6) Transferred schools	(7) Dropped out of school	(8) Family index (std.)
Panel A: 2015								
Treatment	-0.048	0.034	-0.395	-0.028	-0.046	-0.006	-0.014	0.017
	(0.103)	(0.082)	(0.456)	(2.291)	(0.036)	(0.048)	(0.019)	(0.078)
Observations	300	300	295	290	317	317	317	320
R^2	0.093	0.091	0.081	0.137	0.106	0.060	0.029	0.099
Control mean	0.000	0.000	2.688	35.582	0.253	0.088	0.024	0.000
Panel B: 2016								
Treatment	-0.180	-0.128	-0.116	1.896	0.075*	-0.037	-0.046	-0.190
	(0.121)	(0.119)	(0.095)	(3.256)	(0.040)	(0.025)	(0.029)	(0.145)
Observations	275	274	269	277	303	303	303	304
R^2	0.070	0.112	0.031	0.229	0.063	0.070	0.029	0.078
Control mean	0.000	-0.000	0.912	35.607	0.162	0.108	0.074	0.000

Notes: (1) This table shows the intent-to-treat (ITT) effect of the intervention for each year in which these outcomes were measured (see Table 2), weighted by the inverse probability of each student participating in a data collection round. 2014 is omitted because all students participated on round 2 of data collection. (2) The dependent variables take the value of 1 if a student has adopted at least one of the behaviors included in the survey and 0 otherwise. (3) All control means can be interpreted as proportions and all impact estimates as marginal effects. (4) The observations indicate the total number of students who participated in data collection each year in which these outcomes were measured. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.20: Lee bounds estimates of ITT effects on school performance, by year (2015-2016)

	2015				2016			
	Lower	Upper	95% CI	N	Lower	Upper	95% CI	N
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Language grade (std.)	-.227	.094	[-.491, .346]	408	-.385	.077	[-6.84, 10.058]	408
	(.16)	(.153)			(3.366)	(5.204)		
Math grade (std.)	-.137	.161	[-.384, .375]	408	-.355*	.005	[-.689, .375]	408
	(.15)	(.13)			(.203)	(.224)		
No. of pending subjects	-.668	-.112	[-1.667, .725]	408	-.292	-.032	[-.631, .252]	408
	(.592)	(.496)			(.204)	(.171)		
No. of absences	-2.596	1.617	[-13.419, 9.713]	408	-3.912	4.207	[-16.059, 15.392]	408
	(6.257)	(4.68)			(7.27)	(6.694)		
Failed grade	-.061	-.022	[-.176, .081]	408	.032	.074	[-.129, .186]	408
	(.066)	(.059)			(.091)	(.063)		
Transferred schools	-.023	-.004	[-.103, .066]	408	.003	-.04	[-.131, .038]	408
	(.044)	(.04)			(.068)	(.04)		
Dropped out of school	-.024**	-.012	[-.044, .017]	408	-.068**	-.034	[-.115, .018]	408
	(.012)	(.017)			(.029)	(.031)		

Notes: (1) The table shows the Lee bound estimates of the ITT effects. It shows the lower (columns 1 and 5) and upper (columns 2 and 6) bounds, the 95% confidence interval for the ITT effect (columns 3 and 7), and the number of observations (columns 4 and 8). (2) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.21: Association between enrollment in program and school performance, by year (2014-2016)

	(1) Language grade (std.)	(2) Math grade (std.)	(3) Number of pending subjects	(4) Number of absences	(5) Failed grade	(6) Transferred schools	(7) Dropped out of school	(8) Family index (std.)
Panel A: 2014								
Enrolled	0.244**	0.149	-0.607**	-8.211**	-0.071**	-0.019	-0.055**	0.305***
	(0.100)	(0.089)	(0.226)	(2.520)	(0.026)	(0.013)	(0.022)	(0.076)
Observations	396	396	382	337	395	395	395	397
R^2	0.067	0.098	0.155	0.345	0.116	0.025	0.069	0.148
Control mean								
Panel B: 2015								
Enrolled	0.107	0.118	-0.979*	-5.635	-0.086*	-0.064*	-0.023	0.174
	(0.112)	(0.100)	(0.500)	(3.275)	(0.046)	(0.034)	(0.013)	(0.105)
Observations	311	310	308	299	322	322	322	325
R^2	0.102	0.115	0.103	0.198	0.117	0.072	0.045	0.130
Control mean								
Panel C: 2016								
Enrolled	0.175	0.270***	-0.271**	-10.248***	-0.048*	-0.082**	-0.101***	0.302***
	(0.107)	(0.069)	(0.100)	(2.973)	(0.026)	(0.031)	(0.023)	(0.089)
Observations	248	247	244	250	273	273	273	274
R^2	0.049	0.174	0.056	0.254	0.057	0.119	0.070	0.095
Control mean								

Notes: (1) This table shows the association between being enrolled in the program (defined as being assigned to the treatment group and not being expelled) for each year in which these outcomes were measured (see Table 2).

Table A.22: Association between expulsion from the program and school performance, by year (2014-2016)

	(1) Language grade (std.)	(2) Math grade (std.)	(3) Number of pending subjects	(4) Number of absences	(5) Failed grade	(6) Transferred schools	(7) Dropped out of school	(8) Family index (std.)
Panel A: 2014								
Expelled	-0.731**	-0.798**	4.727***	13.085	0.147	0.148	0.455**	-1.212***
	(0.295)	(0.256)	(0.578)	(10.456)	(0.123)	(0.119)	(0.156)	(0.212)
Observations	213	213	194	182	214	214	214	215
R^2	0.088	0.118	0.247	0.316	0.151	0.085	0.227	0.192
Control mean								
Panel B: 2015								
Expelled	-0.920***	-0.664***	4.678***	26.050**	0.177	0.289**	0.043	-0.840***
	(0.233)	(0.170)	(0.748)	(8.159)	(0.156)	(0.121)	(0.072)	(0.220)
Observations	181	181	176	176	198	198	198	201
R^2	0.170	0.181	0.268	0.270	0.164	0.169	0.053	0.213
Control mean								
Panel C: 2016								
Expelled	-1.074***	-0.868***	0.034	21.982***	0.396***	0.068	-0.062	-1.017***
	(0.178)	(0.214)	(0.159)	(5.979)	(0.082)	(0.086)	(0.049)	(0.253)
Observations	164	163	159	166	192	192	192	193
R^2	0.206	0.207	0.041	0.316	0.173	0.087	0.040	0.187
Control mean								

Notes: (1) UPDATE

Table A.23: TOT effects on school performance (2014-2016)

	(1) Months/year with scholarship	(2) Months/year with mentoring	(3) Months/year with both					
Panel A: First stage								
Treatment	6.786***	6.577***	6.321***					
	(0.386)	(0.348)	(0.363)					
Observations	707	707	707					
R^2	0.644	0.639	0.626					
	(1) Language grade (std.)	(2) Math grade (std.)	(3) Number of pending subjects	(4) Number of absences	(5) Failed grade	(6) Transferred schools	(7) Dropped out of school	(8) Family index (std.)
Panel B: Second stage								
Scholarships	0.003	0.001	-0.042*	-0.335**	-0.002	-0.003	-0.003*	0.006
	(0.010)	(0.007)	(0.022)	(0.170)	(0.002)	(0.002)	(0.002)	(0.008)
Observations	1012	1011	980	944	1057	1057	1057	1064
R^2	0.048	0.061	0.149	0.212	0.082	0.050	0.035	0.086
Mentoring	0.004	0.001	-0.044*	-0.357**	-0.002	-0.003	-0.004*	0.006
	(0.011)	(0.007)	(0.023)	(0.178)	(0.002)	(0.002)	(0.002)	(0.008)
Observations	1012	1011	980	944	1057	1057	1057	1064
R^2	0.048	0.061	0.149	0.215	0.082	0.051	0.035	0.087
Combined	0.004	0.001	-0.045*	-0.367**	-0.002	-0.003	-0.004*	0.007
	(0.011)	(0.007)	(0.024)	(0.184)	(0.002)	(0.002)	(0.002)	(0.009)
Observations	1012	1011	980	944	1057	1057	1057	1064
R^2	0.049	0.061	0.151	0.215	0.082	0.051	0.036	0.088

