

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

Alejandro J. Ganimian
Felipe Barrera-Osorio
María Loreto Biehl
María Cortelezzi

Education Division
Social Sector

TECHNICAL
NOTE N°
IDB-TN-01508

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

Alejandro J. Ganimian
Felipe Barrera-Osorio
María Loreto Biehl
María Cortelezzi

August, 2018

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Hard cash and soft skills: experimental evidence on combining scholarships and mentoring in Argentina / Alejandro J. Ganimian, Felipe Barrera-Osorio, María Loreto Biehl, María Cortelezzi.

p. cm. — (IDB Technical Note ; 1508)

Includes bibliographic references.

1. Transfer payments-Argentina. 2. Scholarships-Argentina. 3. Soft skills-Argentina. 4. Mentoring-Argentina. 5. Academic achievement-Argentina. 6. Educational attainment-Argentina. I. Ganimian, Alejandro J. II. Barrera Osorio, Felipe. III. Biehl, María Loreto. IV. Cortelezzi, María. V. Inter-American Development Bank. Education Division. VI. Series.

IDB-TN-1508

<http://www.iadb.org>

Copyright © 2018 Inter-American Development Bank. This work is licensed under a Creative Commons IGO 3.0 Attribution-NonCommercial-NoDerivatives (CC-IGO BY-NC-ND 3.0 IGO) license (<http://creativecommons.org/licenses/by-nc-nd/3.0/igo/legalcode>) and may be reproduced with attribution to the IDB and for any non-commercial purpose. No derivative work is allowed.

Any dispute related to the use of the works of the IDB that cannot be settled amicably shall be submitted to arbitration pursuant to the UNCITRAL rules. The use of the IDB's name for any purpose other than for attribution, and the use of IDB's logo shall be subject to a separate written license agreement between the IDB and the user and is not authorized as part of this CC-IGO license.

Note that link provided above includes additional terms and conditions of the license.

The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the Inter-American Development Bank, its Board of Directors, or the countries they represent.



1300 New York Ave NW, Washington, DC 20577

Loreto Biehl, email: loretob@IADB.ORG

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

Alejandro J. Ganimian[†]

Felipe Barrera-Osorio[‡]

María Loreto Biehl[§]

María Cortelezzi[¶]

August 22, 2018

Abstract

Many developing countries offer cash to low-income families to encourage children to attend school. These initiatives have increased educational attainment, but they have rarely improved student achievement. One potential reason may be that beneficiaries may need additional support to develop the requisite “soft” skills to succeed in school. We conducted a three-year randomized evaluation of a program that provides secondary school students with scholarships and non-academic mentoring in the Province of Buenos Aires, Argentina. The program improved students’ academic behaviors (e.g., starting to study early before an exam or catching up on schoolwork missed due to absences). Yet, we find little evidence that it improved their academic mindsets (e.g., self-beliefs about performance and efficacy), perseverance (e.g., grit), or learning strategies (e.g., metacognition). The program also improved some metrics of school performance (e.g., language grades, student absenteeism, grade failure, and the number of failed subjects) on its first year, but we cannot detect similar gains in subsequent years. This fadeout may be due to the fact that a large share of treatment students were expelled from the program for not meeting its requirements. We do not find any evidence that the program improved students’ achievement in math and reading or their personality traits.

JEL codes: C93, I21, I22, I25.

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

---

*We gratefully acknowledge the funding provided by the Inter-American Development Bank and an anonymous donor for this project. We thank Marina Bassi, Mercedes Mateo Díaz, Qayam Jetha, Costas Meghir, Hugo Ñopo, Ernesto Pais, and Emiliana Vegas for their input at different stages of this project. We also thank Marina Bassi, Rafael de Hoyos, Timothée Demont, Gregory Elacqua, Isy Faingold Vigil, Ariel Fiszbein, Verónica Frisancho, Brad Hershbein, Richard Murnane, Ernesto Pais, Jeffrey Puryear, Laura Trucco, Michael Walker, and seminar participants at AEFPP, APPAM, Brookings, GDN, IDB, LACEA-IEN, the Ministry of Education and Sports of Argentina, UNU-WIDER, and UTDT for comments. The usual disclaimers apply.

[†]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.

[¶]Executive Director, *Proyecto Educar 2050*. E-mail: mcortelezzi@educar2050.org.ar.

1 Introduction

Many developing countries provide cash to low-income families to encourage children to attend school. Some of these initiatives are called “scholarships” and others “cash transfers”, but they operate under the same theory of change. Low-income parents may refrain from sending their children to school, or pull them out of school early, if they perceive that the costs of schooling are too high, its benefits are too low (or take too long to materialize), or they lack access to credit (Banerjee et al. 2013). Thus, these programs aim to (partly) cover the costs of and raise the (immediate) returns to schooling, while relaxing credit constraints by offering their beneficiaries cash to enroll and stay in school (Fiszbein et al. 2009).

Scholarship and cash transfers are among the most rigorously evaluated educational interventions in developing countries. According to a recent review, there are 47 evaluations of these programs in 20 countries (Ganimian and Murnane 2016). They have typically increased enrollment in school, but with few exceptions, they have not increased student achievement.

One potential reason why scholarships and cash transfers have not impacted student achievement is that the parents of beneficiaries have had little experience with schooling, which may limit their capacity to transfer productive academic mindsets, perseverance, and behaviors, as well as learning strategies and social skills to their children (Borghans et al. 2008; Farrington et al. 2012; Gabrieli et al. 2015). If the lack of these “character”, “socio-emotional”, or “soft” skills is a binding constraint for these children to succeed in school, mentoring that seeks to develop these skills could be a useful complement to scholarships or cash transfers.¹

This paper reports the results of a three-year randomized evaluation of a program that provides secondary school students in the Province of Buenos Aires, Argentina with a scholarship and non-academic mentoring. We assess the impact of the program on the outcomes that it seeks to influence directly (e.g., academic behaviors) and indirectly (e.g., academic perseverance and mindsets), as well as on other outcomes that may be plausibly impacted by this type of program (e.g., school performance and student achievement).

We present four main sets of findings. First, the program positively impacted students’ (self-reported) academic behaviors (e.g., starting to study early before an exam or catching up on schoolwork missed due to absences). These effects were large and consistent across years. Second, we find little evidence that the program improved students’ academic mindsets (e.g., self-beliefs about performance and efficacy), perseverance (e.g., grit), or learning strategies (e.g., metacognition), as measured by self-reports and performance assessments, but we cannot

---

¹There is a long-standing debate among economists, psychologists, and scholars in other fields over the correct label for such skills (Duckworth and Yeager 2015). In this paper, we use the three aforementioned terms interchangeably to refer to “patterns of thought, feelings, and behavior” (Borghans et al. 2008, p. 974) other than cognitive ability that lead to student success.

discard the possibility that it had small to moderate effects on these outcomes. Third, the program improved students' language grades and it reduced student absenteeism, grade failure, and the number of subjects failed on its first year. Yet, we cannot detect these effects in subsequent years, possibly due to the fact that a large share of treatment students were expelled from the program for not meeting its requirements. Fourth, we find no evidence that the program positively impacted students' achievement or personality traits. These were not target outcomes of the program, but we wanted to understand whether they would result from the expected improvements in school performance and socio-emotional skills.

The rest of the paper is organized as follows. Section 2 reviews prior research. Section 3 describes the context, intervention, sampling strategy, and randomization. Section 4 presents the data collected for this study. Section 5 discusses the empirical strategy. Section 6 reports the results. Section 7 discusses the policy implications.

2 Prior research

There are three common obstacles that low-income parents face when deciding whether to send their children to school (see Banerjee et al. 2013). First, the costs of doing so may be too high. These include *direct* costs (e.g., fees) (Barrera-Osorio et al. 2007; Borkum 2012; Liu et al. 2012; Lucas and Mbiti 2012), the costs of *complements* to schooling (e.g., transportation, uniforms, or textbooks) (Evans et al. 2009; Glewwe et al. 2009; Muralidharan and Prakash 2013), and the *opportunity* costs of not employing children at home or in the (informal) labor market (Del Carpio and Macours 2010; Skoufias et al. 2001). Second, the benefits from schooling may be too low or take too long to accrue. Specifically, the returns that parents expect for their children may be too low, given their understanding of these returns in the population, their private assessment of their children's skills, and their judgment of available schooling options (Dizon-Ross 2015; Jensen 2010, 2012; Loyalka et al. 2013). Third, parents may lack access to credit to cover schooling costs (Karlan and Linden 2014).

Scholarships and cash transfers were conceived to tackle these barriers to schooling (Fiszbein et al. 2009). They provide cash to low-income parents for enrolling and keeping their children in school. They aim to cover the costs of schooling, provide a short-term reward for a behavior that pays off over the long-term, and relax (or lift altogether) existing credit constraints.

Nearly every one of these programs that has been rigorously evaluated has increased schooling. Yet, the scope and magnitude of their impact has depended on the design of such initiatives and the characteristics of their beneficiaries (Ganimian and Murnane 2016).²

---

²Some design features that make a difference are whether cash is made conditional (Baird et al. 2011; Benhassine et al. 2013), the outcomes upon which it is made conditional (Barrera-Osorio et al. 2011), and

These initiatives, however, have been less successful in improving student achievement. Several studies that measured the impact of such programs on student learning have found no effect (Baez and Camacho 2011; Filmer and Schady 2014). There are two exceptions. There is some evidence that cash transfers may impact student learning in the long-run (Barham et al. 2014). Merit-based scholarships (i.e., awarded based on students' performance on an exam) have also increased students' test scores (Barrera-Osorio and Filmer 2016; Kremer et al. 2009).

Admittedly, most scholarships and cash transfers were not designed to improve student achievement. Yet, it seems reasonable to expect that if these programs increase beneficiaries' participation in school, they should also learn more.³

One potential reason why these programs have typically failed to impact student achievement (especially, in secondary school) is that the parents of beneficiaries have had little experience with schooling, which may limit their capacity to transfer academic mindsets (e.g., the beliefs that effort can increase competence in school, that it is possible to succeed in school, and that doing well in school matters), perseverance (e.g., grit, delayed gratification, self-discipline, self-control), and behaviors (e.g., doing homework, organizing materials, participating, studying), as well as learning strategies (e.g., study skills, metacognitive strategies, self-regulated learning, and goal-setting) and social skills (e.g., interpersonal skills, empathy, cooperation, assertion, and responsibility) to their children (Farrington et al. 2012). If the lack of these mindsets, behaviors, and skills is a binding constraint for these children to succeed in school, increasing their access to school is unlikely to improve their achievement.

Mentoring could potentially improve students' socio-emotional skills, but to our knowledge, there is only one randomized evaluation of non-academic mentoring in developing countries. Huan et al. (2014) evaluated the effect of a program that designated a music, art, or physical education teacher to deliver 32 fully-scripted, 45-minute weekly counseling sessions to students in grades 7 and 8 in Shaanxi, China in 2012. The intervention sought to help students deal with "learning anxiety" and stressful relationships with teachers and peers. On average, it reduced learning anxiety and dropout rates after six months, but effects faded after a year. Only students at high risk of dropping out continued to benefit after that period. The authors argued that this fadeout was due to decreasing student interest in the program. This study leaves a number of questions unanswered, including whether mentoring would be more effective if it catered to students' individual needs and whether it could affect other outcomes. Further, while this mentoring program seems appropriate for school systems that place considerable

---

treatment exposure (Behrman et al. 2009, 2011; Dammert 2009; Perova and Vakis 2012). Some characteristics of beneficiaries that matter are their age (Maluccio and Flores 2005; Schultz 2004) and socio-economic status (Galiani and McEwan 2013).

³This expectation seems less reasonable in lower-middle income countries where disadvantaged students already lag far behind their peers by primary school and have little chance of understanding the material taught in school (Muralidharan et al. 2016; Muralidharan and Zieleniak 2014; Pritchett and Beatty 2015). Yet, it seems more reasonable in upper-middle income countries such as the one we study.

emphasis on school performance and high-stakes exams, it is less relevant for other systems. Our study contributes to this literature by addressing these unanswered questions in a context of greater relevance for upper-middle income countries.

3 Experiment

3.1 Context

Schooling in Argentina is compulsory and free from age 4 until the end of secondary school. In 12 out 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 (2013b).⁴ According to the latest official figures from 2013, the Argentine school system serves nearly 11 million students, including 1.7 million in pre-school, 4.6 million in primary school, and 3.9 million in secondary school (DiNIECE 2013a). The school calendar starts in February and ends in December.

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

Argentina is an interesting setting for exploring the potential complementarities of scholarships and mentoring because it combines high access to secondary school with low graduation and achievement. The country expanded access to secondary before most other Latin American nations. 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 (Busso et al. 2013).⁵ Yet, its secondary graduation rate lags behind those of its middle-income neighbors. In 2011, it stood at 41%, compared to 64% in Brazil, 84% in Chile, and 44% in Mexico (OECD 2014).⁶ Most secondary school students in Argentina do not reach national standards of achievement. The latest national student assessment, *Aprender* 2016, found that 42% of sixth graders performed at the two lowest levels in math and 33% in language (SEE-MEDN 2017).

---

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

⁵By the late 2000s, Argentina’s enrollment advantage persisted: 75% of secondary school age youths were enrolled on time, compared to 59% in the average country in the region (Busso et al. 2013).

⁶Further, youths from low- and high-income families have very different chances of graduating from secondary school. In 2011, 39% of secondary school age youths from the lowest income quintile graduated from school, compared to 81% of their peers in the highest income quintile (Alfonso et al. 2011).

3.2 Treatment

The Scholarship and Mentoring Program (SMP) that we evaluate in this study provides a scholarship and a non-academic mentoring session each month to secondary school students throughout the academic year. It run by one of the largest education non-profits in Argentina.⁷ It is the longest-standing program of its kind in the country: it has been in place since 1997. It is also the largest such program run by a non-profit in Argentina: in 2015, it reached 2,544 students across 16 of the 24 provinces of the country.

Typically, students are invited to apply to the program in their last year of secondary school (grade 7). Representatives from the non-profit interview all applicants (with their parents) and select which students will be invited to join.⁸ If students are admitted, they can remain in the program from the first (grade 8) to the last grade (grade 12) of secondary school, provided that they comply with three requirements: (a) they remain enrolled in a program-affiliated school; (b) they do not repeat grades; and (c) they are not suspended from school.⁹ For the purposes of this evaluation, the non-profit agreed to forgo its regular selection process, invite students to apply, and allow us to randomly select those who would be admitted. However, it maintained its three requirements for students to remain in the program.

We considered, but decided against, adding two experimental arms to this study to evaluate the impact of scholarships and non-academic mentoring by themselves for three main reasons. First, as we discussed in Section 2, cash transfers and scholarships are by far the educational intervention that has been evaluated most often in developing countries, and in nearly all cases, they have increased educational attainment and have not improved student achievement. Thus, we did not believe what we would learn anything from evaluating the scholarship by itself that we did not already know from prior work. Second, although there are far fewer evaluations of non-academic mentoring in developing countries, our main motivation for evaluating this program was to shed light on the question of why cash transfers and scholarships have rarely improved learning, so we decided to focus on testing their complementarity with mentoring. Third, the non-profit that provided the program had never offered mentoring by itself, and we anticipated that it might encounter several “growing pains” from doing it for the first time (especially, ensuring that students take up the mentoring without any financial incentives).

---

⁷The non-profit has decided to remain anonymous.

⁸The non-profit admits students based on six criteria: (a) whether they are enrolled in a school that is affiliated with the program; (b) the extent to which they have maintained a “good academic standing”; (c) whether they are from a “low-income family”; (d) the extent to which they show “interest in and commitment to schooling”; (e) whether they have a parent or legal guardian (with valid identification); and (f) whether they or they siblings already receive another scholarship. The non-profit does not have a standard set of metrics to measure criteria (b), (c), or (d). Admissions are made on a case-by-case basis by an internal committee.

⁹The first criterion is due to logistical constraints; mentors cannot follow all students who switch schools. The second and third criteria are meant to encourage students to do well in school.

Therefore, we wanted to avoid confounding the (potentially positive) effects of the mentoring component with the (potentially negative) effects of the learning curve for the non-profit.

3.2.1 Scholarship

Each student who participates in the SMP is supposed to receive a monthly scholarship during the school year (i.e., 10 scholarships in total, from March to December of each school year). Each scholarship is about USD 40, which is on par with the amount offered in cash transfer programs for secondary school in developing countries (Fiszbein et al. 2009). The scholarship is deposited on a month-by-month basis in a bank account in the name of each student’s parent or legal guardian. The funds may be withdrawn at any time and used for any purpose.

3.2.2 Mentoring

Each student who participates in the SMP is also supposed to receive a monthly non-academic mentoring session during the school year (i.e., 10 sessions in total, from March to December of each school year). Each session lasts 30-45 minutes and is typically held at the school, before or after the school day.¹⁰ Sessions may be one-on-one (i.e., between one mentor and one student) or group-based (i.e., between one mentor and multiple students from the same school). Each mentor decides the breakdown of individual and group meetings for each student. Mentors may also invite parents or legal guardians to join these meetings.

It is important to note that mentors in this program differ considerably from their counterparts in similar initiatives. First, they are paid employees of the non-profit, they are not volunteers. Second, they have all completed tertiary or university education, they are not high school or college students. Most mentors have a bachelor’s degree in psychology, developmental psychology, social work, or education, or they have graduated from a teacher training program. Third, mentors receive both pre-service and in-service training and support. They undergo an induction process when they join the organization and they participate in two training sessions each year (one that is conducted in-person and another one that is delivered online). They are also assigned to a coach who has experience as a mentor and supervises their work. Finally, they have access to an online platform through which they share resources and ideas. Accordingly, mentors are granted considerable discretion in how to run mentoring sessions. The content of the sessions is not standardized and/or pre-determined by the non-profit; it is decided upon by each mentor based on the needs of each individual student. However, all mentoring sessions share a common structure: (a) they begin with an “icebreaker”, in which the mentor seeks to earn the trust of the student, and the student shares his/her schoolwork,

---

¹⁰Students are only pulled out of the classroom when it is not possible to meet them before or after school.

as well as a number of reports from teachers and school staff required by the program;¹¹ (b) they proceed with a “diagnosis”, in which the student discusses his/her experience at school and shares his/her strengths and weaknesses; and (c) they end with an “action plan”, in which the student and mentor agree on goals for the next session (e.g., studying for a math exam). Finally, mentors also have full discretion to suspend or expel students from the program.¹² They typically make this decision on the first and last mentoring session of each year, when they review the progress made by their mentees. The grounds for expelling students from the program are determined by the non-profit, but the grounds for suspending students are not clearly-defined, and as we discuss in Section 4.2.1, mentors vary widely in the extent to which they resort to expulsions and suspensions.