Notes: (1) The table shows the treatment-on-the-treated (TOT) effect of the intervention, pooled across all years in which these outcomes were measured (see Table 2). Panel A shows the results from the first stage regression of the number of months per year with a scholarship (column 1), a mentoring session (column 2), or both (column 3). Panel B shows the dose-response relationship of scholarships and mentoring, and the TOT effect of both. (2) For an explanation of how each outcome was measured, see Appendix B. (3) All dependent variables in Panel B have been standardized with respect to the control group in the first year that these outcomes were measured. (4) The observations indicate the total number of students who participated in data collection across all years in which these outcomes were measured. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.24: ITT effects on student achievement with covariates (2015-2016)

	(1) Math test score (std.)	(2) Reading test score (std.)	(3) Family index (std.)
Treatment	-0.078	-0.036	-0.049
	(0.091)	(0.066)	(0.069)
SES index	0.081***	0.090*	0.104**
	(0.024)	(0.046)	(0.037)
Observations	683	706	714
R^2	0.203	0.184	0.244

Notes: (1) This table shows the intent-to-treat (ITT) effect of the intervention, accounting for students' socio-economic status at baseline (see section 4), pooled across all years in which these outcomes were measured (see Table 2). (2) For an explanation of how each outcome was measured, see Appendix B. (3) All dependent variables have been standardized with respect to the control group in the first year that these outcomes were measured. (4) The observations indicate the total number of students who participated in data collection across all years in which these outcomes were measured. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.25: ITT effects on student achievement, by year (2015-2016)

	(1) Math test score (std.)	(2) Reading test score (std.)	(3) Family index (std.)
Panel A: 2015			
Treatment	0.054	-0.067	-0.002
	(0.140)	(0.085)	(0.072)
Observations	335	349	355
R^2	0.212	0.217	0.250
Control mean	0.000	0.000	0.000
Panel B: 2016			
Treatment	-0.161*	0.050	-0.037
	(0.081)	(0.099)	(0.089)
Observations	348	357	359
R^2	0.199	0.154	0.221
Control mean	0.007	-0.158	-0.000

Notes: (1) This table shows the intent-to-treat (ITT) effect of the intervention for each year in which these outcomes were measured (see Table 2). (2) For an explanation of how each outcome was measured, see Appendix B. (3) All dependent variables have been standardized with respect to the control group in the first year that these outcomes were measured. (4) The observations indicate the total number of students who participated in data collection each year in which these outcomes were measured. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.26: TOT effects on student achievement (2015-2016)

	(1) Months/year with scholarship	(2) Months/year with mentoring	(3) Months/year with both
Panel A: First stage			
Treatment	6.786***	6.577***	6.321***
	(0.386)	(0.348)	(0.363)
Observations	707	707	707
R^2	0.644	0.639	0.626
	(1) Math test score (std.)	(2) Reading test score (std.)	(3) Family index (std.)
Panel B: Second stage			
Scholarships	-0.008	-0.001	-0.003
	(0.013)	(0.009)	(0.010)
Observations	683	706	714
R^2	0.184	0.170	0.221
Mentoring	-0.008	-0.001	-0.003
	(0.013)	(0.010)	(0.010)
Observations	683	706	714
R^2	0.184	0.170	0.222
Combined	-0.009	-0.001	-0.003
	(0.014)	(0.010)	(0.010)
Observations	683	706	714
R^2	0.183	0.170	0.221

Notes: (1) The table shows the treatment-on-the-treated (TOT) effect of the intervention, pooled across all years in which these outcomes were measured (see Table 2). Panel A shows the results from the first stage regression of the number of months per year with a scholarship (column 1), a mentoring session (column 2), or both (column 3). Panel B shows the dose-response relationship of scholarships and mentoring, and the TOT effect of both. (2) For an explanation of how each outcome was measured, see Appendix B. (3) All dependent variables in Panel B have been standardized with respect to the control group in the first year that these outcomes were measured. (4) The observations indicate the total number of students who participated in data collection across all years in which these outcomes were measured. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.27: ITT effects on student achievement with inverse probability weighting, by year (2015-2016)

	(1) Math test score (std.)	(2) Reading test score (std.)	(3) Family index (std.)
Panel A: 2015			
Treatment	0.060	-0.080	-0.003
	(0.141)	(0.087)	(0.074)
Observations	335	349	355
R^2	0.222	0.220	0.256
Control mean	0.000	0.000	0.000
Panel B: 2016			
Treatment	-0.132	0.083	0.003
	(0.087)	(0.124)	(0.107)
Observations	311	320	322
R^2	0.222	0.167	0.242
Control mean	0.007	-0.158	-0.000

Notes: (1) This table shows the intent-to-treat (ITT) effect of the intervention for each year in which these outcomes were measured (see Table 2), weighted by the inverse probability of each student participating in a data collection round. (2) The dependent variables take the value of 1 if a student has adopted at least one of the behaviors included in the survey and 0 otherwise. (3) All control means can be interpreted as proportions and all impact estimates as marginal effects. (4) The observations indicate the total number of students who participated in data collection each year in which these outcomes were measured. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.28: Lee bounds estimates of ITT effects on student achievement, by year (2015-2016)

	2015				2016			
	Lower	Upper	95% CI	N	Lower	Upper	95% CI	N
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Math test score (std.)	.011	.15	[-.343, .458]	408	-.232	-.034	[-.497, .237]	408
	(.204)	(.179)			(.159)	(.163)		
Reading test score (std.)	-.12	.013	[-.448, .323]	408	-.037	.188	[-.292, .433]	408
	(.19)	(.18)			(.154)	(.148)		

Notes: (1) The table shows the Lee bound estimates of the intent-to-treat (ITT) effect of the intervention for each year in which these outcomes were measured (see Table 2). It shows the lower (columns 1 and 5) and upper (columns 2 and 6) bounds, the 95% confidence interval (columns 3 and 7), and the number of observations (columns 4 and 8). (2) For an explanation of how each outcome was measured, see Appendix B. (3) All dependent variables have been standardized with respect to the control group in the first year that these outcomes were measured. (4) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.29: ITT effects on personality traits with covariates (2016)

	(1) Extraversion (std.)	(2) Agreeableness (std.)	(3) Conscientiousness (std.)	(4) Neuroticism (std.)	(5) Openness (std.)	(6) Family index (std.)
Treatment	-0.042	0.155	-0.221*	-0.099	0.159	-0.022
	(0.096)	(0.118)	(0.114)	(0.087)	(0.106)	(0.104)
SES index	-0.039	0.081**	-0.071	-0.014	0.078	0.016
	(0.030)	(0.027)	(0.042)	(0.050)	(0.047)	(0.048)
Observations	341	341	341	341	341	341
R^2	0.034	0.031	0.046	0.022	0.048	0.020

Notes: (1) This table shows the intent-to-treat (ITT) effect of the intervention, accounting for students' socio-economic status at baseline (see section 4), on the only year in which these outcomes were measured (see Table 2). (2) For an explanation of how each outcome was measured, see Appendix B. (3) All dependent variables have been standardized with respect to the control group in the year that these outcomes were measured. (4) The observations indicate the total number of students who participated in data collection. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.30: ITT effects on personality traits with inverse probability weighting (2016)

	(1) Extraversion (std.)	(2) Agreeableness (std.)	(3) Conscientiousness (std.)	(4) Neuroticism (std.)	(5) Openness (std.)	(6) Family index (std.)
Treatment	-0.064	0.162	-0.254**	-0.118	0.166	-0.050
	(0.089)	(0.126)	(0.109)	(0.084)	(0.101)	(0.092)
Observations	341	341	341	341	341	341
R^2	0.029	0.018	0.039	0.024	0.040	0.021
Control mean	-0.000	0.000	-0.000	0.000	-0.000	-0.000

Notes: (1) This table shows the intent-to-treat (ITT) effect of the intervention for each year in which these outcomes were measured (see Table 2), weighted by the inverse probability of each student participating in a data collection round. (2) The dependent variables take the value of 1 if a student has adopted at least one of the behaviors included in the survey and 0 otherwise. (3) All control means can be interpreted as proportions and all impact estimates as marginal effects. (4) The observations indicate the total number of students who participated in data collection each year in which these outcomes were measured. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.31: Lee bounds estimates of ITT effects on personality traits (2016)

	Lower	Upper	95% CI	N
	(1)	(2)	(3)	(4)
Extraversion (std.)	-.182	.058	[-.441, .303]	408
	(.157)	(.148)		
Agreeableness (std.)	.045	.273*	[-.208, .508]	408
	(.153)	(.142)		
Conscientiousness (std.)	-.351**	-.148	[-.6, .123]	408
	(.15)	(.163)		
Neuroticism (std.)	-.212	-.014	[-.469, .246]	408
	(.155)	(.156)		
Openness (std.)	.084	.257	[-.223, .534]	408
	(.182)	(.164)		

Notes: (1) The table shows the Lee bound estimates of the intent-to-treat (ITT) effect of the intervention for each year in which these outcomes were measured (see Table 2). It shows the lower (columns 1 and 5) and upper (columns 2 and 6) bounds, the 95% confidence interval (columns 3 and 7), and the number of observations (columns 4 and 8). (2) For an explanation of how each outcome was measured, see Appendix B. (3) All dependent variables have been standardized with respect to the control group in the first year that these outcomes were measured. (4) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.32: TOT effects on personality traits (2016)

	(1) Months/year with scholarship	(2) Months/year with mentoring	(3) Months/year with both			
Panel A: First stage						
Treatment	6.786***	6.577***	6.321***			
	(0.386)	(0.348)	(0.363)			
Observations	707	707	707			
R^2	0.644	0.639	0.626			
	(1) Extraversion (std.)	(2) Agreeableness (std.)	(3) Conscientiousness (std.)	(4) Neuroticism (std.)	(5) Openness (std.)	(6) Family index (std.)
Panel B: Second stage						
Scholarships	-0.010	0.032*	-0.043**	-0.018	0.033*	-0.003
	(0.015)	(0.018)	(0.018)	(0.014)	(0.019)	(0.016)
Observations	341	341	341	341	341	341
R^2	0.026	0.007	0.024	0.018	0.025	0.019
Mentoring	-0.010	0.033*	-0.044**	-0.019	0.033*	-0.003
	(0.015)	(0.019)	(0.019)	(0.014)	(0.019)	(0.016)
Observations	341	341	341	341	341	341
R^2	0.026	0.010	0.025	0.016	0.029	0.019
Combined	-0.010	0.034*	-0.046**	-0.020	0.035*	-0.003
	(0.016)	(0.019)	(0.019)	(0.015)	(0.020)	(0.017)
Observations	341	341	341	341	341	341
R^2	0.027	0.009	0.025	0.016	0.027	0.019

Notes: (1) The table shows the treatment-on-the-treated (TOT) effect of the intervention, pooled across all years in which these outcomes were measured (see Table 2). Panel A shows the results from the first stage regression of the number of months per year with a scholarship (column 1), a mentoring session (column 2), or both (column 3). Panel B shows the dose-response relationship of scholarships and mentoring, and the TOT effect of both. (2) For an explanation of how each outcome was measured, see Appendix B. (3) All dependent variables in Panel B have been standardized with respect to the control group in the first year that these outcomes were measured. (4) The observations indicate the total number of students who participated in data collection across all years in which these outcomes were measured. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.33: Heterogeneous ITT effects on academic behaviors (2015-2016)

	Family index (std.)			
	(1)	(2)	(3)	(4)
Treatment	0.434***	0.373***	0.273**	0.281
	(0.069)	(0.103)	(0.096)	(0.251)
Female	0.090			
	(0.119)			
Treatment × Female	-0.235**			
	(0.089)			
Low SES		-0.005		
		(0.121)		
Treatment × Low SES		-0.302		
		(0.236)		
Repeated a grade			-0.266	
			(0.251)	
Treatment × Repeated			0.080	
			(0.325)	
Eligible				0.229
				(0.190)
Treatment × Eligible				0.039
				(0.327)
Observations	707	707	702	707
R^2	0.053	0.057	0.059	0.060
Control mean	-0.000	-0.000	-0.000	-0.000

Notes: (1) This table shows the heterogeneous intent-to-treat (ITT) effects of the intervention, pooled across all years in which these outcomes were measured (see Table 2). (2) Low socio-economic status (SES) students are those in the first quartile of the first principal component from a principal component analysis of household assets. Students who repeated a grade are those who did so prior to the start of the program and study (i.e., during elementary school). Eligible students are those that would have been admitted to the program through its regular admissions process (see section 2.2). (3) The dependent variable is standardized with respect to the control group in the first year that these outcomes were measured. (4) The observations indicate the total number of students who participated in data collection across all years in which these outcomes were measured. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.34: Heterogeneous ITT effects on academic mindsets, perseverance, and learning strategies (2014-2016)

	Family index (std.)			
	(1)	(2)	(3)	(4)
Treatment	0.035	-0.100	0.146	0.378**
	(0.158)	(0.055)	(0.125)	(0.156)
Female	-0.012			
	(0.169)			
Treatment × Female	0.094			
	(0.217)			
Low SES		-0.206*		
		(0.091)		
Treatment × Low SES		0.754***		
		(0.210)		
Repeated a grade			-0.229	
			(0.130)	
Treatment × Repeated			-0.250	
			(0.221)	
Eligible				0.223
				(0.130)
Treatment × Eligible				-0.401
				(0.229)
Observations	1102	1102	1094	1102
R^2	0.011	0.037	0.034	0.017
Control mean	0.000	0.000	0.000	0.000

Notes: (1) This table shows the heterogeneous intent-to-treat (ITT) effects of the intervention, pooled across all years in which these outcomes were measured (see Table 2). (2) Low socio-economic status (SES) students are those in the first quartile of the first principal component from a principal component analysis of household assets. Students who repeated a grade are those who did so prior to the start of the program and study (i.e., during elementary school). Eligible students are those that would have been admitted to the program through its regular admissions process (see section 2.2). (3) The dependent variable is standardized with respect to the control group in the first year that these outcomes were measured. (4) The observations indicate the total number of students who participated in data collection across all years in which these outcomes were measured. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.35: Heterogeneous ITT effects on student achievement (2015-2016)