3.2.3 Theory of change

Table 1 presents the theory of change of the SMP. The program seeks to address two sets of factors that may prevent youth from disadvantaged backgrounds from staying and succeeding in secondary schools: (a) the multiple costs of schooling, their potential misperceptions about the benefits from graduating from secondary school, and their lack of access to credit; and (b) the potential lack of experience that their parents may have had with secondary school, which limits their capacity to support their children. The scholarship seeks to address the first problem and mentoring sessions the second problem.

[Insert Table 1 here.]

As the table indicates, the non-profit expects parents to use the scholarship to offset the costs of their children’s schooling and students to participate in the mentoring sessions regularly. As a result, the program is meant to make students more likely to attend school regularly, stay in school for longer, and adopt academic behaviors that will help them do better in school. The regular exercise of these behaviors is in turn expected to lead students to increasingly adopt academic mindsets (e.g., update their beliefs about their own performance and efficacy), persevere in the face of challenges, and improve their learning strategies. The expectations of the non-profit seem to be well-aligned with existing theory and evidence on the role of socio-emotional skills (see, for example, Farrington et al. 2012).

The non-profit does not expect the program to necessarily impact other outcomes. However, if the program had an impact on students’ academic behaviors, as well as on their mindsets,

---

¹¹Students are expected to show: (a) their folders, which contain their work on all the subjects that they take at school (monthly); (b) their attendance and discipline certificates, which are completed by a staff member of the school (monthly); (c) a report from the same staff member (biannually); (d) their school report cards; and (e) a report from one of their teachers (biannually).

¹²When students are suspended from the program, they do not receive the scholarships or mentoring.

perseverance, and learning strategies, we thought that it was important to understand whether it ultimately impacted students' performance in school, how much they learned at school, and potentially even more stable personality traits. These potential effects of the program are also in line with evolving theory and evidence on the complementarity between academic and socio-emotional skills (see, for example, Borghans et al. 2008).

3.2.4 Costs

We collected the costs of the program using the approach outlined in Dhaliwal et al. (2013). It costs USD 733 per student per year (Table A.1 in Appendix A). More than half of the costs are due to the scholarships (i.e., the cash transfers and the costs associated with distributing it). The other half is spent on the mentoring sessions (27%), administration of the program (7%), supervision and monitoring (6%), staff training (4%), and identifying/selecting students (3%).

3.3 Sample

We conducted this study in the Province of Buenos Aires (PBA), which is an ideal setting to pilot programs that could be scaled to the rest of Argentina. First, it by far is the largest sub-national school system in the country. In 2012, it had 4,442 secondary schools and nearly 1.5 million students from grades 8 to 12 (DiNIECE 2013a). Second, PBA students perform on par with the average student in the country. In the latest national student assessment, 42% of grade 6 students in PBA performed at the two lowest achievement levels in math and 33% of them performed in the lowest two levels in language (SEE-MEDN 2017).

We invited 10 public secondary schools in the PBA to participate in the study based on three criteria set out by the non-profit: (a) they had to serve students from low-income families (because this is the target demographic of the program); (b) they had to have previously participated in the SMP (because it was preferable that they were familiar with the requirements of the program); and (c) they could not be participating in the SMP at the time of recruitment (because the non-profit wanted to avoid having students chosen by the regular admission process and by lottery in the same school). A representative from the non-profit met with the principal of each of these schools to invite them to participate the evaluation.

All schools accepted to participate. Each school was located in a different district (*localidad*) within the PBA, including: Campana, Ensenada, Gregorio de Laferrere, Guernica, José C. Paz, Merlo, Quilmes, Santos Lugares, Virrey del Pino, and Zárate.

We invited a random subset of grade 7 students at each school to participate in the study. We randomly selected two grade 7 sections per school, the non-profit sent the parents of these

students invitations to attend a recruiting session, and representatives from the non-profit held multiple sessions at each school to accommodate parents' scheduling needs. Then, a representative from the non-profit met with all interested parents and their children to explain the requirements of the program and the evaluation and collect their baseline data. This process resulted in 408 interested students who were entered into a lottery roster. The number of available spots at each school was determined by the amount of funds that the non-profit had raised to cover the costs of the program at that school.¹³

3.4 Lottery

We randomly assigned students in the lottery roster to a treatment group, which was offered the program, or to a control group, which was not offered the program. This lottery resulted in 204 students in each group. As it is typically the case with the program, the offer of the program allowed students to potentially remain in the program until they graduated from high school, provided that they complied with the program's requirements (see Section 3.2).

4 Data

We designed our data collection strategy to track the impact of the program on each step of our hypothesized theory of change (see Section 3.2.3), including the outcomes that the non-profit expected to impact directly and indirectly, as well as the outcomes that we believed could be plausibly affected by the program. Table 2 provides an overview of all rounds of data collection, including the date, share of participants, and location of each round.

[Insert Table 2 here.]

4.1 Baseline

We administered a student and a household survey prior to randomization. The survey of students included questions on their demographic and educational background. The survey of households asked one parent or legal guardian (typically, the mother) about the educational background of the student's relatives, the amount and distribution of funds spent on his/her education, and his/her housing conditions and household assets.

Table 3 presents selected summary statistics for both surveys. As the table shows, the non-profit succeeded in recruiting the type of students for which it designed the program.

---

¹³Many donors of the program are individuals or corporations who have some type of relationship to the school or the community it serves, so donations are typically school-specific.

Almost a third of students in our sample had repeated at least one grade and 5% of them had dropped out of school prior to the start of the evaluation. The assets that these students had at their homes, as well as their living conditions, indicate that these students were from low-income families. Only 21% of students in our sample had a car, 72% had a refrigerator, and 55% had a computer. Just 30% of them had natural gas, 83% had running water, and only 60% of their parents were homeowners.

[Insert Table 3 here.]

As the table shows, control and treatment students were comparable at baseline. We only found five statistically significant differences between experimental groups in six variables: whether students had previously dropped out of school; whether they had a car, natural gas, a fridge, or a cell phone at home; and whether their parents were homeowners. To check whether differences on these individual variables (which are what we would expect simply by chance) are indicative of broader contrasts between groups, we run joint tests using seemingly unrelated regressions (SURs). As the last two rows of the table show, we cannot reject the null hypothesis of no difference across groups in these variables ($\chi^2 = .055$, $p = .815$).

4.2 Follow-ups

As stated above, we tracked students for three years to collect data on our outcomes of interest. At each data collection round, we tried to locate and survey all students in our lottery roster, regardless of whether they remain enrolled in their original school, lived in their original home, or remained enrolled in the program (in the case of treatment students). This section offers an overview of the instruments that we used, but Appendices B through D include further details and show the distribution of responses by experimental group and data collection round.

4.2.1 Program participation

The program was implemented mostly as intended. In 2014, the average treatment student received eight scholarships and was offered nine mentoring sessions (seven individual and two group sessions), of which they attended eight. On average, parents were invited to six sessions, of which they attended five. Nearly all students had the same mentor throughout the year. These figures were similar for students still in the program in 2015 and 2016 (Table A.3).

Importantly, however, the exposure of the average treatment student to the program decreased over the course of the evaluation due to three main reasons. First, two students decided not to join the program in 2014. Second, a large share of students was suspended at least once

from the program for not complying with its requirements: 26% of the total in 2014, 34% of students still in the program in 2015, and 31% of students still in the program in 2016. The average treatment student received 1.4 suspensions in 2014, 2 in 2015, and 2 in 2016.¹⁴ Third, a large proportion of students were expelled from the program, either during the school year (6 students in 2014, 10 in 2015, and 3 in 2016) or at the end of the year (5 students in 2014, 26 in 2015, and 34 in 2016) for not complying with the requirements discussed in Section 3.2.¹⁵ By the end of 2016, only 120 of the original 204 treatment students remained in the program.

4.2.2 Academic behaviors

We collected data on students' academic behaviors in 2015 and 2016. The instrument was developed by the research team and it enquired about students' self-reported preventive and corrective behaviors at school. It included questions about general proactive behaviors (e.g., asking teachers to explain confusing concepts); preventive behaviors related to homework, tests, and absenteeism (e.g., reviewing the textbook before a test); and corrective behaviors related to homework, tests, failing a subject, failing a grade, absenteeism, and free periods (e.g., asking a peer for missed schoolwork). Appendix B describes each domain of the survey and displays the distribution of raw scores for each set of questions in 2015 and 2016.

4.2.3 Academic mindsets, perseverance, and learning strategies

We also collected data on students' academic mindsets, perseverance, and learning strategies in 2014, 2015, and 2016. All rounds included the same six instruments, which we selected based on three main criteria. First, we identified instruments that captured students' academic mindsets, perseverance, and learning strategies, as defined by Farrington et al. (2012). Then, we shortlisted those instruments that allowed us to contrast self-reports and performance assessments, in light of ongoing debates about the perils of relying solely on the former (see Alan et al. 2016; Borghans et al. 2008; Duckworth and Yeager 2015; West et al. 2015). Lastly, we chose instruments that had already been administered and validated in Argentina.

To measure students' academic mindsets, we used: (a) a survey of self-beliefs about academics, including beliefs about performance and efficacy; and (b) the questions of the Learning and Study Strategies Inventory (LASSI) that measure students' motivation. To capture students' academic perseverance, we used: (a) the short Grit survey (Grit-S), which focuses

---

¹⁴Suspensions 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.

¹⁵Expulsions 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.

on perseverance and passion for long-term goals; (b) the Domain-Specific Impulsivity Scale for Children (DSIS-C), a survey of self-control; (c) *Caras* (Smileys), an assessment of self-control; and (d) Labs, an assessment of students’ planning skills. Finally, to measure students’ learning strategies, we used the questions of the LASSI on students’ organization and planning skills. Appendix C briefly describes each instrument.

4.2.4 School performance

We collected data on students’ performance at school from 2014 to 2016, including: (a) the number of absences; (b) their math and language grades; (c) whether they passed their grade; (d) the number of subjects that they failed;¹⁶ (e) whether they transferred schools; and (f) whether they dropped out of school.

4.2.5 Student achievement

We administered standardized tests of math and reading in 2015 and 2016.¹⁷ They were designed by the *Centro de Medición de la Universidad Católica de Chile* (MIDE-UC) to assess what students ought to know and be able to do according to Argentina’s standards, including: (a) the *Núcleos de Aprendizaje Prioritario* (NAPs), the contents that the government has prioritized from the national curriculum; and (b) publicly-released items from national student assessment. We scaled the results using Item Response Theory (IRT) to place students on a common scale across both rounds of data collection. Appendix D discusses test design and scoring and shows the distribution of scores for both experimental groups in 2015 and 2016.

4.2.6 Personality traits

Lastly, we administered the Big 5 Inventory to measure students’ personality traits in 2016. This widely-used survey focuses on five facets: extraversion, agreeableness, conscientiousness, neuroticism, and openness (John et al. 2008; John and Srivastava 1999). The version that we administered asked students to use a scale ranging from 1 (“totally disagree”) to 5 (“totally agree”) to express whether they self-identified with 44 statements (e.g., “I see myself as someone who is talkative”). Figure A.1 shows the raw score distributions on each facet.

---

¹⁶In 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, 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.

¹⁷We administered the same assessment on both occasions, as they included items for grades 9 and 10, the corresponding grades for students in the second and third year of our evaluation cohort.

4.3 Attrition

We find that control students are more likely to fail to participate from follow-up rounds of data collection than their treatment peers. We estimate differential attrition across experimental groups in two ways. First, we compare attrition in each round *with respect to the baseline*.¹⁸ We find that control students are more likely to attrit from all rounds than their control peers, but that the difference between groups is only statistically significant in follow-up round 1. Control students were 4.4 percentage points more likely to attrit. Then, we compare attrition in each round *with respect to the previous round*.¹⁹ We find that control students are less likely to attrit from nearly all rounds, but that the difference between groups is only statistically significant in follow-up rounds 1, 4, and 5.²⁰ Control students were 4.2 and 2.4 percentage points more likely to attrit in rounds 4 and 5, respectively (Table A.4).

Yet, we also find that this differential attrition does not change considerably the student composition at each follow-up round of data collection. We check balance across experimental groups on selected variables from the baseline student survey at each round.²¹ We find that control students are more likely to be female, to be older, to attend school in the morning, and to have repeated a grade prior to the study, but none of these differences are statistically significant. The only statistically significant difference is that control students are between 0.2 and 4.2 percentage points more likely than their control peers to have dropped out of school prior to the study, depending on the data collection round (Table A.5).

In light of these results, we do not see many reasons to believe that differential attrition has compromised the equality of expectation assumption between experimental groups. Nevertheless, to address this possibility, we estimate both inverse-probability weighted treatment effects and Lee (2009) bounds for each set of outcomes in our study.

---

¹⁸Specifically, we run a regression of a dummy for attritors in each round (which equals 1 if a student did not participate in that round and 0 otherwise) on the treatment dummy and randomization fixed effects, adjusting the standard errors to account for within-school correlations across students in outcomes.

¹⁹We run the same regression as above, but we set the attrition dummy to equal 1 if a student participated in the previous round but not in the following one and 0 otherwise.

²⁰Note that the attrition dummies for follow-up round 1 are identical in both cases because they are both defined with respect to baseline.

²¹We run a regression of each variable on the treatment dummy and the randomization fixed effects separately for each round, adjusting the standard errors to account for within-school correlations across students in outcomes.

5 Empirical strategy

5.1 Intent-to-treat (ITT) effect

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

$$Y_{it} = \alpha_j + \lambda_t + \beta T_i + \gamma X_i + \epsilon_{ijt} \quad (1)$$

where Y_{it} is the outcome of interest for student i at time t , α_j are school (i.e., randomization block) fixed effects, λ_t are year fixed effects, T_i is a dummy indicating whether the student was offered a spot in the SMP, X_i is an index of family income at baseline,²² and ϵ_{ijt} is the idiosyncratic error term. We adjust the standard errors to account for within-school correlations across students in outcomes. The coefficient of interest is β , which indicates the magnitude of the effect of the offer of a spot in the SMP across all rounds of data collection for a given outcome. We also estimate a version of equation (1) that accounts for the location of data collection (i.e., whether it was conducted at home or at school).²³

We also estimate the ITT effect separately by year by fitting:

$$Y_i = \alpha_j + \beta T_i + \gamma X_i + \epsilon_{ij} \quad (2)$$

where everything is defined as above. However, we do not present these as our main estimates to minimize the possibility of focusing on false positives due to multiple hypothesis testing. We include these results for each set of outcomes in Appendix A and discuss them in the text.

We also address the possibility of false positives in our pooled ITT estimates in two ways. First, following Barrera-Orsorio and Filmer (2016), we run joint tests using Seemingly Unrelated Regressions (SURs) to check whether effects on individual outcomes are indicative of broader impacts in the construct being measured (e.g., academic behaviors). Second, following Kling et al. (2005), we also aggregate all outcomes measuring the same construct into a family index, which is standardized with respect to the control group.

---

²²This index is the first principal component from a principal component analysis of household assets (dummies for whether students had a car, natural gas, running water, a bathroom, a solid floor, a fridge, a computer, Internet, and a cell phone at home, and whose parents were homeowners) collected at baseline.

²³This specification does not affect any of our impact estimates, so we exclude them from this paper. However, all results are available from the authors upon request.

5.2 Heterogeneous ITT effect

We also explore whether the offer of the program differentially impacts girls, students who had previously repeated a grade, students from low-income families,²⁴ or students who would have been admitted to the program through its regular selection process²⁵ by fitting:

$$Y_{it} = \alpha_j + \lambda_t + \beta T_i + \delta G_i + \kappa I_i + \epsilon_{ijt} \quad (3)$$

where G_i indicates whether student i belongs to the subgroup of interest, I_i is the interaction between T_i and G_i , and everything else is defined as above. The coefficient of interest is κ , which indicates the magnitude of the differential effect of the program on the sub-group. Tables with statistically insignificant effects are omitted but available from the authors.

5.3 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 two-stage least squares (2SLS) instrumental variables (IV) model:

$$\begin{aligned} A_{it} &= \phi_j + \zeta_t + \mu T_i + \eta_{ijt} \\ Y_{it} &= \psi_j + \omega_t + \nu \hat{A}_i + \varepsilon_{ijt} \end{aligned} \quad (4)$$

where ϕ_j and ψ_j are school (i.e., randomization block) fixed effects, ζ_t and ω_t are year fixed effects, A_{it} is the number of months in which a student has received both a scholarship and a mentoring session (which is zero for all control students), η_{ijt} and ε_{ijt} are the idiosyncratic error terms, and everything else is defined as above. The coefficient ν indicates the relationship between each month of the combined intervention and the outcome.

We also estimate the dose-response relationship between the number of scholarships or the number of mentoring sessions received and the outcomes of interest. Yet, that is not the TOT effect of each component of the program. On any given month of the school year, a student may receive a scholarship, a mentoring session, or both.²⁶ Therefore, an estimation of the dose-response relationship between the number of scholarships and an outcome of interest

---

²⁴We define students from low-income families as those in the lowest quartile of the first principal component of a principal component analysis of household assets (dummies for students who had a fridge, a computer, Internet connection, natural gas, and running water at home) collected at baseline.

²⁵Even if, as specified in Section 3.2, the non-profit agreed to forgo its regular selection process, it still conducted non-binding admissions interviews and recorded which students would have been admitted if the usual process were followed. This allows to investigate whether the program differentially impacts the type of student who is typically admitted to the program.

²⁶For 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.

instrumented by the random assignment does not meet the exclusion restriction because the instrument also affects the outcome through the number of scholarships (Angrist et al. 1996). We could control for the number of mentoring sessions, but the correlation between the number of scholarships and mentoring sessions among treatment students is very high (between .83 and .96, depending on the year), so this estimation would rely on variation among a few students. The same argument can be made about the estimation of the dose-response relationship between the number of scholarships and the outcome of interest.

6 Results

6.1 Academic behaviors

We find evidence that the SMP improved students' academic behaviors. As Table 4 shows, treatment students reported engaging in preventive and corrective behaviors more often than their control peers in 2015 and 2016. The effects were positive, moderate to large (between .15 and $.31\sigma$), and remained statistically significant after the inclusion of baseline covariates. All but one of these effects emerged both in 2015 and 2016 and were consistent across years in their sign, magnitude, and statistical significance (Table A.6).

[Insert Table 4 here.]

There are two patterns in the distribution of students' responses worth highlighting. First, the differences between experimental groups on some indexes (e.g., corrective test behavior, corrective failing behavior, and corrective absenteeism behavior) are partly driven by a large share of control students who did not engage in any of the behaviors about which we enquired. This suggests that many students from low-income families are unaware of academic behaviors that may help them succeed in school, and that mentoring exposes students to these behaviors. Second, while treatment students may have felt under pressure to report engaging in behaviors that they were taught during the mentoring sessions even if they did not actually do so, this possibility is hard to reconcile with the fact that the overall distribution of students' self-reported behaviors shifted to the left between 2015 and 2016 (i.e., both control and treatment students reported engaging in fewer behaviors) (Figures B.1 and B.2).

Our results are similar if we account for attrition at each round of data collection. The ITT estimates remain virtually unchanged when we weigh each student's response by the inverse probability of his/her participation in this round of data collection (Table A.7). The Lee (2009) bounds are wide, but the treatment effects are consistent in sign, magnitude, and statistical significance with our main estimates (Table A.8).

We find heterogeneous effects by students' gender and socio-economic status. The program had a smaller effect on corrective test, failing, flunking, and absenteeism behavior for girls (by $.2$ to $.3\sigma$) (Table A.9). It also had a smaller effect on corrective homework, test, and failing behavior among students from low-income families (by $.31$ to $.44\sigma$) (Table A.10). We do not find heterogeneous effects for students who had previously repeated a grade or students who would have been admitted to the program through its regular admissions process.

Exposure to the program is associated with engagement in academic behaviors. Specifically, each month that a student received a scholarship and a mentoring session, he/she improved on average between $.02$ and $.04\sigma$ on all but two of the indexes of academic behavior (Table A.11). The indexes impacted and the magnitude of these effects are consistent with the ITT effects.

6.2 Academic mindsets, perseverance, and learning strategies

We find little evidence that the SMP improved academic mindsets, perseverance, and learning strategies. As Table 5 shows, we only find a marginally statistically significant positive effect on the questions of the LASSI measuring motivation (of $.12\sigma$). Yet, this difference is driven by a one-time effect in 2015, which does not emerge in 2014 or 2016 (Table A.12). We do not find any statistically significant effects in any of the other surveys or performance assessments that we administered in 2014-2016, but we cannot discard small to moderate effects.

[Insert Table 5 here.]

There is little indication that the program affected the distribution of academic mindsets, perseverance, and learning strategies. In nearly all metrics, treatment students performed on par with their control peers at all parts of the score distribution (Figures C.1 and C.3).

Accounting for attrition does not change this pattern of results. The inverse-probability weighted ITT estimates are very close to our unweighted estimates (Table A.13). The bounds on the ITT estimates confirm that we cannot discard small to moderate effects (Table A.14).

We find heterogeneous effects by socio-economic status. The program had a larger effect on grit, self-control, self-beliefs, and learning strategies among students from low-income families (by $.45$ to $.8\sigma$) (Table A.15). We do not, however, find evidence of heterogeneity by gender, prior repetition status, or eligibility under the regular admissions process.

The TOT effects aligned with the ITT effects in sign, magnitude, and statistical significance. Exposure to the program is associated with small and statistically insignificant effects, except for the questions of the LASSI measuring motivation. In that case, each month of exposure to the program is associated with a $.02\sigma$ effect (Table A.16).

6.3 School performance

We find almost no evidence that the SMP impacted students' performance in school when we estimate the ITT effects across all three years of the study. As Table 6 shows, we detect a marginally statistically significant reduction in the number of student absences from school (by 2.6 absences, over a control group mean of 33), but this reduction is no longer statistically significant once we include covariates in our estimation.

[Insert Table 6 here.]

Yet, these pooled effects mask important differences in the impact of the program across years. In 2014, the program improved students' grades in language (by .18 to $.2\sigma$) and reduced their absences to school (by 7.1 absences), the number of subjects that they failed (by .46 to .49, over a control group mean of 1.5), and their propensity to fail a grade (by 5 to 6 percentage points, over a control group mean of 15%). In 2015, the sign of most effects stayed the same, but their magnitude decreased and they were no longer statistically significant. In 2016, the sign of some effects changed. In fact, that year, we find that the program had a positive and marginally statistically significant effect in students' propensity to fail a grade (Table A.17). We do not find heterogeneous effects for any sub-groups.

We observe the same pattern of results once we account for attrition.²⁷ The inverse-probability weighted ITT estimates are nearly identical to the unweighted estimates (Table A.18). The bounds on the ITT effects are wide, but they rule out moderate to large effects in most outcomes in 2015 and 2016 (Table A.19). It is hard to argue that attrition from the study sample explains why the impacts from 2014 fade out by 2015 and 2016.

It seems more likely that this fadeout is due to the large share of students who were suspended and/or expelled from the program during the course of the impact evaluation (see Table A.3). It is possible that the program was improving students' performance in school, but that after nearly a third of treatment students were expelled, the average treatment student was no longer outperforming his/her control peer.

This explanation is consistent with the TOT effects. We find that each month receiving a scholarship and a mentoring session is associated with a statistically significant reduction in the number of failed subjects and absences, even when we pool our estimation across years (Table A.20). This pattern of results lends further support to the possibility that the effects found in 2014 may have been sustained if most students had remained in the program.

---

²⁷We have school performance data for all students in 2014, so Tables A.18 and A.18 exclude that year.

6.4 Student achievement

We find no evidence that the SMP impacted student achievement, as measured by standardized tests of math and reading. As Table 7 shows, the program had a negative but statistically insignificant effect in both subjects (by $.01$ to $.08\sigma$, depending on the subject). In fact, we find that the program had a negative and marginally statistically significant effect in math in 2016 (by $.16$ to $.17\sigma$), which is robust to the inclusion of covariates (Table A.21).

[Insert Table 7 here.]

The distribution of student achievement differs by subject and year. In 2015, there was a slightly larger share of treatment students (as compared to control students) performing above average in math, but not in reading. In 2016, we observe the opposite pattern (Figure D.1). These results are consistent with our discussion about the potential consequences of the large share of students who were expelled from the program during the study.

Accounting for attrition does not change the general pattern of results. The inverse-probability weighted ITT effects are similar to the unweighted estimates, but the negative effect in math is not statistically significant (Table A.22). The bounds allow us to discard small to moderate positive effects in math and moderate to large positive effects in reading (Table A.23).

We find evidence of heterogeneous effects by students' socio-economic status. The program had a larger (albeit marginally statistically significant) effect in both subjects among low-income students (by $.37$ to $.42\sigma$, depending on the subject) (Table A.24). We do not, however, find heterogeneous effects for other sub-groups.

The TOT effects are consistent with the ITT estimates in sign, magnitude, and statistical significance. Each month of program exposure is associated with a negative, but near-zero and statistically insignificant reduction in math and reading achievement (Table A.25).

6.5 Personality traits

Finally, we do not find that the SMP improved students' personality traits. In fact, as Table 8 shows, we find that the program had a negative and marginally statistically significant impact on conscientiousness (of $.22$ to $.25\sigma$) in 2016, the only year in which we measured these traits. This was partly due to a larger share of control students scoring high in this facet (Figure A.1).

[Insert Table 8 here.]

Our results remain virtually unchanged once we account for attrition. The inverse-probability weighted ITT estimates are nearly identical to the unweighted estimates, and the negative effect on conscientiousness remains statistically significant (Table A.26). In fact, the bounds on the ITT estimates rule out moderate to large positive effects on this facet (Table A.27).

We did not find heterogenous effects for any sub-groups. The TOT effects were consistent with the ITT effects in sign, magnitude, and statistical significance (Table A.28).

7 Discussion

This paper presents the results of a three-year randomized evaluation of a program that provided a monthly scholarship and a monthly mentoring session to secondary school students in the Province of Buenos Aires, Argentina. Our main motivation to evaluate this program was to understand whether cash transfers and scholarships have typically failed to improve student achievement (even if they have consistently increased educational attainment) because the students who typically benefit from these programs need additional support to develop the socio-emotional skills that they need to succeed in school. Yet, the lessons from our study are of considerable interest to Argentina, since the program that we evaluated is the longest-standing of its kind and it is run by the largest education non-profit in the country.

We find that the program positively impacted students’ academic behaviors (e.g., starting to study early before an exam or catching up on schoolwork missed due to absences). These effects were large and consistent across years. Yet, they did not translate into broader improvements on academic mindsets (e.g., self-beliefs about performance and efficacy), perseverance (e.g., grit), or learning strategies (e.g., metacognition), as measured by self-reports and performance assessments, or personality traits (although we cannot discard small to moderate effects).

There are at least three possible explanations for these seemingly contradictory results. One possibility is that the intervention made treatment students differentially susceptible to “social desirability bias” (i.e., they might have felt under pressure to report engaging in the behaviors that they were taught during the mentoring sessions even if they were not actually doing so) (see Bertrand and Mullainathan 2001; Mummolo and Peterson 2018; Zizzo 2010). This is plausible but hard to reconcile with the fact that both control and treatment students were less likely to report engaging in these behaviors in the third and final year of the study.

Another possibility is that academic mindsets, perseverance, and learning strategies are not malleable, or at least take a long time to change. This explanation seems better suited to explain the lack of effects on personality traits, which have been argued to be less malleable (Borghans et al. 2008). Yet, a number of recent field experiments in developing countries have

found that relatively brief and simple interventions have positively impacted mindsets and perseverance (see, for example, Alan et al. 2016; Outes et al. 2017).

A third possibility is that academic behaviors may not be related to mindsets, perseverance, and learning strategies as we had expected (i.e., students who adopt academic behaviors do not necessarily develop a more positive perception of themselves, persevere in the face of obstacles, or become more adept at studying). We cannot rigorously test this claim, but it is consistent with recent reviews of existing research (see, for example, Farrington et al. 2012), as well as with recent studies in developed countries, which find that students' (self-reported) mindsets and perseverance worsen even if their school performance improves (West et al. 2015).

We also find that the program improves students' performance in school in the first year of the evaluation, but that these effects fade out by the third year. There is no evidence suggesting that this fadeout is due to attrition from our study sample. It seems more likely that this fadeout is (at least in part) due to the large share of students who were suspended and/or expelled from the program. Each month of exposure to the program was associated with improvements in school performance, but by the end of our study, nearly a third of the treatment group had been expelled. Therefore, it is possible that we would have continued to observe effects on school performance if more students had remained in the program.

Lastly, we find that the program did not improve student achievement (as measured by standardized tests), even when it was positively impacting students' performance in school. This may seem surprising, but it is actually consistent with two sets of results that have been documented in other evaluations in developing countries. First, an intervention may yield improvements in one student assessment and not in another, even if both assessments focus on the same subject, due to differences in the assessments and in the incentives that students (and teachers) face to exert effort to do well in them (see, for example, Glewwe et al. 2010). Second, students who do well in school may not perform well in standardized tests designed for impact evaluations purposes and vice versa. School-based grades and/or assessments focus on whether students meet grade-based curricular expectations, whereas independent assessments try to minimize "floor" and "ceiling" effects, and thus capture a much wider range of content and cognitive domains at different difficulty levels (see Muralidharan et al. 2016).

Our results offer several lessons for research, policy, and practice. First, they suggest that the complementarities between scholarships and mentoring are not as important as we had expected or the type of mentoring that we considered was not enough to bring them to bear. Given what we know about the positive effects of cash transfers and scholarships by themselves, it is hard to argue that our study provides much reason to believe that combining scholarships with mentoring is a promising (let alone cost-effective) way of improving learning outcomes. It is, however, possible that the specific type of mentoring that we evaluated might have been

too infrequent and/or discretionary to meaningfully improve students' socio-emotional skills. Future studies in developing countries should focus on evaluating mentoring programs that provide more intensive support and/or are purposefully structured to impact specific skills. This type of program has produced promising results in developed countries (see, for example, Guryan et al. 2017; Heller et al. 2017; Kraft 2015).

Our results also draw attention to design features of mentoring programs in general (and of this program in particular) that may moderate the extent to which they improve learning. For example, our results seem to suggest that the program that we studied would benefit from reconsidering the type of student that it typically admits. We do not find that the program worked any better for those students who met the regular admissions criteria, which appears to favor students who are already doing relatively well in school. In fact, the program had a larger effect on low-income students, who are more likely to struggle in school. Similarly, our results indicate that imposing onerous conditions for students to receive support does not necessarily improve student effort and/or performance. Quite the opposite, the requirements for students to remain in the program that we evaluated ultimately resulted in nearly a third of them being expelled and hence losing access to potentially valuable support. Finally, although it is reasonable to grant autonomy to mentors with relatively high levels of education and experience (as compared to volunteer mentors), the ample variability in mentors' propensity to suspend and expel students suggests that it might be useful to establish some common guidelines for these sanctions to ensure that they are adjudicated fairly and consistently.

References

- Alan, S., T. Boneva, and S. Ertac (2016). Ever failed, try again, succeed better: Results from a randomized educational intervention on grit. *Unpublished manuscript*. Essex, England: University of Essex.
- Alfonso, M., M. S. Bos, J. Duarte, and C. Rondón (2011). Panorama general de la educación en América Latina y el Caribe. In Cabrol, M. and Székely, M. (Eds.), *Educación para la transformación*. Washington, DC: Banco Interamericano de Desarrollo (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.
- Baez, J. E. and A. Camacho (2011). Assessing the long-term effects of conditional cash transfers on human capital: Evidence from Colombia. (World Bank Policy Research Working Paper No. 5681). The World Bank. Washington, DC.
- Baird, S., C. McIntosh, and B. Özler (2011). Cash or condition? Evidence from a cash transfer experiment. *The Quarterly Journal of Economics* 126, 1709–1753.
- Banerjee, A., 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. Abdul Latif Jameel Latif Poverty Action Lab (J-PAL). Cambridge, MA.
- Barham, T., K. Macours, and J. A. Maluccio (2014). Assessing long-term impacts of conditional cash transfers on children and young adults in rural Nicaragua. (Impact Evaluation Report No. 17). International Initiative for Impact Evaluation (3ie). New Delhi, India.
- Barrera-Osorio, F., M. Bertrand, L. L. Linden, and F. Perez-Calle (2011). Improving the design of conditional transfer programs: Evidence from a randomized education experiment in Colombia. *American Economic Journal: Applied Economics*, 167–195.

- 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.
- Barrera-Osorio, F., L. L. Linden, and M. Urquiola (2007). The effects of user fee reductions on enrollment: Evidence from a quasi-experiment. *Unpublished manuscript*. Department of Economics at The University of Texas at Austin. Austin, TX.
- Behrman, J. R., S. W. Parker, and P. E. Todd (2009). Schooling impacts of conditional cash transfers on young children: Evidence from Mexico. *Economic Development and Cultural Change* 57(3), 439–477.
- Behrman, J. R., S. W. Parker, and P. E. Todd (2011). Do conditional cash transfers for schooling generate lasting benefits? A five-year followup of PROGRESA/Oportunidades. *Journal of Human Resources* 46(1), 93–122.
- Benhassine, N., F. Devoto, E. Duflo, P. Dupas, and V. Pouliquen (2013). Turning a shove into a nudge? A “labeled cash transfer” for education. (NBER Working Paper No. 19227). National Bureau of Economic Research (NBER). Cambridge, MA.
- Bertrand, M. and S. Mullainathan (2001). Do people mean what they say? implications for subjective survey data. *American Economic Review* 91(2), 67–72.
- Borghans, L., A. L. Duckworth, J. J. Heckman, and B. Ter Weel (2008). The economics and psychology of personality traits. *Journal of Human Resources* 43(4), 972–1059.
- Borkum, E. (2012). Can eliminating school fees in poor districts boost enrollment? Evidence from South Africa. *Economic Development and Cultural Change* 60(2), 359–398.
- Busso, M., M. Bassi, 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). Inter-American Development Bank. Washington, DC.
- Cayssials, A. N. (2003). *La escala de inteligencia WISC-III en la evaluación psicológica infanto-juvenil*. Buenos Aires, Argentina.
- Dammert, A. C. (2009). Heterogeneous impacts of conditional cash transfers: Evidence from Nicaragua. *Economic Development and Cultural Change* 58(1), 53–83.
- Del Carpio, X. V. and K. Macours (2010). Leveling the intra-household playing field: Compensation and specialization in child labor allocation. *Research in Labor Economics* 31, 259–295.