	Family index (std.)			
	(1)	(2)	(3)	(4)
Treatment	-0.042	-0.153**	-0.006	0.122
	(0.120)	(0.056)	(0.084)	(0.177)
Female	0.032			
	(0.122)			
Treatment × Female	0.052			
	(0.155)			
Low SES		-0.529***		
		(0.156)		
Treatment × Low SES		0.455**		
		(0.173)		
Repeated a grade			-0.411*	
			(0.182)	
Treatment × Repeated			-0.078	
			(0.181)	
Eligible				0.184
				(0.225)
Treatment × Eligible				-0.188
				(0.175)
Observations	714	714	709	714
R^2	0.225	0.253	0.261	0.227
Control mean	0.000	0.000	0.000	0.000

Notes: (1) This table shows the heterogeneous intent-to-treat (ITT) effects of the intervention, pooled across all years in which these outcomes were measured (see Table 2). (2) Low socio-economic status (SES) students are those in the first quartile of the first principal component from a principal component analysis of household assets. Students who repeated a grade are those who did so prior to the start of the program and study (i.e., during elementary school). Eligible students are those that would have been admitted to the program through its regular admissions process (see section 2.2). (3) The dependent variable is standardized with respect to the control group in the first year that these outcomes were measured. (4) The observations indicate the total number of students who participated in data collection across all years in which these outcomes were measured. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.36: Heterogeneous ITT effects on school performance (2014-2016)

	Family index (std.)			
	(1)	(2)	(3)	(4)
Treatment	-0.009	0.037	-0.009	0.121
	(0.106)	(0.074)	(0.071)	(0.098)
Female	0.227*			
	(0.121)			
Treatment × Female	0.135			
	(0.121)			
Low SES		-0.215		
		(0.142)		
Treatment × Low SES		-0.023		
		(0.194)		
Repeated a grade			-0.762***	
			(0.129)	
Treatment × Repeated			0.153	
			(0.135)	
Eligible				0.294***
				(0.076)
Treatment × Eligible				-0.099
				(0.132)
Observations	1064	1064	1055	1064
R^2	0.101	0.087	0.172	0.090
Control mean	0.000	0.000	0.000	0.000

Notes: (1) This table shows the heterogeneous intent-to-treat (ITT) effects of the intervention, pooled across all years in which these outcomes were measured (see Table 2). (2) Low socio-economic status (SES) students are those in the first quartile of the first principal component from a principal component analysis of household assets. Students who repeated a grade are those who did so prior to the start of the program and study (i.e., during elementary school). Eligible students are those that would have been admitted to the program through its regular admissions process (see section 2.2). (3) The dependent variable is standardized with respect to the control group in the first year that these outcomes were measured. (4) The observations indicate the total number of students who participated in data collection across all years in which these outcomes were measured. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.37: Heterogeneous ITT effects on personality traits (2016)

	Family index (std.)			
	(1)	(2)	(3)	(4)
Treatment	-0.168	0.112	0.109	0.323
	(0.144)	(0.141)	(0.130)	(0.224)
Female	0.081			
	(0.188)			
Treatment × Female	0.307			
	(0.172)			
Low SES		0.240*		
		(0.125)		
Treatment × Low SES		-0.509		
		(0.343)		
Repeated a grade			0.048	
			(0.291)	
Treatment × Repeated			-0.454	
			(0.347)	
Eligible				0.228
				(0.209)
Treatment × Eligible				-0.450
				(0.252)
Observations	341	341	339	341
R^2	0.035	0.029	0.033	0.027
Control mean	.	.	.	.

Notes: (1) This table shows the heterogeneous intent-to-treat (ITT) effects of the intervention, pooled across all years in which these outcomes were measured (see Table 2). (2) Low socio-economic status (SES) students are those in the first quartile of the first principal component from a principal component analysis of household assets. Students who repeated a grade are those who did so prior to the start of the program and study (i.e., during elementary school). Eligible students are those that would have been admitted to the program through its regular admissions process (see section 2.2). (3) The dependent variable is standardized with respect to the control group in the first year that these outcomes were measured. (4) The observations indicate the total number of students who participated in data collection across all years in which these outcomes were measured. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix B Additional information on data collection

B.1 Academic behaviors

The survey of academic behaviors asked students to recall the last time that they faced a challenging situation at school and specify the steps they had taken to prevent or address it. For example, the first question asked students to recall the last time they did not something covered in class and report whether they had: asked their teacher to explain the topic again, asked a relative to explain it, asked a peer, consulted a book/Internet on the topic, sought a private tutor, or attended after-school lessons. The score for this item ranged from 0 (if a student had not engaged in any of the listed behaviors) to 6 (if he/she engaged in all of them).

The nine other items asked students what they did when they: were assigned homework, failed a homework assignment, had to study for a test, failed a test, failed a term of a subject,[26] failed a subject, planned to miss a schoolday, were absent to school, or had a "free" period.[27] The maximum score for each item depended on the number of behaviors presented to students. Figures B.1 and B.2 display the distribution of raw scores for all questions in 2015 and 2016.

B.2 Academic mindsets

To measure academic mindsets, we used a survey of students' self-beliefs about academics. The survey, which was developed by psychologists at the University of Buenos Aires, asks students to indicate the extent to which they agreed with statements about themselves using a scale ranging from 1 ("strongly disagree") to 5 ("strongly agree"). It includes statements about performance (e.g., "I think I will get good grades this year") and efficacy (e.g., "I am capable of doing school assignments well, even if they are difficult"). It had already been administered to students in Argentina (see, e.g., Pais et al. 2013; Schmidt et al. 2008).

We also used the items from the Learning and Study Strategies Inventory (LASSI) that focus on motivation. The LASSI, which was developed by psychologists at the University of Texas at Austin (see Weinstein and Palmer 1988), asks students how often they find themselves in certain situations using a scale ranging from 1 ("never") to 5 ("always"). It includes situations about organization and planning (e.g., "I have trouble putting together a plan and sticking to it") and motivation (e.g., "I try hard to get good grades, even in subjects that I do not like"). Figures B.3 to B.5 display the distribution of raw scores for all questions from 2014 to 2016.

[26] In Argentina, students are typically awarded three grades per year (one per term) during the school year.
[27] When a teacher is absent, students are regularly allowed to stay in their classroom without studying.

B.3 Academic perseverance

To measure academic perseverance, we used the short Grit scale (Grit-S), which was developed by Angela Duckworth at the University of Pennsylvania (see Duckworth et al. 2007; Duckworth and Quinn 2009). The scale asks students to indicate the extent to which a series of statements describe them using a scale ranging from 1 ("not like me at all") to 5 ("very much like me"). It includes statements about consistency of effort (e.g., "I often set a goal but later choose to pursue a different one") and about perseverance of effort (e.g., "I finish whatever I begin").

We also used the Domain-Specific Impulsivity Scale for Children (DSIS-C), created by psychologists at the University of Pennsylvania (see Tsukayama et al. 2013). It asks students how often they engage in behaviors using a scale from 1 ("almost never") to 5 ("at least once a day"). It includes statements about inter-personal (e.g., "I interrupted others while they were speaking") and schoolwork impulsivity (e.g., "I forgot something I needed for school").

We also used a performance assessment known as CARAS in Spanish (i.e., "faces" in English), which was developed by American and Spanish psychologists (see Thurstone and Yela 2001). It shows students 45 sets of three smileys and asks them to cross out the smiley in each set that does not look like the others. Students are encouraged to complete as many as possible. Each student's "reflexivity index" is calculated by subtracting the number of incorrect answers from the number of correct answers. It had also been administered in Argentina (see Arán-Filipetti 2012; Arán-Filipetti and López 2013; Arán-Filipetti and Richaud de Minzi 2011). Figures B.3 to B.5 display the distribution of raw scores for all questions from 2014 to 2016.