- Dhaliwal, I., E. Duffo, 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.
- DiNIECE (2013a). Anuario Estadístico 2013. Buenos Aires, Argentina: Dirección Nacional de Información de la Calidad Educativa (DiNIECE).
- DiNIECE (2013b). 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).
- Dizon-Ross, R. (2015). Parents’ beliefs and children’s education: Experimental evidence from Malawi. *Unpublished manuscript*. Chicago, IL: University of Chicago Booth School of Business.
- 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.
- Duckworth, A. L. and D. S. Yeager (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher* 44(4), 237–251.
- Evans, D., M. Kremer, and M. Ngatia (2009). The impact of distributing school uniforms on children’s education in Kenya. Unpublished manuscript. The World Bank. Washington, DC.
- 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 University of Chicago Consortium on Chicago School Research (CCSR). Chicago, IL.
- Fernandez Liporace, M. M. and M. M. Casullo (2009). Factores salugénicos, ajuste psicológico y rendimiento académico en estudiantes de nivel medio y universitarios. Unpublished manuscript. Universidad de Buenos Aires (UBA). Buenos Aires, Argentina.
- Filmer, D. and N. Schady (2014). The medium-term effects of scholarships in a low-income country. *Journal of Human Resources* 49(3), 663–694.
- 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.

- Gabrieli, C., D. Ansel, and S. Bartolino Krachman (2015). Ready to be counted: The research case for education policy action on non-cognitive skills. Unpublished manuscript. Transforming Education. Boston, MA.
- Galiani, S. and P. J. McEwan (2013). The heterogeneous impact of conditional cash transfers. *Journal of Public Economics* 103, 85–96.
- Ganimian, A. J. and R. J. Murnane (2016). Improving educational outcomes in developing countries: Lessons from rigorous impact evaluations. *Review of Educational Research* 86(3), 719–755.
- Glewwe, P., N. Ilias, and M. Kremer (2010). Teacher incentives. *American Economic Journal: Applied Economics*, 205–227.
- Glewwe, P., M. Kremer, and S. Moulin (2009). Many children left behind? Textbooks and test scores in Kenya. *American Economic Journal: Applied Economics* 1(1), 112–135.
- 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.
- 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.
- Jensen, R. (2010). The (perceived) returns to education and the demand for schooling. *The Quarterly Journal of Economics* 125(2), 515–548.
- Jensen, R. (2012). Do labor market opportunities affect young women’s work and family decisions? Experimental evidence from India. *The Quarterly Journal of Economics* 127(2), 753–792.
- 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.
- Karlan, D. and L. L. Linden (2014). Loose knots: Strong versus weak commitments to save for education in uganda. (NBER Working Paper No. 19863). National Bureau of Economic Research (NBER). Cambridge, MA.
- Kling, J. R., J. B. Liebman, and L. F. Katz (2005). Experimental analysis of neighborhood effects. *Econometrica* 75(1), 83–119.
- 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. Thornton (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.
- Liu, C., H. Yi, L. Zhang, R. Luo, Y. Shi, J. Chu, and S. Rozelle (2012). The effect of early commitment of financial aid on matriculation to senior high school among poor junior high students in rural China. (REAP Working Paper No. 254). Rural Education Action Program (REAP). Stanford, CA.
- Loyalka, P., C. Liu, Y. Song, H. Yi, X. Huang, J. Wei, L. Zhang, Y. Shi, J. Chu, and S. Rozelle (2013). Can information and counseling help students from poor rural areas go to high school? Evidence from China. *Journal of Comparative Economics* 41(4), 1012–1025.
- Lucas, A. M. and I. M. Mbiti (2012). Access, sorting, and achievement: the short-run effects of free primary education in Kenya. *American Economic Journal: Applied Economics* 4, 226–253.
- Maluccio, J. and R. Flores (2005). Impact evaluation of a conditional cash transfer program: The Nicaraguan Red de Protección Social. Unpublished manuscript. International Food Policy Research Institute.
- Martos Mula, A., O. Saavedra, N. Wierna, M. Ruggeri, J. Tschambler, N. Ávila Carreras, M. Bonillo, and M. Bovi Mitre (2013). Afectación de las funciones cognitivas y motoras en

- niños residentes de zonas rurales de Jujuy y su relación con plaguicidas inhibidores de la colinesterasa. Un estudio piloto. *Acta toxicológica Argentina* 21, 15–24.
- Mummolo, J. and E. Peterson (2018). Demand effects in survey experiments: An empirical assessment. *Unpublished manuscript*. Princeton University: Princeton, NJ.
- Muralidharan, K. and N. Prakash (2013). Cycling to school: Increasing secondary school enrollment for girls in India. (Working Paper No. 19305). National Bureau of Economic Research (NBER). Cambridge, MA.
- Muralidharan, K., A. Singh, and A. J. Ganimian (2016). Disrupting education? Experimental evidence on technology-aided instruction in India. (NBER Working Paper No. 22923). National Bureau of Economic Research (NBER). Cambridge, MA.
- Muralidharan, K. and Y. Zieleniak (2014). Chasing the syllabus: Measuring learning trajectories in developing countries with longitudinal data and item response theory. *Unpublished manuscript*. University of California, San Diego. San Diego, CA.
- OECD (2014). *Education at a Glance 2014: OECD Indicators*. Paris, France: Organisation for Economic Co-operation and Deveopment.
- Outes, I., A. Sánchez, and R. Vakis (2017). Growth mindset at scale: Impact of a psychosocial intervention on secondary school attainment in Peru. *Unpublished manuscript*. Oxford, England: Blavatnik School of Government, Oxford University.
- Pais, E. F. (2015). Informe de evaluación de impacto en habilidades socioemocionales del programa extra-clase, Año 2014, Argentina y México. Unpublished manuscript. Techint. Buenos Aires, Argentina.
- 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.
- Perova, E. and R. Vakis (2012). 5 years in Juntos: New evidence on the program’s short and long-term impacts. *Economía* 35(69), 53–82.
- Pritchett, L. and A. Beatty (2015). Slow down, you’re going too fast: Matching curricula to student skill levels. *International Journal of Educational Development* 40, 276–288.

- 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).
- Schultz, T. P. (2004). School subsidies for the poor: Evaluating the Mexican Progresa poverty program. *Journal of development Economics* 74(1), 199–250.
- SEE-MEDN (2017). Aprender 2016: Análisis de desempeños por capacidades y contenidos. Nivel primario. (Serie de documentos técnicos, Nro. 7.) Ciudad Autónoma de Buenos Aires: Secretaría de Evaluación Educativa. Ministerio de Educación y Deportes de la Nación.
- Skoufias, E., S. W. Parker, J. R. Behrman, and C. Pessino (2001). Conditional cash transfers and their impact on child work and schooling: Evidence from the Progresa program in Mexico. *Economia*, 45–96.
- Soprano, A. (2003). Evaluación de las funciones ejecutivas en el niño. *Revista Neurología* 37(44-50).
- 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).
- Wechsler, D. (1994). *Test de inteligencia para niños WISC-III: manual*. Buenos Aires, Argentina: Paidós.
- Weinstein, C. E. and D. R. Palmer (1988). *LASSI: The learning and study strategies inventory*. Miami, FL: Publishing Company.
- West, M. R., M. A. Kraft, A. S. Finn, R. Martin, A. L. Duckworth, C. F. Gabrieli, J. D. Gabrieli, et al. (2015). Promise and paradox: Measuring students’ non-cognitive skills and the impact of schooling. *Educational Evaluation and Policy Analysis* 38(1), 148–170.
- 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.
- Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics* 13(1), 75–98.

Table 1: SMP’s theory of change

Problem	Inputs/Activities	Outputs	Direct outcomes	Indirect outcomes	Plausible outcomes
Low-income families perceive the costs of schooling to be too high, its benefits to be too low, and/or they lack access to credit	Scholarship: USD 414 per year, disbursed over 10 months to students’ parents or legal guardians	Parents use the scholarships to offset the costs of their children’s schooling	- Lower student absenteeism - Lower dropout rates		
Parents in low-income families have little experience with schooling, which limits their capacity to support their children	Mentoring: 10 mentoring sessions before/after school per year, conducted monthly individually or in groups, led by EAs (parents attend some sessions)	- Students attend mentoring sessions regularly - Students bring the required materials to the mentoring sessions - Parents attend mentoring sessions when invited	Improved academic behaviors (e.g., going to class, homework, organizing materials, participating, studying)	- Improved academic mindsets (e.g., self-beliefs about performance and efficacy) - Improved academic perseverance (e.g., grit) - Improved learning strategies (e.g., metacognition)	- Improved school performance (e.g., higher grades, lower grade failure) - Improved student achievement (e.g., math and reading test scores) - Improved personality traits (e.g., conscientiousness)
Assumptions:	- The amount of the scholarship is enough to cover school costs - Scholarship does not lead to substitution effects in parents’ investments - External factors do not prevent students or parents from attending sessions - Costs, benefits, and/or access to credit are binding constraints to student attendance and permanence in school - Students understand and are able to implement mentors’ recommendations - Academic mindsets and perseverance are malleable over the short- or medium-term - Improved academic behaviors influence development of academic mindsets and perseverance - Academic behaviors are binding constraint to school performance - Better school performance leads to better student achievement				

Table 2: Data collection timeline

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

Notes: (1) We collected SMP data directly from the non-profit, which is why we have not specified the location of data collection. (2) SMP data are only available for treatment students (because control students did not participate in the program).

Table 3: Balancing checks (baseline)

Variable	All (1)	Control (2)	Treatment (3)	Difference (4)	N (5)
<i>Panel A. Student survey</i>					
Argentine	.977 (.149)	.98 (.141)	.975 (.157)	-.005 (.02)	397
Female	.52 (.5)	.544 (.499)	.495 (.501)	-.049 (.051)	408
Age	12.435 (1.062)	12.502 (1.153)	12.368 (.961)	-.131 (.11)	407
Attends morning shift	.578 (.494)	.583 (.494)	.574 (.496)	-.008 (.045)	408
Previously repeated grade(s)	.309 (.463)	.322 (.468)	.297 (.458)	-.024 (.044)	404
Previously dropped out of school	.05 (.218)	.073 (.262)	.027 (.163)	-.047* (.022)	360
<i>Panel B. Household survey</i>					
Has car	.208 (.407)	.162 (.369)	.255 (.437)	.095*** (.025)	408
Has natural gas	.294 (.456)	.265 (.442)	.324 (.469)	.06* (.032)	408
Has running water	.811 (.392)	.789 (.409)	.833 (.374)	.053 (.05)	408
Has in-house bathroom	.824 (.382)	.809 (.394)	.838 (.369)	.03 (.045)	408
Has solid floor	.985 (.121)	.98 (.139)	.99 (.099)	.01 (.006)	408
Has fridge	.713 (.453)	.667 (.473)	.76 (.428)	.091*** (.027)	408
Has computer	.539 (.499)	.539 (.5)	.539 (.5)	.002 (.027)	408
Has Internet	.382 (.487)	.377 (.486)	.387 (.488)	.012 (.033)	408
Has cell phone	.904 (.294)	.877 (.329)	.931 (.253)	.054* (.029)	408
Parent is homeowner	.598 (.491)	.564 (.497)	.632 (.483)	.07* (.033)	408
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) and shows the number of non-missing observations (column 5). (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 4: ITT effects on academic behaviors (2015-2016)

	Control (1)	Effect size	
	(2)	(3)	
Proactive school behavior (std.)	0 (1) 180	.158 (.091) 707	.133 (.087) 707
Preventive homework behavior (std.)	0 (1) 180	.222** (.088) 707	.206** (.076) 707
Corrective homework behavior (std.)	0 (1) 180	.223** (.078) 707	.213** (.078) 707
Preventive test behavior (std.)	0 (1) 180	.193*** (.04) 707	.186*** (.039) 707
Corrective test behavior (std.)	0 (1) 180	.204*** (.054) 707	.188*** (.056) 707
Corrective failing behavior (std.)	0 (1) 180	.299*** (.051) 707	.279*** (.052) 707
Corrective flunking behavior (std.)	0 (1) 180	.154** (.057) 707	.143** (.056) 707
Preventive absenteeism behavior (std.)	0 (1) 180	.251*** (.063) 707	.254*** (.066) 707
Corrective absenteeism behavior (std.)	0 (1) 180	.279*** (.063) 707	.273*** (.064) 707
Corrective free period behavior (std.)	0 (1) 180	.021 (.093) 707	.041 (.099) 707
Family index (std.)	0 (1) 180	.307*** (.064) 707	.292*** (.063) 707
χ^2		3.055	2.378
p-value		.08	.123
School FE?	Y	Y	Y
Year FE?	Y	Y	Y
Controls?		N	Y

Notes: (1) The table shows the mean and standard deviations of control group students in 2015 (column 1) and the average ITT effect without (column 2) and with covariates (column 3), pooled across 2015-2016. (2) * significant at 10%; ** significant at 5%; *** significant at 1%. (3) Standard errors in columns 2-3 are clustered at the school level. (4) The family index is a standardized average of all the individual outcomes in this table. (5) The table also shows the χ^2 and associated p-value from a joint test using seemingly unrelated regressions. This test only includes the individual outcomes, not the family index.

Table 5: ITT effects on academic mindsets, perseverance, and learning strategies (2014-2016)

	Control (1)	Effect size (2) (3)	
Grit (std.)	0 (1) 193	.052 (.053) 1102	.062 (.056) 1102
DSISC (std.)	0 (1) 193	.093 (.073) 1102	.1 (.073) 1102
Self-beliefs (std.)	0 (1) 193	.073 (.089) 1102	.073 (.084) 1102
LASSI - Organization and Planning (std.)	0 (1) 193	.028 (.083) 1102	.041 (.087) 1102
LASSI - Motivation (std.)	0 (1) 193	.124* (.065) 1102	.122* (.065) 1102
CARAS - Reflexivity Index (std.)	0 (1) 193	.002 (.074) 1094	.025 (.075) 1094
LABS (std.)	0 (1) 193	-.054 (.044) 1102	-.066 (.043) 1102
Family index (std.)	0 (1) 193	.083 (.082) 1102	.093 (.083) 1102
χ^2		.967	1.227
p-value		.325	.268
School FE?	Y	Y	Y
Year FE?	Y	Y	Y
Controls?		N	Y

Notes: (1) The table shows the mean and standard deviations of control group students in 2014 (column 1) and the average ITT effect without (column 2) and with covariates (column 3), pooled across 2014-2016. (2) * significant at 10%; ** significant at 5%; *** significant at 1%. (3) Standard errors in columns 2-3 are clustered at the school level. (4) The family index is a standardized average of all the individual outcomes in this table. (5) The table also shows the χ^2 and associated p-value from a joint test using seemingly unrelated regressions. This test only includes the individual outcomes, not the family index.

Table 6: ITT effects on school performance (2014-2016)

	Control (1)	Effect size (2) (3)	
Language grade (std.)	0 (1) 203	.028 (.086) 1012	.003 (.086) 1012
Math grade (std.)	0 (1) 203	.009 (.054) 1011	-.007 (.056) 1011
No. of pending subjects	1.516 (2.511) 190	-.332 (.182) 980	-.301 (.188) 980
No. of absences	33.4 (28.056) 173	-2.643* (1.427) 944	-1.964 (1.561) 944
Failed grade	.148 (.356) 203	-.015 (.018) 1057	-.009 (.02) 1057
Transferred schools	.025 (.155) 203	-.019 (.013) 1057	-.017 (.012) 1057
Dropped out of school	.054 (.227) 203	-.025 (.014) 1057	-.024 (.014) 1057
Family index (std.)	0 (1) 204	.047 (.065) 1064	.025 (.068) 1064
χ^2		.103	.001
p-value		.748	.976
School FE?	Y	Y	Y
Year FE?	Y	Y	Y
Controls?		N	Y

Notes: (1) The table shows the mean and standard deviations of control group students in 2014 (column 1) and the average ITT effect without (column 2) and with covariates (column 3), pooled across 2014-2015. (2) * significant at 10%; ** significant at 5%; *** significant at 1%. (3) Standard errors in columns 2-3 are clustered at the school level. (4) The family index is a standardized average of all the individual outcomes in this table. (5) The table also shows the χ^2 and associated p-value from a joint test using seemingly unrelated regressions. This test only includes the individual outcomes, not the family index.

Table 7: ITT effects on student achievement (2015-2016)

	Control (1)	Effect size (2) (3)	
Math achievement (std.)	0 (1) 164	-.055 (.093) 683	-.078 (.091) 683
Reading achievement (std.)	0 (1) 173	-.008 (.066) 706	-.036 (.066) 706
Family index (std.)	0 (1) 174	-.018 (.07) 714	-.049 (.069) 714
χ^2		.357	.749
p-value		.55	.387
School FE?	Y	Y	Y
Year FE?	Y	Y	Y
Controls?		N	Y

Notes: (1) The table shows the mean and standard deviations of control group students in 2014 (column 1) and the average ITT effect without (column 2) and with covariates (column 3), pooled across 2014-2015. (2) * significant at 10%; ** significant at 5%; *** significant at 1%. (3) Standard errors in columns 2-3 are clustered at the school level. (4) Test scores in math and reading are scaled using Item Response Theory and standardized with respect to the control group at baseline. (5) The bottom rows display the result from a joint test using Seemingly Unrelated Regressions (SURs).

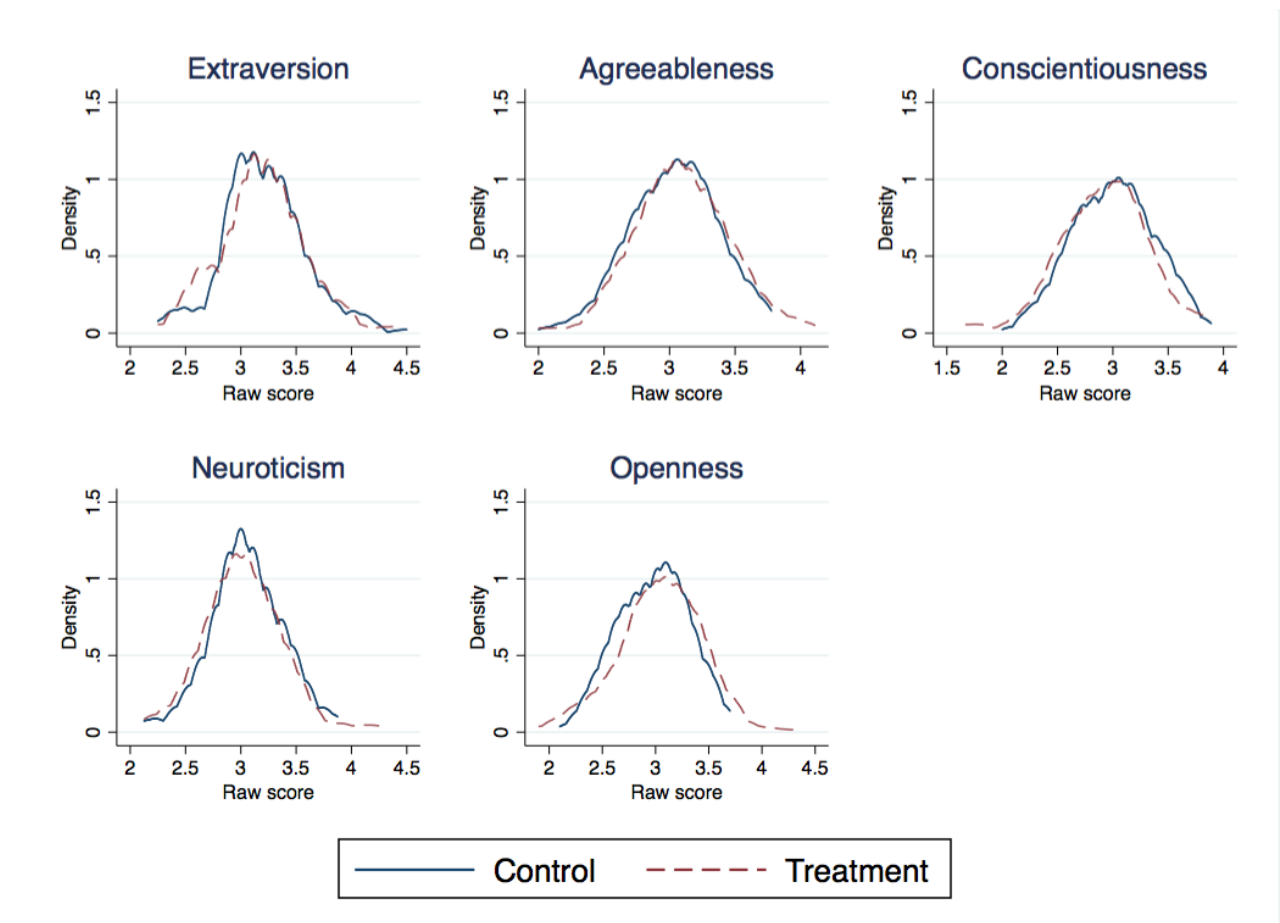
Table 8: ITT effects on personality traits (2016)

Index	Control (1)	Effect size	
		(2)	(3)
Extraversion (std.)	0 (1) 165	-.055 (.09) 341	-.042 (.096) 341
Agreeableness (std.)	0 (1) 165	.182 (.116) 341	.155 (.118) 341
Conscientiousness (std.)	0 (1) 165	-.245* (.111) 341	-.221* (.114) 341
Neuroticism (std.)	0 (1) 165	-.104 (.087) 341	-.099 (.087) 341
Openness (std.)	0 (1) 165	.186 (.106) 341	.159 (.106) 341
Family index (std.)	0 (1) 165	-.017 (.097) 341	-.022 (.104) 341
χ^2		.392	.196
p-value		.531	.658
School FE?		Y	Y
Controls?		N	Y

Notes: (1) The table shows the mean and standard deviations of control group students (column 1) and the average ITT effect without (column 2) and with covariates (column 3). (2) * significant at 10%; ** significant at 5%; *** significant at 1%. (3) Standard errors in columns 2 and 3 are clustered at the school level. (4) The bottom rows display the result from a joint test using Seemingly Unrelated Regressions (SURs).

Appendix A Additional graphs and tables

Figure A.1: Distributions of survey of students' raw scores on Big Five Inventory (2016)



Note: This figure shows the distribution of raw scores of each facet in the Big Five Inventory, separately for the control and treatment groups, in 2015.

Table A.1: Program costs per year (2014)

Budget line	Cost per year		Cost per student		Share of total
	ARS	USD	ARS	USD	
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.2: Comparison of RCT and non-RCT schools on administrative data

Variable	Non-RCT schools					N (7)
	All PBA schools (1)	Group 1: Urban schools (2)	Group 2: Urban + public schools (3)	RCT schools (4)	RCT-Group 1 schools (5)	RCT-Group 2 schools (6)
Enrolled students	313.993 (256.547)	332.795 (257.938)	345.31 (291.128)	691.7 (371.646)	359.74*** (81.487)	347.651*** (92.009)
Passing rate	78.515 (14.143)	78.188 (14.289)	71.922 (12.913)	66.857 (6.188)	-11.355** (4.765)	-5.082 (4.311)
Dropout rate	2.272 (4.271)	2.173 (4.198)	3.204 (4.859)	2.823 (4.802)	.652 (1.401)	-.382 (1.623)
Repetition rate	11.082 (9.631)	11.233 (9.687)	15.351 (9.419)	9.505 (5.173)	-1.732 (3.067)	-5.868** (2.983)
Overage rate	17.03 (13.266)	17.26 (13.533)	22.968 (12.093)	18.411 (8.506)	1.154 (4.285)	-4.574 (3.831)

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) and shows the number of non-missing observations (column 5). (2) * significant at 10%; ** significant at 5%; *** significant at 1%. (3) Standard errors in column 4 are clustered at the school level.

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

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

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

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

Baseline									
Control	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 (.141)	-.005 (.02)	0 (.017)	.001 (.017)	.013 (.017)	0 (.017)	.001 (.018)	.002 (.014)	
Female	.544 (.499)	-.049 (.051)	-.068 (.057)	-.094 (.054)	-.059 (.06)	-.072 (.058)	-.074 (.056)	-.049 (.08)	
Age	12.502 (1.153)	-.131 (.11)	-.096 (.12)	-.09 (.116)	-.008 (.128)	-.098 (.115)	-.061 (.107)	.058 (.092)	
Attends morning shift	.583 (.494)	-.008 (.045)	-.014 (.049)	-.02 (.052)	-.047 (.056)	-.024 (.05)	.003 (.059)	-.027 (.061)	
Previously repeated grade(s)	.322 (.468)	-.024 (.044)	-.038 (.044)	-.032 (.042)	.007 (.048)	-.028 (.042)	-.032 (.046)	-.013 (.039)	
Previously dropped out of school	.073 (.262)	-.047* (.022)	-.042* (.021)	-.032* (.015)	-.002 (.026)	-.037* (.02)	-.04** (.014)	.004 (.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, by year (2015-2016)

	2015			2016		
	Control (1)	Effect size (2)	Effect size (3)	Control (4)	Effect size (5)	Effect size (6)
Proactive school behavior (std.)	0 (1) 180	.115 (.161) 366	.1 (.154) 366	-.217 (.816) 165	.208* (.1) 341	.173 (.099) 341
Preventive homework behavior (std.)	0 (1) 180	.235* (.125) 366	.234* (.115) 366	-.468 (.867) 165	.216** (.087) 341	.183* (.084) 341
Corrective homework behavior (std.)	0 (1) 180	.205** (.081) 366	.209** (.08) 366	-.347 (.81) 165	.253** (.102) 341	.228* (.105) 341
Preventive test behavior (std.)	0 (1) 180	.209** (.07) 366	.214*** (.065) 366	-.413 (.768) 165	.186*** (.048) 341	.165** (.054) 341
Corrective test behavior (std.)	0 (1) 180	.216** (.081) 366	.205** (.083) 366	-.37 (.749) 165	.198** (.068) 341	.176** (.068) 341
Corrective failing behavior (std.)	0 (1) 180	.263** (.084) 366	.246** (.082) 366	-.292 (.695) 165	.348*** (.046) 341	.324*** (.042) 341
Corrective flunking behavior (std.)	0 (1) 180	.12 (.087) 366	.119 (.083) 366	-.286 (.729) 165	.2*** (.047) 341	.18*** (.048) 341
Preventive absenteeism behavior (std.)	0 (1) 180	.176* (.096) 366	.183* (.095) 366	-.418 (.862) 165	.339*** (.061) 341	.34*** (.073) 341
Corrective absenteeism behavior (std.)	0 (1) 180	.255** (.089) 366	.257** (.086) 366	-.342 (.74) 165	.313*** (.062) 341	.299*** (.064) 341
Corrective free period behavior (std.)	0 (1) 180	.039 (.11) 366	.051 (.116) 366	-.277 (.889) 165	.005 (.121) 341	.034 (.123) 341
School FE?		Y	Y		Y	Y
Controls?		N	Y		N	Y

Notes: (1) The table shows the mean and standard deviations of control group students (columns 1 and 4) and the average ITT effect without (columns 2 and 5) and with covariates (columns 3 and 6). (2) * significant at 10%; ** significant at 5%; *** significant at 1%. (3) Standard errors in columns 2-3 and 5-6 are clustered at the school level.

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

	2015		2016	
	Effect size		Effect size	
	(1)	(2)	(3)	(4)
Proactive school behavior (std.)	.131 (.165)	.117 (.159)	.211* (.101)	.171 (.101)
	366	366	341	341
Preventive homework behavior (std.)	.251* (.126)	.251* (.115)	.222** (.084)	.184** (.081)
	366	366	341	341
Corrective homework behavior (std.)	.209** (.084)	.214** (.083)	.249** (.107)	.22* (.113)
	366	366	341	341
Preventive test behavior (std.)	.217** (.068)	.223*** (.064)	.188*** (.05)	.164** (.057)
	366	366	341	341
Corrective test behavior (std.)	.224** (.078)	.214** (.081)	.203** (.072)	.177** (.073)
	366	366	341	341
Corrective failing behavior (std.)	.269*** (.082)	.251** (.08)	.354*** (.05)	.326*** (.047)
	366	366	341	341
Corrective flunking behavior (std.)	.132 (.084)	.13 (.081)	.208*** (.049)	.185*** (.05)
	366	366	341	341
Preventive absenteeism behavior (std.)	.187* (.1)	.194* (.099)	.35*** (.058)	.347*** (.071)
	366	366	341	341
Corrective absenteeism behavior (std.)	.266** (.089)	.267** (.087)	.322*** (.058)	.305*** (.058)
	366	366	341	341
Corrective free period behavior (std.)	.039 (.105)	.051 (.111)	.02 (.12)	.051 (.122)
	366	366	341	341
School FE?	Y	Y	Y	Y
Controls?	N	Y	N	Y

Notes: (1) The table shows the mean and standard deviations of control group students (columns 1 and 4) and the average ITT effect without (columns 2 and 5) and with covariates (columns 3 and 6). (2) * significant at 10%; ** significant at 5%; *** significant at 1%. (3) Standard errors in columns 2-3 and 5-6 are clustered at the school level.

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

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 (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.9: ITT effects on academic behaviors by gender (2015-2016)

	Constant (1)	Treatment (2)	Female (3)	Interaction (4)	N (5)
Proactive school behavior (std.)	-.26** (.1)	.165 (.121)	.025 (.091)	-.009 (.073)	707
Preventive homework behavior (std.)	-.382*** (.106)	.216 (.124)	.03 (.105)	.016 (.138)	707
Corrective homework behavior (std.)	-.106 (.093)	.249** (.101)	-.045 (.103)	-.059 (.092)	707
Preventive test behavior (std.)	-.32*** (.081)	.232** (.078)	.042 (.103)	-.07 (.123)	707
Corrective test behavior (std.)	-.156* (.082)	.315*** (.058)	.062 (.113)	-.209*** (.063)	707
Corrective failing behavior (std.)	-.043 (.101)	.409*** (.08)	.07 (.154)	-.204** (.081)	707
Corrective flunking behavior (std.)	-.033 (.099)	.295*** (.075)	.057 (.132)	-.269** (.088)	707
Preventive absenteeism behavior (std.)	-.455*** (.089)	.408*** (.108)	.268** (.085)	-.266 (.149)	707
Corrective absenteeism behavior (std.)	-.369*** (.07)	.45*** (.057)	.207** (.091)	-.304*** (.087)	707
Corrective free period behavior (std.)	-.125 (.107)	.109 (.092)	-.099 (.173)	-.191 (.202)	707

Notes: (1) The table examines whether the program differentially impacted students by gender. Column 4 shows the interaction between the treatment and female dummies for each variable at baseline. (2) * significant at 10%; ** significant at 5%; *** significant at 1%. (3) Standard errors in columns 2-3 are clustered at the school level. (4) All estimations include school and fixed effects, but no controls.

Table A.10: ITT effects on academic behaviors by income (2015-2016)

	Constant (1)	Treatment (2)	Low SES (3)	Interaction (4)	N (5)
Proactive school behavior (std.)	-.219*** (.056)	.195* (.09)	-.096 (.123)	-.2 (.135)	707
Preventive homework behavior (std.)	-.306*** (.059)	.201 (.116)	-.222 (.123)	.024 (.202)	707
Corrective homework behavior (std.)	-.122 (.083)	.29** (.104)	-.029 (.126)	-.31* (.154)	707
Preventive test behavior (std.)	-.297*** (.064)	.236*** (.072)	.027 (.111)	-.186 (.192)	707
Corrective test behavior (std.)	-.097* (.046)	.27*** (.055)	-.032 (.094)	-.305* (.161)	707
Corrective failing behavior (std.)	.009 (.062)	.398*** (.054)	.014 (.105)	-.44** (.157)	707
Corrective flunking behavior (std.)	.029 (.069)	.198** (.08)	-.044 (.111)	-.214 (.17)	707
Preventive absenteeism behavior (std.)	-.31*** (.038)	.282** (.09)	.123 (.105)	-.1 (.216)	707
Corrective absenteeism behavior (std.)	-.237*** (.043)	.327*** (.093)	.043 (.104)	-.205 (.203)	707
Corrective free period behavior (std.)	-.236** (.074)	.055 (.127)	.232* (.11)	-.077 (.204)	707

Notes: (1) The table examines whether the program differentially impacted students by income. Column 4 shows the interaction between the treatment and female dummies for each variable at baseline. (2) * significant at 10%; ** significant at 5%; *** significant at 1%. (3) Standard errors in columns 2-3 are clustered at the school level. (4) All estimations include school and fixed effects, but no controls.

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

<i>Panel A. First stage</i>						
	Months/year with scholarship		Months/year with mentoring		Months/year with both	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	6.786*** (.386) 707	6.759*** (.381) 707	6.577*** (.348) 707	6.542*** (.345) 707	6.321*** (.363) 707	6.286*** (.357) 707
<i>Panel B. Second stage</i>						
	Scholarships		Mentoring		Combined	
	(1)	(2)	(3)	(4)	(5)	(6)
Proactive school behavior (std.)	.023* (.012) 707	.02* (.012) 707	.024* (.013) 707	.02* (.012) 707	.025* (.013) 707	.021 (.013) 707
Preventive homework behavior (std.)	.033*** (.012) 707	.03*** (.01) 707	.034*** (.012) 707	.031*** (.011) 707	.035*** (.013) 707	.033*** (.011) 707
Corrective homework behavior (std.)	.033*** (.01) 707	.032*** (.01) 707	.034*** (.01) 707	.033*** (.01) 707	.035*** (.011) 707	.034*** (.011) 707
Preventive test behavior (std.)	.028*** (.006) 707	.028*** (.005) 707	.029*** (.006) 707	.028*** (.005) 707	.031*** (.006) 707	.03*** (.006) 707
Corrective test behavior (std.)	.03*** (.008) 707	.028*** (.008) 707	.031*** (.008) 707	.029*** (.008) 707	.032*** (.008) 707	.03*** (.008) 707
Corrective failing behavior (std.)	.044*** (.007) 707	.041*** (.007) 707	.045*** (.007) 707	.043*** (.007) 707	.047*** (.008) 707	.044*** (.007) 707
Corrective flunking behavior (std.)	.023*** (.008) 707	.021*** (.008) 707	.023*** (.008) 707	.022*** (.008) 707	.024*** (.009) 707	.023*** (.008) 707
Preventive absenteeism behavior (std.)	.037*** (.008) 707	.038*** (.009) 707	.038*** (.009) 707	.039*** (.009) 707	.04*** (.009) 707	.04*** (.01) 707
Corrective absenteeism behavior (std.)	.041*** (.008) 707	.04*** (.008) 707	.042*** (.008) 707	.042*** (.008) 707	.044*** (.008) 707	.043*** (.008) 707
Corrective free period behavior (std.)	.003 (.013) 707	.006 (.014) 707	.003 (.013) 707	.006 (.014) 707	.003 (.014) 707	.006 (.015) 707
School FE?	Y	Y	Y	Y	Y	Y
Year FE?	Y	Y	Y	Y	Y	Y
Controls?	N	Y	N	Y	N	Y

Notes: (1) Panel A shows the results from the first stage regression of the number of months per year with a scholarship (columns 1-2), a mentoring session (column 3-4), or both (column 5-6), with and without covariates, pooled across 2015-2016. (2) Panel B shows the dose-response relationship of scholarships (columns 1-2) and mentoring (columns 3-4) and the TOT effect of both (columns 5-6), without and with covariates. (3) * significant at 10%; ** significant at 5%; *** significant at 1%. (4) Standard errors are clustered at the school level.