B.4 Learning strategies

To measure learning strategies, we used the items from the LASSI that focus on organization and planning (see section B.2).

B.5 Student achievement

B.5.1 Design

The student assessments that we used were developed by psychometricians at the *Centro de Medición de la Universidad Católica de Chile* (MIDE-UC). It created its own items, drawing on the *Núcleos de Aprendizaje Prioritarios*, the contents from the national curriculum prioritized by the federal government, and publicly-released items from the national student assessment (called the *Operativo Nacional de Evaluación* at the time of the study).

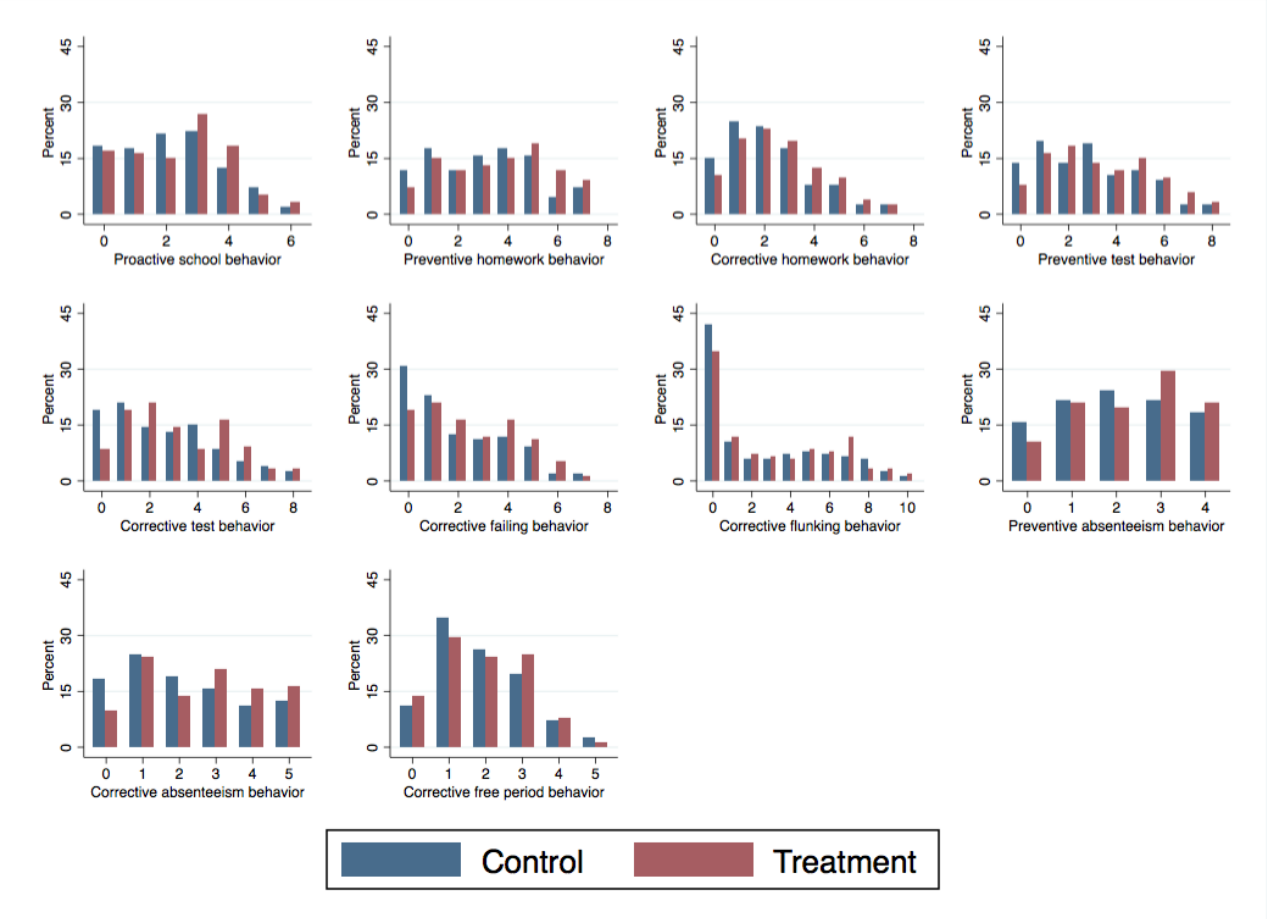
The tests were designed to assess a wide array of domains and skills in math and reading. The math test included 30 multiple-choice items that covered number properties, equations, probability, measurement, trigonometry, and statistics. It assessed students' capacity to identify mathematical concepts, understand and use symbolic math, perform calculations using various strategies, and solve abstract and applied problems. The reading test included 30 multiple-choice items that featured a historical passage, a descriptive passage, a poem, two movie reviews, and an excerpt from a fiction book. It assessed students' capacity to locate information in the text, understand the relationship between two parts of a text, identify the main idea of a text, and interpret the meaning of words from context.

B.5.2 Scoring, scaling, and linking

We scored all items dichotomously and used a two-parameter logistic (2PL) Item Response Theory (IRT) model to scale the results from both rounds of assessments (Harris 2005).[28] This model allows us to account for differences between items (specifically, in their difficulty and capacity to distinguish between students of similar ability). It also allows us to capitalize on the fact that the same assessments were administered in both rounds to map assessment results for both years (2015 and 2016) onto a common scale. We standardized scores with respect to the control group in 2015 to have a mean of 0 and a standard deviation of 1. Figure B.6 displays the distribution of scaled scores for 2015 and 2016.

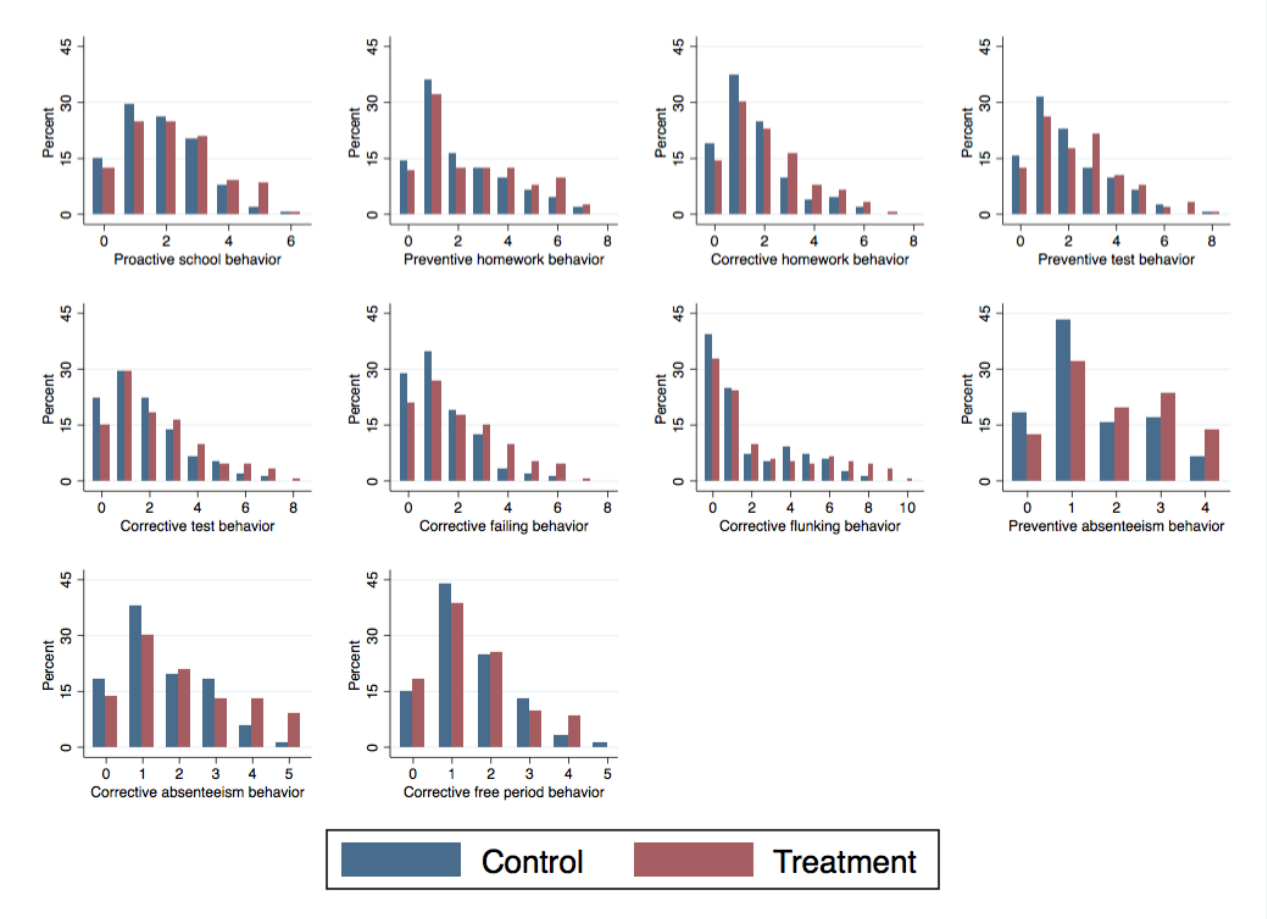
[28] We used the OpenIRT program for Stata written by Tristan Zajonc. Our choice of a 2PL instead of a 3PL model was based partly on the sampling requirements for 3PL models discussed in (Yen and Fitzpatrick 2006).