Table A.12: ITT effects on academic mindsets, perseverance, and learning strategies, by year (2014-2016)

	2014			2015			2016		
	Control (1)	Effect size (2)	Effect size (3)	Control (4)	Effect size (5)	Effect size (6)	Control (7)	Effect size (8)	Effect size (9)
Grit (std.)	0 (1)	.079 (.073)	.088 (.075)	0 (1)	.115 (.088)	.118 (.091)	0 (1)	-.053 (.093)	-.033 (.096)
DSISC (std.)	193 0	395 .099	395 .118	180 0	366 .133	366 .137	165 0	341 .04	341 .035
Self-beliefs (std.)	193 0	.099 (.099)	.101 (.101)	180 0	.076 (.076)	.077 (.077)	165 0	.082 (.082)	.081 (.081)
	193 0	395 .042	395 .039	180 0	366 .114	366 .117	165 0	341 .061	341 .06
	193 0	.102 (.102)	.099 (.099)	180 0	.085 (.085)	.082 (.082)	165 0	.104 (.104)	.102 (.102)
LASSI - Organization and Planning (std.)	193 0	395 .027	395 .046	180 0	366 .03	366 .046	165 0	341 .024	341 .027
	193 0	.101 (.101)	.103 (.103)	180 0	.079 (.079)	.084 (.084)	165 0	.112 (.112)	.117 (.117)
LASSI - Motivation (std.)	193 0	395 .153	395 .138	180 0	366 .173**	366 .18**	165 0	341 .034	341 .039
	193 0	.128 (.128)	.131 (.131)	180 0	.074 (.074)	.078 (.078)	165 0	.056 (.056)	.045 (.045)
CARAS - Reflexivity Index (std.)	193 0	395 -.009	395 .005	180 0	366 .044	366 .088	165 0	341 -.036	341 -.023
	193 0	.082 (.082)	.075 (.075)	180 0	.082 (.082)	.084 (.084)	165 0	.1 (.1)	.1 (.1)
LABS (std.)	193 0	394 -.014	394 -.021	176 0	360 -.114	360 -.129	164 0	340 -.024	340 -.037
	193 0	.066 (.066)	.069 (.069)	180 0	.109 (.109)	.115 (.115)	165 0	.055 (.055)	.056 (.056)
School FE?	193 Y	395 Y	395 Y	180 Y	366 Y	366 Y	165 Y	341 Y	341 Y
Controls?		N	Y		N	Y		N	Y

Notes: (1) The table shows the mean and standard deviations of control group students (columns 1 and 4) and the average ITT effect without (columns 2, 5, and 8) and with covariates (columns 3, 6, and 9). (2) * significant at 10%; ** significant at 5%; *** significant at 1%. (3) Standard errors in columns 2-3, 5-6, and 8-9 are clustered at the school level.

Table A.13: ITT effects on academic mindsets, perseverance, and learning strategies with inverse probability weighting, by year (2014-2016)

	2014		2015		2016	
	Effect size (1)	Effect size (2)	Effect size (3)	Effect size (4)	Effect size (5)	Effect size (6)
Grit (std.)	-.002 (.105)	.006 (.107)	.128 (.096)	.131 (.1)	-.044 (.093)	-.019 (.097)
	247	247	366	366	341	341
DSISC (std.)	.015 (.145)	.03 (.147)	.133 (.088)	.137 (.089)	.068 (.086)	.063 (.086)
	247	247	366	366	341	341
Self-beliefs (std.)	-.009 (.145)	-.023 (.131)	.131 (.094)	.135 (.091)	.083 (.115)	.083 (.113)
	247	247	366	366	341	341
LASSI - Organization and Planning (std.)	-.055 (.148)	-.043 (.149)	.044 (.088)	.061 (.094)	.034 (.12)	.038 (.124)
	247	247	366	366	341	341
LASSI - Motivation (std.)	.152 (.194)	.137 (.196)	.179** (.075)	.186** (.081)	.057 (.056)	.062 (.047)
	247	247	366	366	341	341
CARAS - Reflexivity Index (std.)	-.067 (.056)	-.056 (.044)	.048 (.082)	.089 (.085)	-.039 (.101)	-.028 (.101)
	247	247	360	360	340	340
LABS (std.)	.053 (.076)	.042 (.075)	-.107 (.118)	-.122 (.123)	-.028 (.056)	-.04 (.059)
	247	247	366	366	341	341
School FE?	Y	Y	Y	Y	Y	Y
Controls?	N	Y	N	Y	N	Y

Notes: (1) The table shows the mean and standard deviations of control group students (columns 1 and 4) and the average ITT effect without (columns 2, 5, and 8) and with covariates (columns 3, 6, and 9). (2) * significant at 10%; ** significant at 5%; *** significant at 1%. (3) Standard errors in columns 2-3, 5-6, and 8-9 are clustered at the school level.

Table A.14: Lee bounds estimates of ITT effects on academic mindsets, perseverance, and learning strategies, by year (2015-2016)

	2015				2016			
	Lower (1)	Upper (2)	95% CI (3)	N (4)	Lower (5)	Upper (6)	95% CI (7)	N (8)
Grit (std.)	-.016 (.108)	.191* (.108)	[-.194, .369]	408	.069 (.14)	.244* (.128)	[-.164, .456]	408
DSISC (std.)	-.006 (.103)	.199* (.102)	[-.176, .368]	408	.05 (.102)	.217** (.102)	[-.118, .386]	408
Self-beliefs (std.)	-.095 (.113)	.158 (.111)	[-.281, .34]	408	.017 (.115)	.258*** (.09)	[-.172, .406]	408
LASSI - Organization and Planning (std.)	-.073 (.104)	.126 (.118)	[-.244, .321]	408	-.021 (.122)	.13 (.116)	[-.224, .322]	408
LASSI - Motivation (std.)	.094 (.101)	.222** (.101)	[-.073, .389]	408	.124 (.126)	.214* (.126)	[-.094, .43]	408
CARAS - Reflexivity Index (std.)	-.094 (.099)	.014 (.096)	[-.26, .175]	408	-.032 (.113)	.068 (.112)	[-.224, .258]	408
LABS (std.)	-.134 (.116)	.12 (.105)	[-.326, .293]	408	-.258** (.118)	-.022 (.095)	[-.452, .134]	408

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.15: ITT effects on academic mindsets, perseverance, and learning strategies by income (2014-2016)

	Constant (1)	Treatment (2)	Low SES (3)	Interaction (4)	N (5)
Grit (std.)	-.055 (.058)	-.111** (.049)	-.225** (.074)	.655*** (.155)	1102
DSISC (std.)	.28*** (.063)	-.055 (.061)	-.17 (.135)	.608** (.2)	1102
Self-beliefs (std.)	-.06 (.057)	-.128* (.069)	-.305** (.101)	.801*** (.176)	1102
LASSI - Organization and Planning (std.)	.025 (.084)	-.152* (.082)	-.138 (.101)	.757*** (.189)	1102
LASSI - Motivation (std.)	-.103** (.04)	.011 (.049)	-.16 (.125)	.455* (.214)	1102
CARAS - Reflexivity Index (std.)	-.001 (.047)	.081 (.056)	.365** (.118)	-.232 (.134)	1094
LABS (std.)	.029 (.06)	-.041 (.042)	-.16 (.16)	-.112 (.159)	1102

Notes: (1) The table examines whether the program differentially impacted students by income. Column 4 shows the interaction between the treatment and female dummies for each variable at baseline. (2) * significant at 10%; ** significant at 5%; *** significant at 1%. (3) Standard errors in columns 2-3 are clustered at the school level. (4) All estimations include school and fixed effects, but no controls.

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

<i>Panel A. First stage</i>						
	Months/year with scholarship		Months/year with mentoring		Months/year with both	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	7.431*** (.317) 1102	7.405*** (.315) 1102	7.048*** (.292) 1102	7.013*** (.293) 1102	6.825*** (.303) 1102	6.788*** (.303) 1102
<i>Panel B. Second stage</i>						
	Scholarships		Mentoring		Combined	
	(1)	(2)	(3)	(4)	(5)	(6)
Grit (std.)	.007 (.007) 1102	.008 (.007) 1102	.007 (.007) 1102	.009 (.008) 1102	.008 (.008) 1102	.009 (.008) 1102
DSISC (std.)	.013 (.009) 1102	.014 (.01) 1102	.013 (.01) 1102	.014 (.01) 1102	.014 (.01) 1102	.015 (.01) 1102
Self-beliefs (std.)	.01 (.011) 1102	.01 (.011) 1102	.01 (.012) 1102	.01 (.012) 1102	.011 (.013) 1102	.011 (.012) 1102
LASSI - Organization and Planning (std.)	.004 (.011) 1102	.006 (.011) 1102	.004 (.011) 1102	.006 (.012) 1102	.004 (.012) 1102	.006 (.012) 1102
LASSI - Motivation (std.)	.017* (.009) 1102	.016* (.009) 1102	.018** (.009) 1102	.017* (.009) 1102	.018** (.009) 1102	.018** (.009) 1102
CARAS - Reflexivity Index (std.)	0 (.009) 1094	.003 (.01) 1094	0 (.01) 1094	.004 (.01) 1094	0 (.01) 1094	.004 (.01) 1094
LABS (std.)	-.007 (.006) 1102	-.009 (.006) 1102	-.008 (.006) 1102	-.009 (.006) 1102	-.008 (.006) 1102	-.01* (.006) 1102
School FE?	Y	Y	Y	Y	Y	Y
Year FE?	Y	Y	Y	Y	Y	Y
Controls?	N	Y	N	Y	N	Y

Notes: (1) Panel A shows the results from the first stage regression of the number of months per year with a scholarship (columns 1-2), a mentoring session (column 3-4), or both (column 5-6), with and without covariates, pooled across 2014-2016. (2) Panel B shows the dose-response relationship of scholarships (columns 1-2) and mentoring (columns 3-4) and the TOT effect of both (columns 5-6), without and with covariates. (3) * significant at 10%; ** significant at 5%; *** significant at 1%. (4) Standard errors are clustered at the school level.

Table A.17: ITT effects on school performance, by year (2014-2016)

	2014			2015			2016		
	Control (1)	Effect size (2)	Effect size (3)	Control (4)	Effect size (5)	Effect size (6)	Control (7)	Effect size (8)	Effect size (9)
Language grade (std.)	0 (1)	.195* (.099)	.177* (.093)	0 (1)	-.039 (.092)	-.068 (.094)	0 (1)	-.156 (.108)	-.183 (.111)
Math grade (std.)	203 (1)	406 (.103)	406 (.097)	159 (1)	331 (.077)	331 (.077)	127 (1)	275 (.101)	275 (.107)
No. of pending subjects	0 (1)	.103 (.097)	.092 (.097)	0 (1)	-.005 (.077)	-.017 (.077)	0 (1)	-.108 (.101)	-.133 (.107)
	203 (1)	406 (.103)	406 (.097)	158 (1)	331 (.077)	331 (.077)	127 (1)	274 (.101)	274 (.107)
No. of absences	1.516 (2.511)	-.494** (.191)	-.461** (.169)	2.688 (3.717)	-.307 (.433)	-.265 (.455)	.912 (.907)	-.145 (.095)	-.138 (.091)
	190 (33.4)	386 (-7.098**)	386 (-6.455**)	157 (35.582)	325 (-4.43)	325 (.122)	125 (35.607)	269 (.395)	269 (1.322)
Failed grade	173 (28.056)	346 (2.337)	346 (2.257)	152 (26.657)	321 (2.047)	321 (2.151)	128 (23.622)	277 (2.232)	277 (2.436)
	.148 (.356)	-.06** (.022)	-.054** (.02)	.253 (.436)	-.033 (.035)	-.03 (.038)	.162 (.37)	.062* (.032)	.071* (.034)
Transferred schools	203 (.025)	406 (-.01)	406 (-.009)	170 (.088)	348 (-.012)	348 (-.011)	148 (.108)	303 (-.038)	303 (-.035)
	.155 (.155)	.012 (.012)	.014 (.014)	.284 (.284)	.039 (.039)	.038 (.038)	.312 (.312)	.027 (.027)	.026 (.026)
Dropped out of school	203 (.054)	406 (-.026)	406 (-.025)	170 (.024)	348 (-.013)	348 (-.012)	148 (.074)	303 (-.038)	303 (-.036)
	.227 (.227)	.021 (.021)	.02 (.02)	.152 (.152)	.019 (.019)	.02 (.02)	.263 (.263)	.026 (.026)	.027 (.027)
School FE?	203 (.023)	406 (.023)	406 (.023)	170 (.170)	348 (.348)	348 (.348)	148 (.148)	303 (.303)	303 (.303)
Controls?		Y (N)	Y (Y)		Y (N)	Y (Y)		Y (N)	Y (Y)

Notes: (1) The table shows the mean and standard deviations of control group students (columns 1, 4 and 7) and the average ITT effect without (columns 2, 5, and 8) and with covariates (columns 3, 6, and 9). (2) * significant at 10%; ** significant at 5%; *** significant at 1%. (3) Standard errors in columns 2-3, 5-6, and 8-9 are clustered at the school level.

Table A.18: ITT effects on school performance with inverse probability weighting, by year (2015-2016)

	2015		2016	
	Effect size (1)	Effect size (2)	Control (3)	Effect size (4)
Language grade (std.)	-.048 (.103) 300	-.074 (.103) 300	-.18 (.121) 275	-.205 (.122) 275
Math grade (std.)	.034 (.082) 300	.026 (.077) 300	-.128 (.119) 274	-.146 (.125) 274
No. of pending subjects	-.395 (.456) 295	-.353 (.471) 295	-.116 (.095) 269	-.111 (.091) 269
No. of absences	-.028 (2.291) 290	.395 (2.407) 290	1.896 (3.256) 277	2.764 (3.497) 277
Failed grade	-.046 (.036) 317	-.041 (.039) 317	.075* (.04) 303	.082* (.042) 303
Transferred schools	-.006 (.048) 317	-.005 (.047) 317	-.037 (.025) 303	-.036 (.024) 303
Dropped out of school	-.014 (.019) 317	-.013 (.02) 317	-.046 (.029) 303	-.043 (.029) 303
School FE?	Y	Y	Y	Y
Controls?	N	Y	N	Y

Notes: (1) The table shows the mean and standard deviations of control group students (columns 1, 4 and 7) and the average ITT effect without (columns 2, 5, and 8) and with covariates (columns 3, 6, and 9). (2) * significant at 10%; ** significant at 5%; *** significant at 1%. (3) Standard errors in columns 2-3, 5-6, and 8-9 are clustered at the school level.

Table A.19: Lee bounds estimates of ITT effects on school performance, by year (2015-2016)

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

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.20: TOT effect on school performance (2014-2016)

<i>Panel A. First stage</i>						
	Months/year with scholarship		Months/year with mentoring		Months/year with both	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	7.908*** (.253) 1012	7.866*** (.255) 1012	7.444*** (.239) 1012	7.391*** (.249) 1012	7.218*** (.243) 1012	7.162*** (.252) 1012
<i>Panel B. Second stage</i>						
	Scholarships		Mentoring		Combined	
	(1)	(2)	(3)	(4)	(5)	(6)
Language grade (std.)	.003 (.01) 1012	0 (.01) 1012	.004 (.011) 1012	0 (.011) 1012	.004 (.011) 1012	0 (.011) 1012
Math grade (std.)	.001 (.007) 1011	-.001 (.007) 1011	.001 (.007) 1011	-.001 (.007) 1011	.001 (.007) 1011	-.001 (.007) 1011
No. of pending subjects	-.042* (.022) 980	-.038* (.023) 980	-.044* (.023) 980	-.04* (.024) 980	-.045* (.024) 980	-.041* (.024) 980
No. of absences	-.335** (.17) 944	-.251 (.186) 944	-.357** (.178) 944	-.267 (.195) 944	-.367** (.184) 944	-.275 (.201) 944
Failed grade	-.002 (.002) 1057	-.001 (.002) 1057	-.002 (.002) 1057	-.001 (.003) 1057	-.002 (.002) 1057	-.001 (.003) 1057
Transferred schools	-.003 (.002) 1057	-.002** (.001) 1057	-.003 (.002) 1057	-.002 (.002) 1057	-.003 (.002) 1057	-.002 (.002) 1057
Dropped out of school	-.003 (.002) 1057	-.003 (.002) 1057	-.004** (.002) 1057	-.003 (.002) 1057	-.004** (.002) 1057	-.003 (.002) 1057
School FE?	Y	Y	Y	Y	Y	Y
Year FE?	Y	Y	Y	Y	Y	Y
Controls?	N	Y	N	Y	N	Y

Notes: (1) Panel A shows the results from the first stage regression of the number of months per year with a scholarship (columns 1-2), a mentoring session (column 3-4), or both (column 5-6), with and without covariates, pooled across 2014-2016. (2) Panel B shows the dose-response relationship of scholarships (columns 1-2) and mentoring (columns 3-4) and the TOT effect of both (columns 5-6), without and with covariates. (3) * significant at 10%; ** significant at 5%; *** significant at 1%. (4) Standard errors are clustered at the school level.

Table A.21: ITT effects on student achievement, by year (2015-2016)

	2015			2016		
	Control (1)	Effect size (2)	Effect size (3)	Control (4)	Effect size (5)	Effect size (6)
Math achievement (std.)	0 (1) 164	.054 (.14) 335	.013 (.14) 335	.007 (.927) 168	-.161* (.081) 348	-.171* (.082) 348
Reading achievement (std.)	0 (1) 173	-.067 (.085) 349	-.088 (.087) 349	-.158 (1.094) 173	.05 (.099) 357	.017 (.093) 357
School FE?		Y	Y		Y	Y
Controls?		N	Y		N	Y

Notes: (1) The table shows the mean and standard deviations of control group students (columns 1 and 4) and the average ITT effect without (columns 2 and 5) and with covariates (columns 3 and 6). (2) * significant at 10%; ** significant at 5%; *** significant at 1%. (3) Standard errors in columns 2-3, 5-6, and 8-9 are clustered at the school level.

Table A.22: ITT effects on student achievement with inverse probability weighting, by year (2015-2016)

	2015		2016	
	Effect size (1)	Effect size (2)	Effect size (3)	Effect size (4)
Math achievement (std.)	.06 (.141) 335	.02 (.141) 335	-.132 (.087) 311	-.139 (.087) 311
Reading achievement (std.)	-.08 (.087) 349	-.102 (.092) 349	.083 (.124) 320	.047 (.118) 320
School FE?	Y	Y	Y	Y
Controls?	N	Y	N	Y

Notes: (1) The table shows the mean and standard deviations of control group students (columns 1 and 4) and the average ITT effect without (columns 2 and 5) and with covariates (columns 3 and 6). (2) * significant at 10%; ** significant at 5%; *** significant at 1%. (3) Standard errors in columns 2-3, 5-6, and 8-9 are clustered at the school level.

Table A.23: Lee bounds estimates of ITT effects on student achievement, by year (2015-2016)

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

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.24: ITT effects on student achievement by income (2014-2016)

	Constant (1)	Treatment (2)	Low SES (3)	Interaction (4)	N (5)
Math achievement (std.)	-.449*** (.058)	-.177** (.074)	-.494*** (.142)	.416* (.185)	683
Reading achievement (std.)	-.778*** (.042)	-.118* (.058)	-.422** (.135)	.37* (.183)	706

Notes: (1) The table examines whether the program differentially impacted students by income. Column 4 shows the interaction between the treatment and female dummies for each variable at baseline. (2) * significant at 10%; ** significant at 5%; *** significant at 1%. (3) Standard errors in columns 2-3 are clustered at the school level. (4) All estimations include school and fixed effects, but no controls.

Table A.25: TOT effect on student achievement (2015-2016)

<i>Panel A. First stage</i>						
	Months/year with scholarship		Months/year with mentoring		Months/year with both	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	6.827*** (.355) 683	6.803*** (.348) 683	6.584*** (.328) 683	6.555*** (.325) 683	6.332*** (.333) 683	6.301*** (.326) 683
<i>Panel B. Second stage</i>						
	Scholarships		Mentoring		Combined	
	(1)	(2)	(3)	(4)	(5)	(6)
Math achievement (std.)	-.008 (.013) 683	-.011 (.013) 683	-.008 (.013) 683	-.012 (.013) 683	-.009 (.014) 683	-.012 (.014) 683
Reading achievement (std.)	-.001 (.009) 706	-.005 (.009) 706	-.001 (.01) 706	-.006 (.01) 706	-.001 (.01) 706	-.006 (.01) 706
School FE?	Y	Y	Y	Y	Y	Y
Year FE?	Y	Y	Y	Y	Y	Y
Controls?	N	Y	N	Y	N	Y

Notes: (1) Panel A shows the results from the first stage regression of the number of months per year with a scholarship (columns 1-2), a mentoring session (column 3-4), or both (column 5-6), with and without covariates, pooled across 2015-2016. (2) Panel B shows the dose-response relationship of scholarships (columns 1-2) and mentoring (columns 3-4) and the TOT effect of both (columns 5-6), without and with covariates. (3) * significant at 10%; ** significant at 5%; *** significant at 1%. (4) Standard errors are clustered at the school level.

Table A.26: ITT effects on personality traits with inverse probability weighting (2016)

	Effect size	
	(1)	(2)
Extraversion (std.)	-.064 (.089) 341	-.051 (.096) 341
Agreeableness (std.)	.162 (.126) 341	.136 (.129) 341
Conscientiousness (std.)	-.254** (.109) 341	-.227* (.114) 341
Neuroticism (std.)	-.118 (.084) 341	-.108 (.084) 341
Openness (std.)	.166 (.101) 341	.137 (.103) 341
School FE?	Y	Y
Controls?	N	Y

Notes: (1) The table shows the mean and standard deviations of control group students (column 1) and the average ITT effect without (column 2) and with covariates (column 3). (2) * significant at 10%; ** significant at 5%; *** significant at 1%. (3) Standard errors in columns 2 and 3 are clustered at the school level. (4) The bottom rows display the result from a joint test using Seemingly Unrelated Regressions (SURs).

Table A.27: Lee bounds estimates of ITT effects on personality traits (2016)

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

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

Table A.28: TOT effect on personality traits (2016)

<i>Panel A. First stage</i>						
	Months/year with scholarship		Months/year with mentoring		Months/year with both	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	5.708*** (.439) 341	5.69*** (.434) 341	5.545*** (.415) 341	5.518*** (.413) 341	5.312*** (.415) 341	5.287*** (.407) 341
<i>Panel B. Second stage</i>						
	Scholarships		Mentoring		Combined	
	(1)	(2)	(3)	(4)	(5)	(6)
Extraversion (std.)	-.01 (.015) 341	-.007 (.016) 341	-.01 (.015) 341	-.008 (.016) 341	-.01 (.016) 341	-.008 (.017) 341
Agreeableness (std.)	.032* (.018) 341	.027 (.019) 341	.033* (.019) 341	.028 (.019) 341	.034* (.019) 341	.029 (.02) 341
Conscientiousness (std.)	-.043** (.018) 341	-.039** (.019) 341	-.044** (.019) 341	-.04** (.019) 341	-.046** (.019) 341	-.042** (.02) 341
Neuroticism (std.)	-.018 (.014) 341	-.017 (.014) 341	-.019 (.014) 341	-.018 (.015) 341	-.02 (.015) 341	-.019 (.015) 341
Openness (std.)	.033* (.019) 341	.028 (.018) 341	.033* (.019) 341	.029 (.019) 341	.035* (.02) 341	.03 (.019) 341
School FE?	Y	Y	Y	Y	Y	Y
Year FE?	Y	Y	Y	Y	Y	Y
Controls?	N	Y	N	Y	N	Y

Notes: (1) Panel A shows the results from the first stage regression of the number of months per year with a scholarship (columns 1-2), a mentoring session (column 3-4), or both (column 5-6), with and without covariates in 2016. (2) Panel B shows the dose-response relationship of scholarships (columns 1-2) and mentoring (columns 3-4) and the TOT effect of both (columns 5-6), without and with covariates. (3) * significant at 10%; ** significant at 5%; *** significant at 1%. (4) Standard errors are clustered at the school level.

Appendix B Survey of students' academic behaviors

This appendix describes each set of questions in the survey of students' academic behaviors. Figures B.1 and B.2 display the distribution of raw scores in these questions in 2015 and 2016.

B.1 Proactive school behavior

This set of questions asked students to recall the last time they did not understand something in class and report whether they: (a) asked their teacher to explain a topic again; (b) asked a relative to explain it; (c) asked a peer; (d) consulted a book/Internet on the topic; (e) sought a private tutor; or (f) attended after-school lessons. The score for each item equals 1 if the student engaged in the prompted behavior and 0 otherwise. The score for the index ranges from 0 to 6.

B.2 Preventive homework behavior

This set of questions asked students to recall the last time they were assigned homework and report whether they: (a) started doing it more than a day in advance; (b) got together with peers to do it; (c) asked the teacher clarifying questions; (d) asked the teacher about the resources that could be consulted (e.g., textbooks, calculators); (e) asked the teacher whether their answers were “on the right track” before turning them in; (f) asked the teacher whether a given answer was correct; or (g) compared answers with a peer. The score for each item equals 1 if the student engaged in the prompted behavior and 0 otherwise. The score for the index ranges from 0 to 7.

B.3 Corrective homework behavior

This set of questions asked students to recall the last time they received a failing grade on their homework and report whether they: (a) asked the teacher why some answers were incorrect; (b) asked him/her to explain a topic again; (c) asked him/her to give partial credit on incorrect answers; (d) asked a relative to explain a topic; (e) compared answers with a peer; (f) sought a private tutor; or (g) attended after-school lessons. The score for each item equals 1 if the student engaged in the prompted behavior and 0 otherwise. The score for the index ranges from 0 to 7.

B.4 Preventive test behavior

This set of questions asked students to recall the last time they had to study for a test and report whether they: (a) started studying more than a day in advance; (b) met a peer to study; (c) asked a relative for help studying for the test; (d) reviewed their folder to see which topics will be covered in the test; (e) reviewed a textbook to see which topics will be covered in the test; (f) asked the teacher questions about difficult topics; (g) sought a private tutor; (h) attended after-school lessons. The score for each item equals 1 if the student engaged in the prompted behavior and 0 otherwise. The score for the index ranges from 0 to 8.

B.5 Corrective test behavior

This set of questions asked students to recall the last time they received a failing grade on a test and report whether they: (a) asked the teacher why some answers were incorrect; (b) asked him/her to explain a topic again; (c) asked him/her to give partial credit on incorrect answers; (d) asked him/her for an opportunity to make up the low grade; (e) asked a relative to explain a topic; (f) asked a peer for help (e.g., looking at their folder); (g) sought a private tutor; or (h) attended after-school lessons. The score for each item equals 1 if the student engaged in the prompted behavior and 0 otherwise. The score for the index ranges from 0 to 8.

B.6 Corrective failing behavior

This set of questions asked students to recall the last time they received a failing term grade on a subject on their school report card and report whether they: (a) asked the teacher to explain a topic again; (b) asked him/her to change the grade to a passing grade based on their performance on specific lessons or projects; (c) asked him/her for an opportunity to make up the low grade; (d) asked a relative to explain a topic; (e) asked a peer for help (e.g., looking at their folder); (f) sought a private tutor; or (g) attended after-school lessons. The score for each item equals 1 if the student engaged in the prompted behavior and 0 otherwise. The score for the index ranges from 0 to 8.

B.7 Corrective flunking behavior

This set of questions asked students to recall the last time they failed a subject and report whether they: (a) asked the teacher to explain a topic again; (b) asked him/her to change the grade to a passing grade based on their performance on specific lessons or projects; (c)

asked him/her for an opportunity to make up the low grade; (d) asked a relative to explain a topic; (e) asked the teacher which topics will be covered in the December/March exam;²⁸ (f) asked him/her what types of questions will be asked in the December/March exam; (g) asked him/her which instructors will be present in the December/March exam;²⁹ (h) asked him/her the date of the December/March exam; (i) sought a private tutor; or (j) attended after-school lessons. The score for each item equals 1 if the student engaged in the prompted behavior and 0 otherwise. The score for the index ranges from 0 to 11.

B.8 Preventive absenteeism behavior

This set of questions asked students to recall the last time they were absent to school and report whether they did any of the following before returning to school: (a) asked a peer what was covered in class; (b) caught up on readings done during class; (c) asked a peer for the homework assigned during class; or (d) asked a peer for his/her folder to copy what was done in class. The score for each item equals 1 if the student engaged in the prompted behavior and 0 otherwise. The score for the index ranges from 0 to 4.

B.9 Corrective absenteeism behavior

This set of questions asked students to recall the last time they were absent to school and report whether they did any of the following after returning to school: (a) asked a peer what was covered in class; (b) asked the teacher what was covered in class; (c) caught up on readings done during class; (d) asked a peer for the homework assigned during class; or (e) asked a peer for his/her folder to copy what was done in class. The score for each item equals 1 if the student engaged in the prompted behavior and 0 otherwise. The score for the index ranges from 0 to 5.

B.10 Corrective free period behavior

This set of questions asked students to recall the last time they had a “free period”³⁰ and report whether they: (a) did homework; (b) studied for a test; (c) read for a class; (d) talked to a friend (reverse-coded); or (e) went home (reverse-coded). The score for each item equals

---

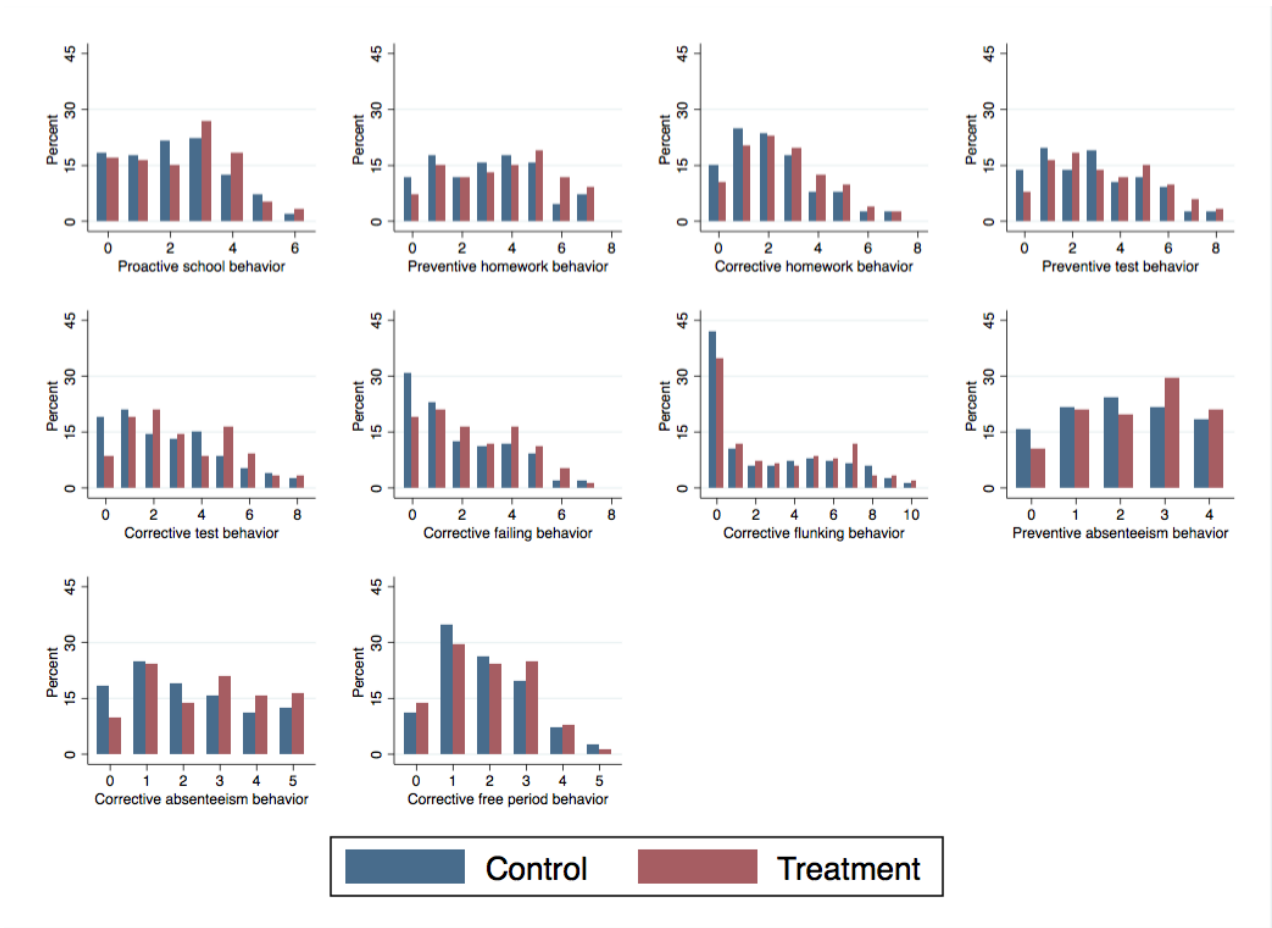
²⁸As we mentioned in a previous footnote, in Argentina, when students fail a subject, they need to take an exam to pass it in December. If they fail this exam, they need to take another exam in March. They can fail up to two subjects in March. If they fail more, they can take these exams once again before the school year begins. If they still fail more than two of these subjects by then, they are supposed to repeat the grade.

²⁹The December/March are typically administered by a group of teachers from different subjects/grades.

³⁰In Argentina, when a teacher is absent, students are sometimes allowed to remain in their classroom without having to take any lessons.

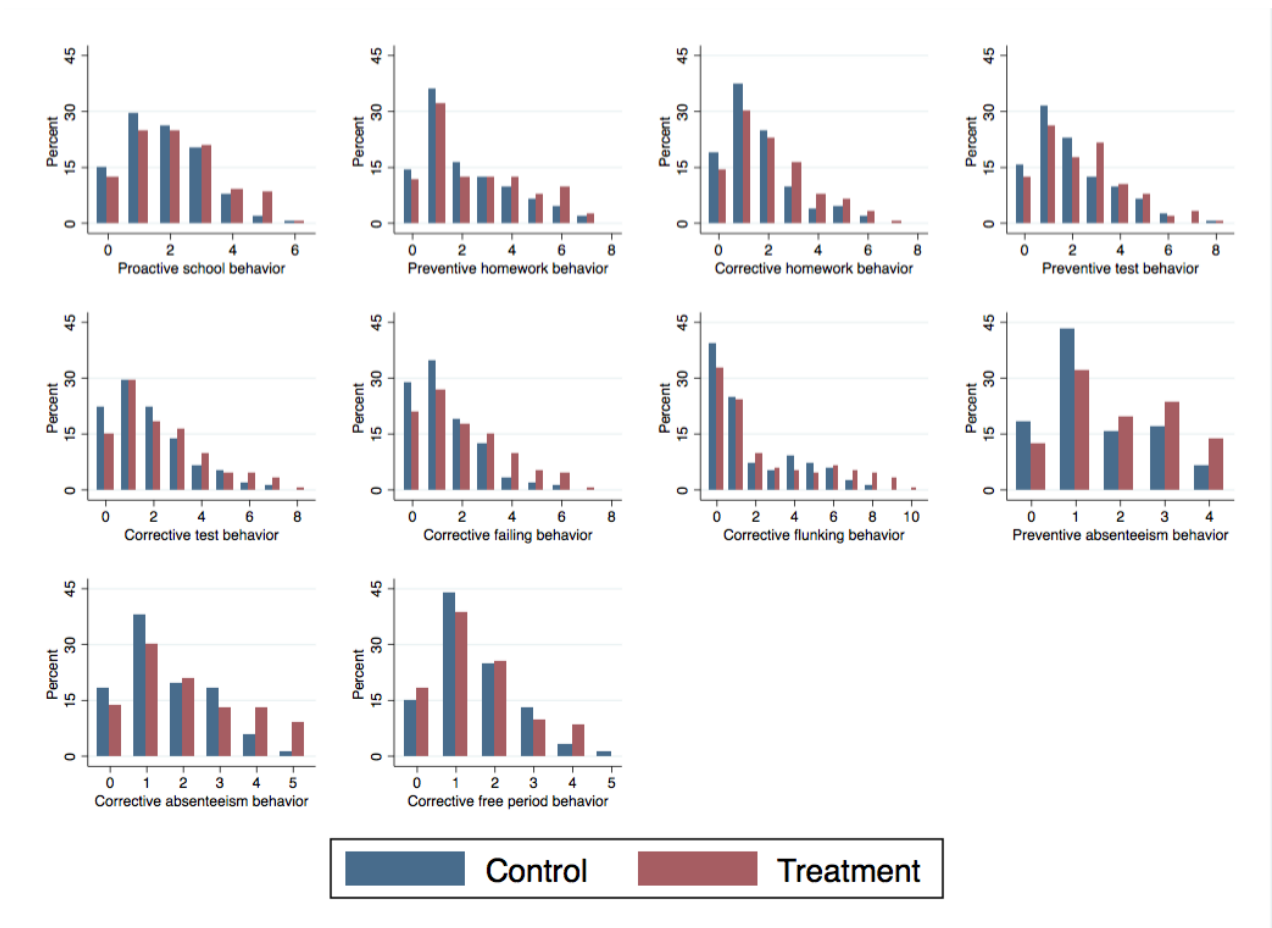
1 if the student engaged in the prompted behavior and 0 otherwise. The score for the index ranges from 0 to 5.