Figure B.1: Distribution of raw scores in survey of academic behaviors (2015)



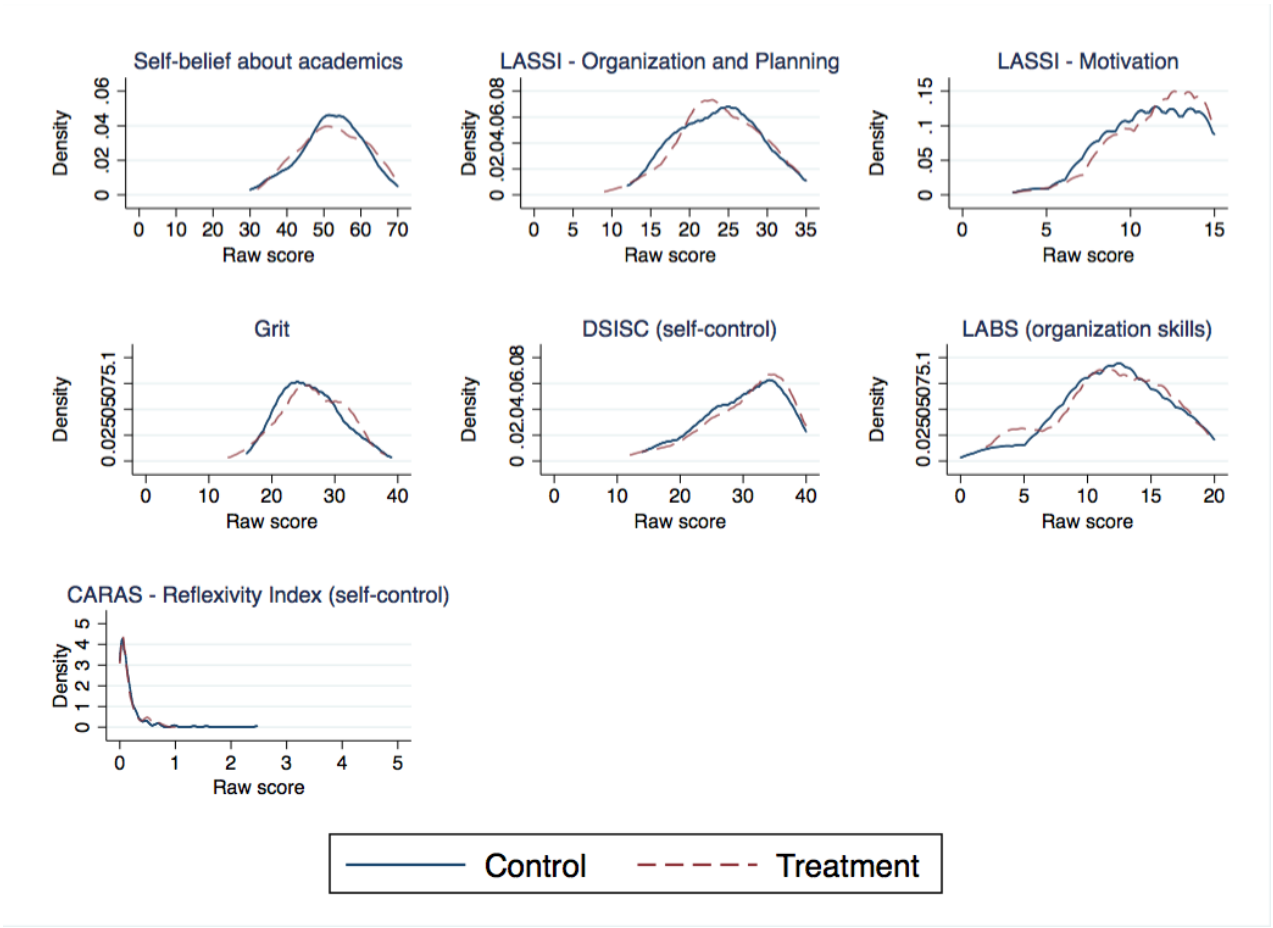
Notes: (1) The figure shows the distribution of raw scores on the survey of academic behaviors in 2015. (2) The maximum score for each item depended on the number of behaviors presented to students (see section B.1).

Figure B.2: Distribution of raw scores in survey of academic behaviors (2016)



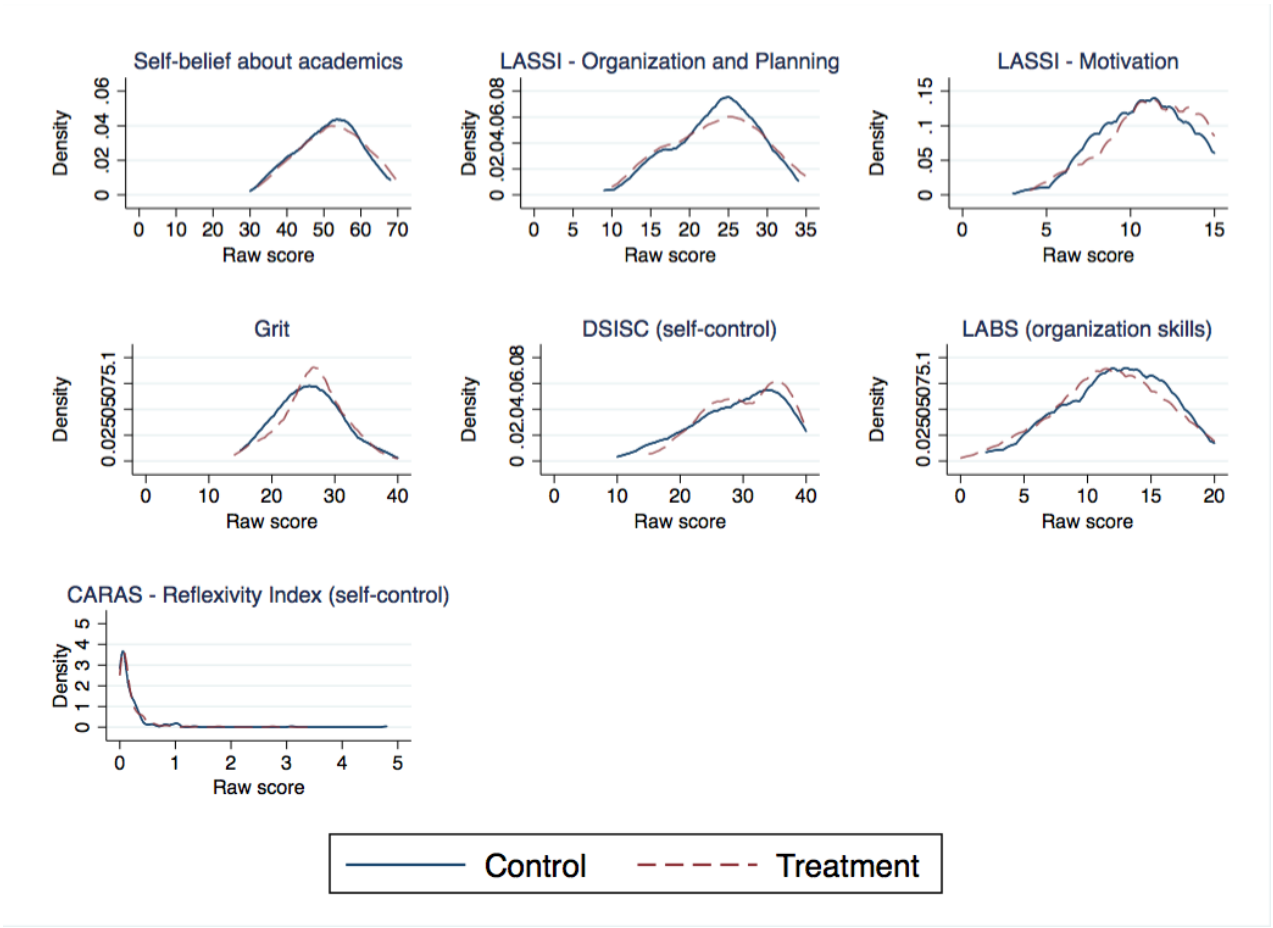
Notes: (1) The figure shows the distribution of raw scores on the survey of academic behaviors in 2016. (2) The maximum score for each item depended on the number of behaviors presented to students (see section B.1).

Figure B.3: Distribution of raw scores in instruments measuring academic mindsets, perseverance, and learning strategies (2014)



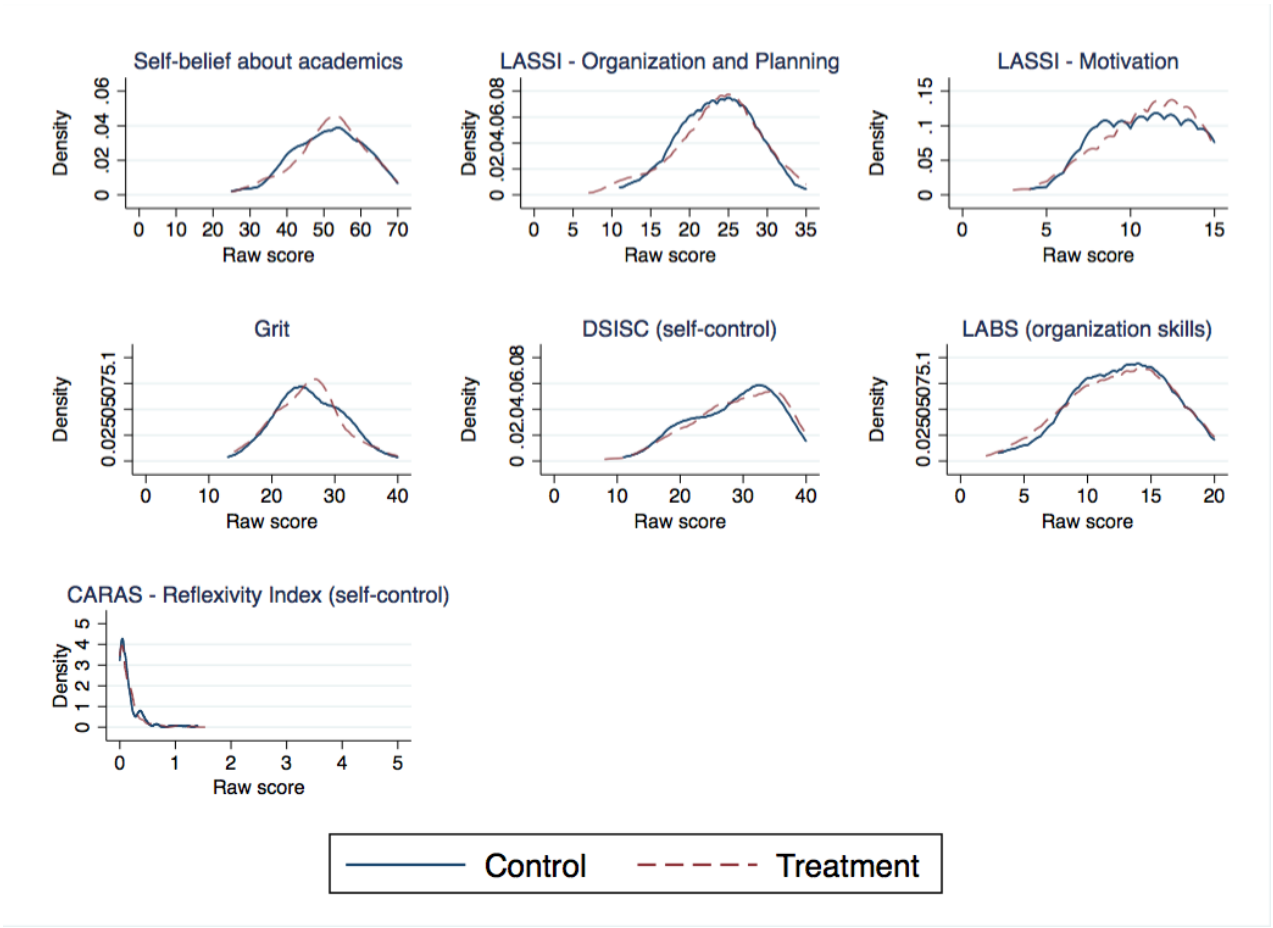
Notes: (1) The figure shows the distribution of raw scores on the survey of academic behaviors in 2014.

Figure B.4: Distribution of raw scores in instruments measuring academic mindsets, perseverance, and learning strategies (2015)



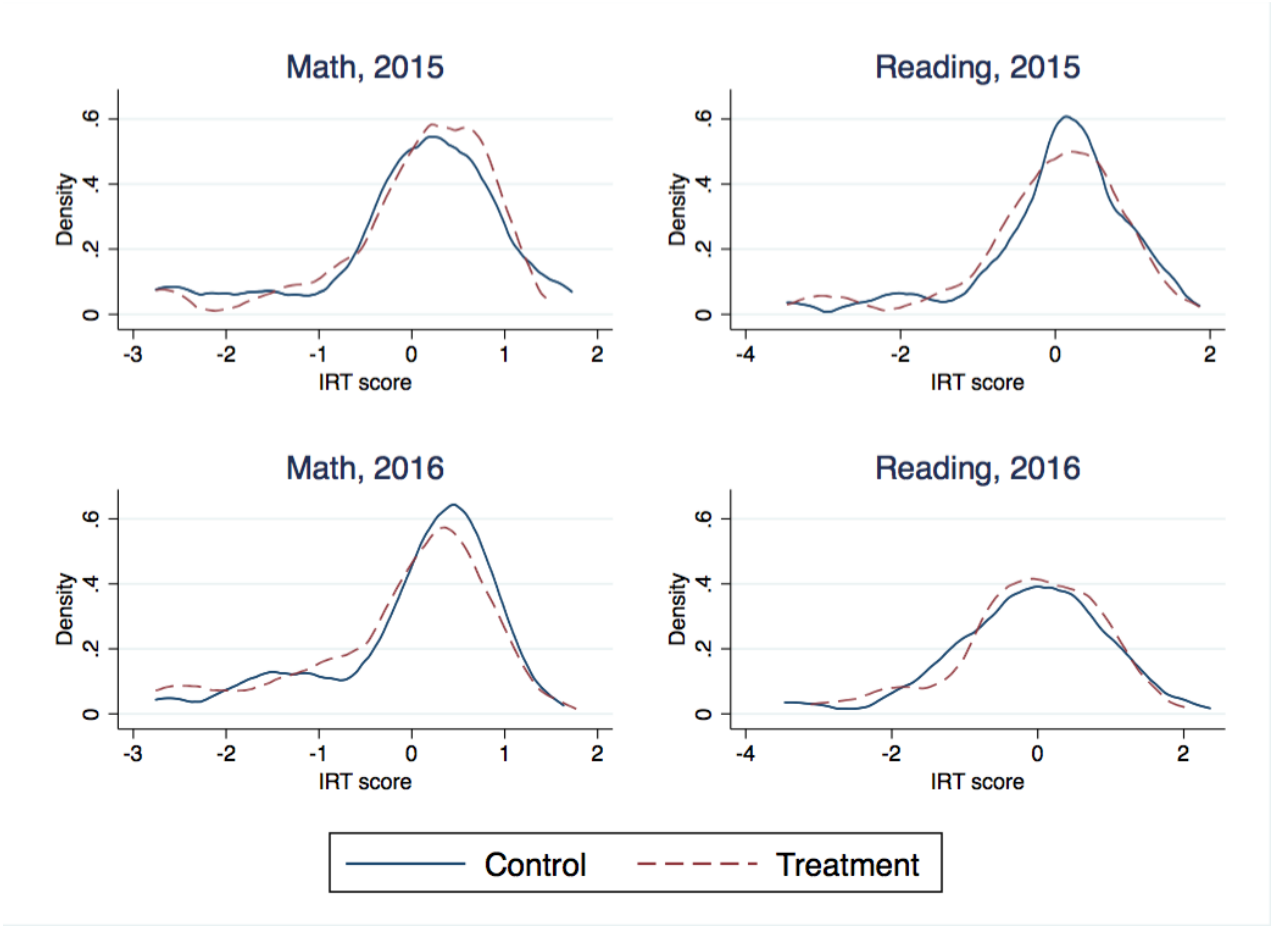
Notes: (1) The figure shows the distribution of raw scores on the survey of academic behaviors in 2015.

Figure B.5: Distribution of raw scores in instruments measuring academic mindsets, perseverance, and learning strategies (2016)



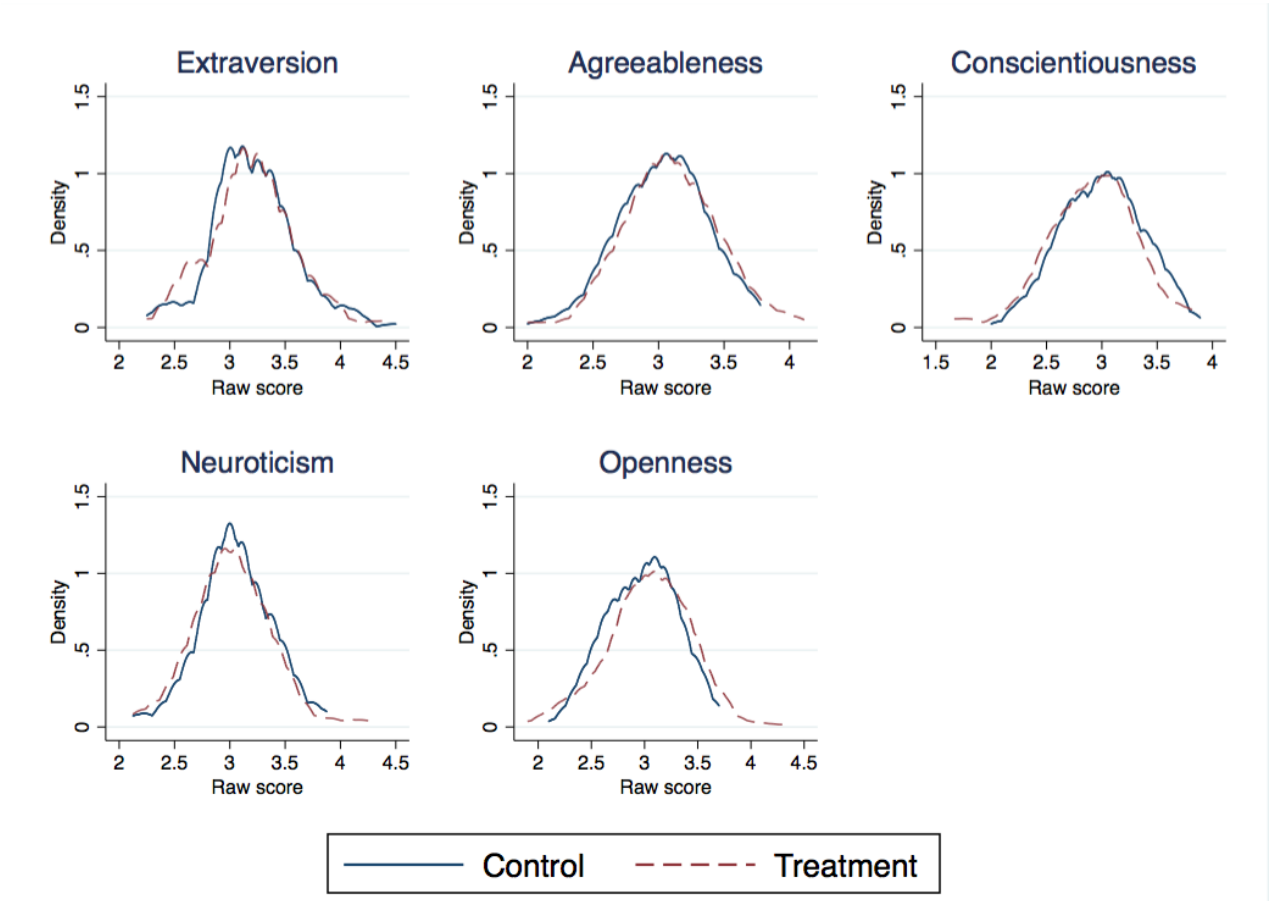
Notes: (1) The figure shows the distribution of raw scores on the survey of academic behaviors in 2016.

Figure B.6: Distribution of IRT-scaled scores on student assessments (2015-2016)



Notes: (1) The figure shows the distribution of scores in student achievement tests of math and reading, separately for the control and treatment groups, in 2015 and 2016. (2) The scores have been estimated using a two-parameter Item Response (IRT) model and standardized with respect to the control group in 2015.

Figure B.7: Distribution of raw scores in survey of personality traits (2016)



Notes: (1) The figure shows the distribution of raw scores on the survey of personality traits in 2016.