Figure B.1: Distributions of survey of students' academic behaviors (2015)



Note: This figure shows discrete histograms for each set of questions in the survey of academic behaviors in 2015, with a superimposed normal distribution.

Figure B.2: Distributions of survey of students' academic behaviors (2016)



Note: This figure shows discrete histograms for each set of questions in the survey of academic behaviors in 2016, with a superimposed normal distribution.

Appendix C   Surveys and assessments of students’ academic mindsets, perseverance, and learning strategies

This appendix describes each survey and assessment of students’ academic mindsets, perseverance, and learning strategies. Figures C.1- C.3 display the distribution of raw scores in these questions from 2014 to 2016.

C.1   Self-beliefs about academics

The survey of self-beliefs about academics asks students to report the extent to which they agree with 14 statements about themselves using a scale that ranges from 1 (“totally disagree”) to 5 (“totally agree”). A confirmatory factor analysis indicates that it measures two distinct types of self-beliefs: those about performance (e.g., “I think I will get good grades this year”) and about self-efficacy (e.g., “I am capable of doing school assignments well, even if they are difficult”). The survey was developed by a team of Argentine psychologists at the University of Buenos Aires (UBA), and it had already administered to secondary school students in the PBA (Schmidt et al. 2008) and to SMP participants (Pais et al. 2013).

C.2   Learning and Study Strategies Inventory (LASSI)

The Learning and Study Strategies Inventory (LASSI) asks students to report the extent to which how frequently they find themselves in 10 different situations, from 1 (“Never”) to 5 (“Always”). According to a factor analysis, seven of these situations measure students’ organization and planning skills (e.g., “I have trouble putting together a study plan and sticking to it”) and three measure their motivation (e.g., “I try hard to get good grades, even in subjects that I do not like”). This survey was developed by psychologists at the University of Texas at Austin (Weinstein and Palmer 1988) and it was later adjusted for and administered to Argentine teenagers and adults by psychologists at the UBA (Fernandez Liporace and Casullo 2009). It has already been administered to SMP participants (Pais et al. 2013).

C.3   Short Grit scale (Grit-S)

The short Grit scale (Grit-S) consists of eight questions that ask students how frequently they find themselves in a given situation, from 1 (“Almost never”) to 5 (“At least once a day”). According to factor analyses, four of these items measure students’ consistency (e.g., “I forget

some of the things I need for school”) and the other four measure students’ perseverance (e.g., “I interrupt others while they are speaking”). The survey was developed by psychologist Angela Duckworth at the University of Pennsylvania (Duckworth and Quinn 2009). To our knowledge, this is the first time that it was administered in Argentina.

C.4 Domain-Specific Impulsivity Scale for Children (DSIS-C)

The Domain-Specific Impulsivity Scale for Children (DSIS-C) describes eight traits or situations to students (e.g., “I am very diligent” or “I have been obsessed with an idea or project for a short period of time, but I later lost interest”) and asks them to indicate whether these descriptions match them, from 1 (“Not at all like me”) to 5 (“Very much like me”). It was developed by a team of psychologists at the University of Pennsylvania (Tsukayama et al. 2013) and it has previously been administered in Argentina (Pais 2015).

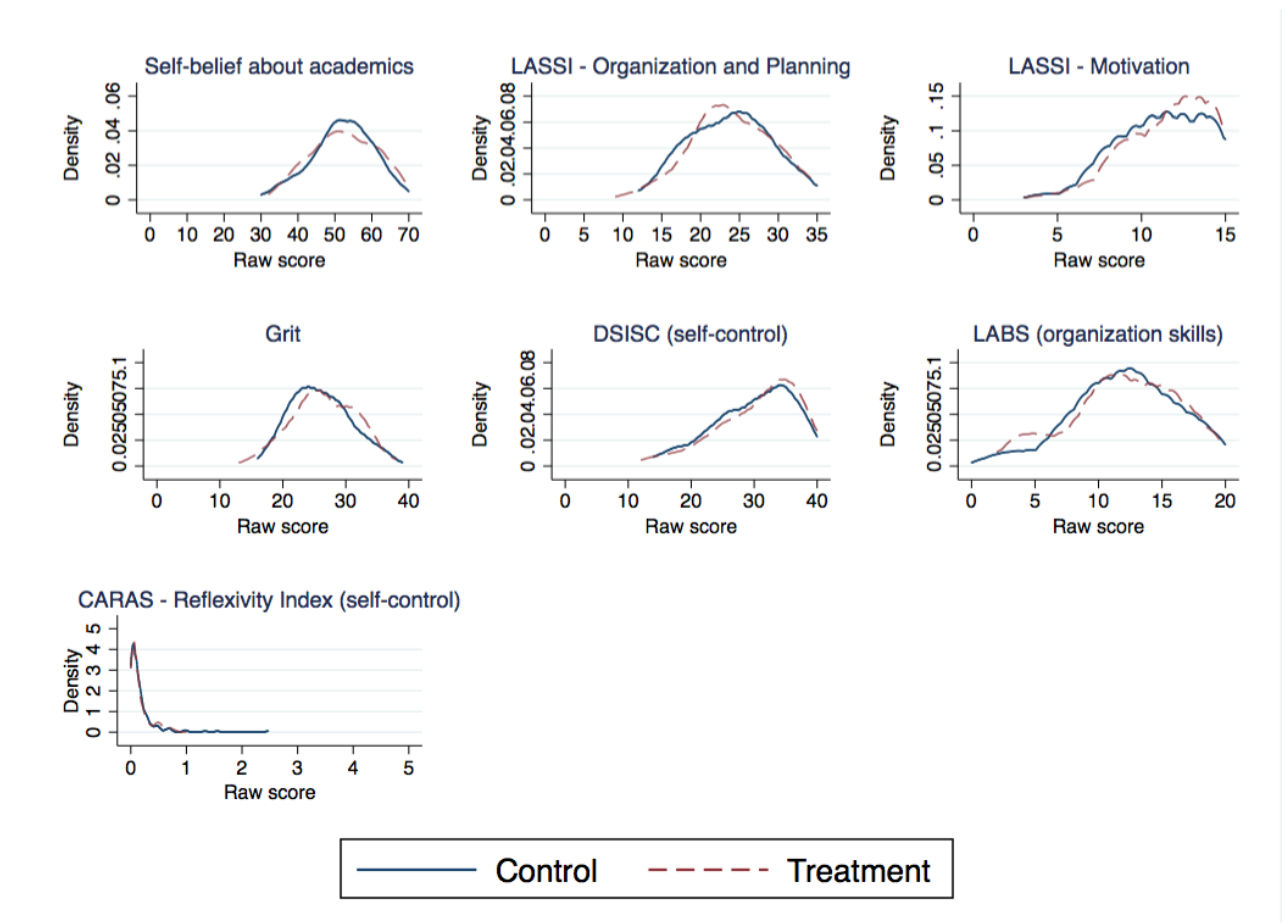
C.5 Labs

The Labs assessment asks students to make their way out of 10 increasingly difficult labyrinths without lifting their pencil. Each student’s score is determined based on the number of mistakes he/she made (i.e., “dead ends” in the labyrinth that they encountered while trying to solve it) as well as the number of labyrinths he/she solved. It was developed by psychologist David Wechsler (Wechsler 1994), and it had previously been administered in Argentina (Arán-Filipetti 2012; Arán-Filipetti and López 2013; Arán-Filipetti and Richaud de Minzi 2011; Cayssials 2003; Martos Mula et al. 2013; Soprano 2003).

C.6 *Caras*

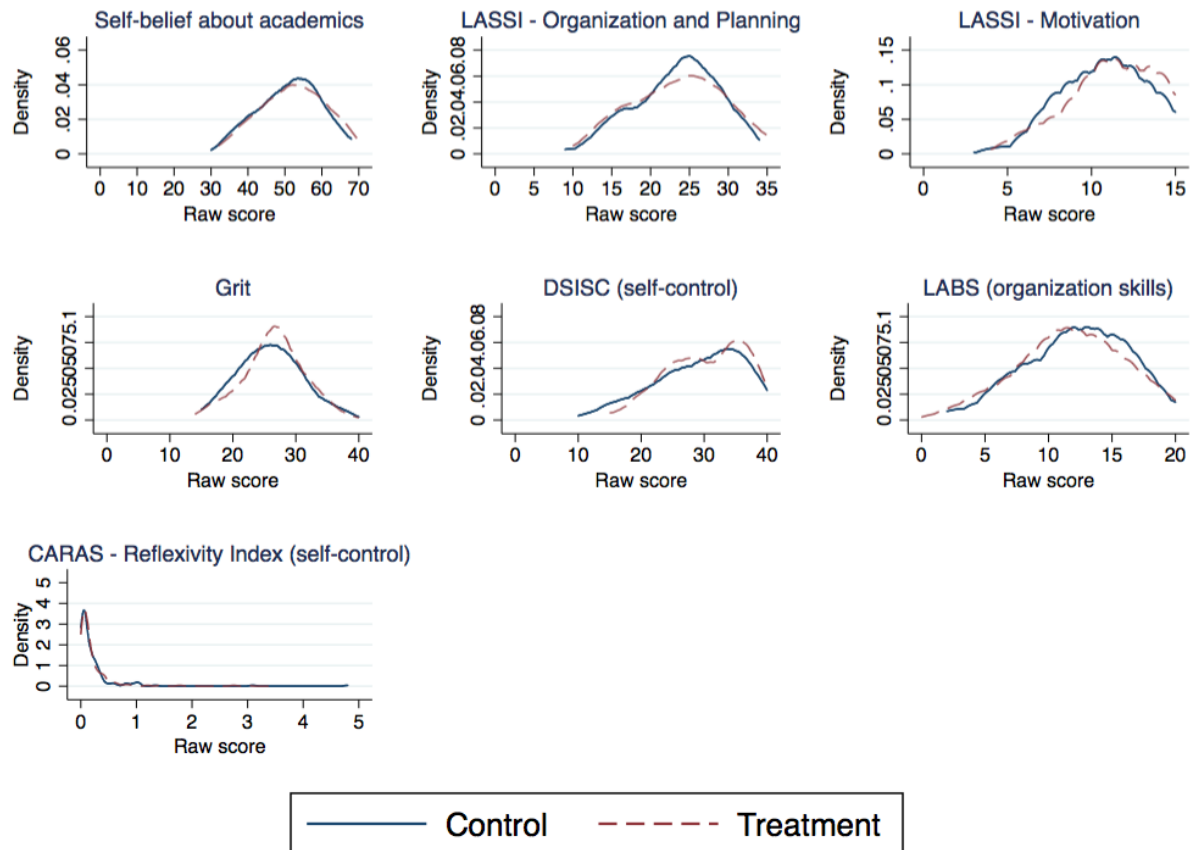
The *Caras* (Smileys) assessment shows students 45 sets of three smileys and asks them to cross out the smiley that is not like the others. The metric of interest is the “reflexivity index”: the number of net (i.e., correct minus incorrect) correct answers, divided by the number of errors. This assessment was developed by American and a Spanish psychologists (Thurstone and Yela 2001) and it has previously been administered in Argentina (Arán-Filipetti 2012; Arán-Filipetti and López 2013; Arán-Filipetti and Richaud de Minzi 2011).

Figure C.1: Distributions of survey of students' academic mindsets, perseverance, and learning strategies (2014)



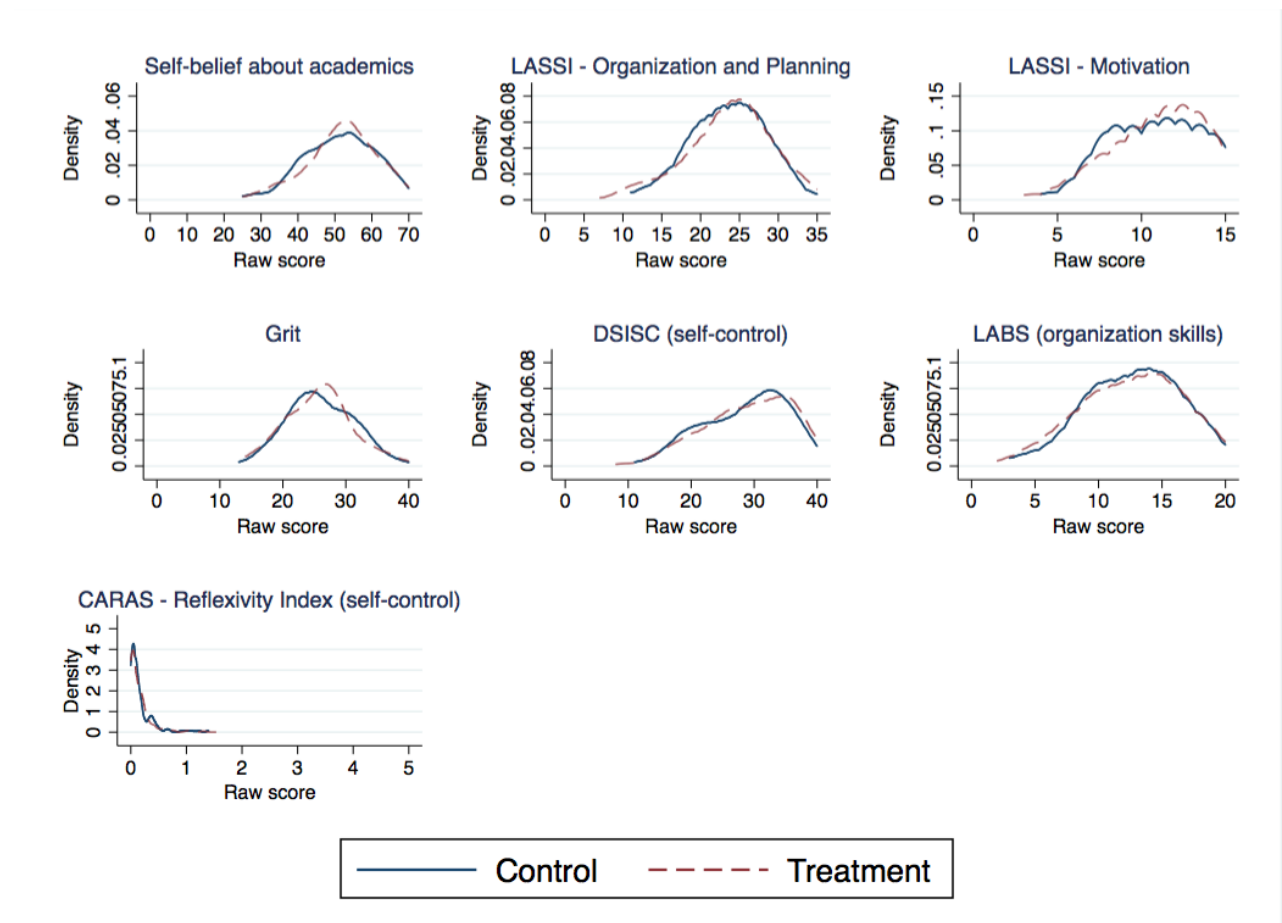
Note: This figure shows the distribution of raw scores for each survey and assessment of academic mindsets, perseverance, and learning strategies in 2014, with a superimposed normal distribution.

Figure C.2: Distributions of survey of students' academic mindsets, perseverance, and learning strategies (2015)



Note: This figure shows the distribution of raw scores for each survey and assessment of academic mindsets, perseverance, and learning strategies in 2015, with a superimposed normal distribution.

Figure C.3: Distributions of survey of students' academic mindsets, perseverance, and learning strategies (2016)



Note: This figure shows the distribution of raw scores for each survey and assessment of academic mindsets, perseverance, and learning strategies in 2016, with a superimposed normal distribution.

Appendix D Student achievement tests

This appendix describes the design and scoring of the student assessments of math and reading. Figure D.1 displays the distribution of scaled scores in these tests in 2015 and 2016.

D.1 Design

The tests were designed to assess a wide array of domains and skills in math and reading at different difficulty levels. 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. It included eight questions of low difficulty, 12 questions of medium difficulty, and 10 questions of high difficulty. 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. It included nine questions of low difficulty, 12 questions of medium difficulty, and nine questions of high difficulty. The item maps are available from the authors.

D.2 Scoring

Items were first scored dichotomously and then scaled using Item Response Theory (IRT). IRT models the relationship between an underlying latent trait (i.e., a student's ability) and the probability that the student will answer a given question on a test (i.e., item) correctly (Yen and Fitzpatrick 2006). IRT is used in large-scale assessments for three main reasons. First, it allows each item to contribute differentially to the estimation of student ability (as opposed to percent-correct scores, which assign the same dichotomous score to each item). Second, it allows researchers to place different (rounds of) assessments on a common scale, provided that they share a subset of items, students, or both. Third, it allows researchers to assess the performance of each individual item, which is particularly useful for test design.

There are three IRT models that are frequently used to scale items scored dichotomously. All assume a single underlying latent trait, but differ in the item characteristics they consider (Harris 2005). We used a three-parameter logistic model, which estimates P_i , the probability that a student will answer item i correctly, based on: θ_p , student p 's ability; a_i , the "discrimination" parameter (i.e., the slope of the Item Characteristic Curve (ICC) at the point of inflection, which reflects how well the item can differentiate between students of

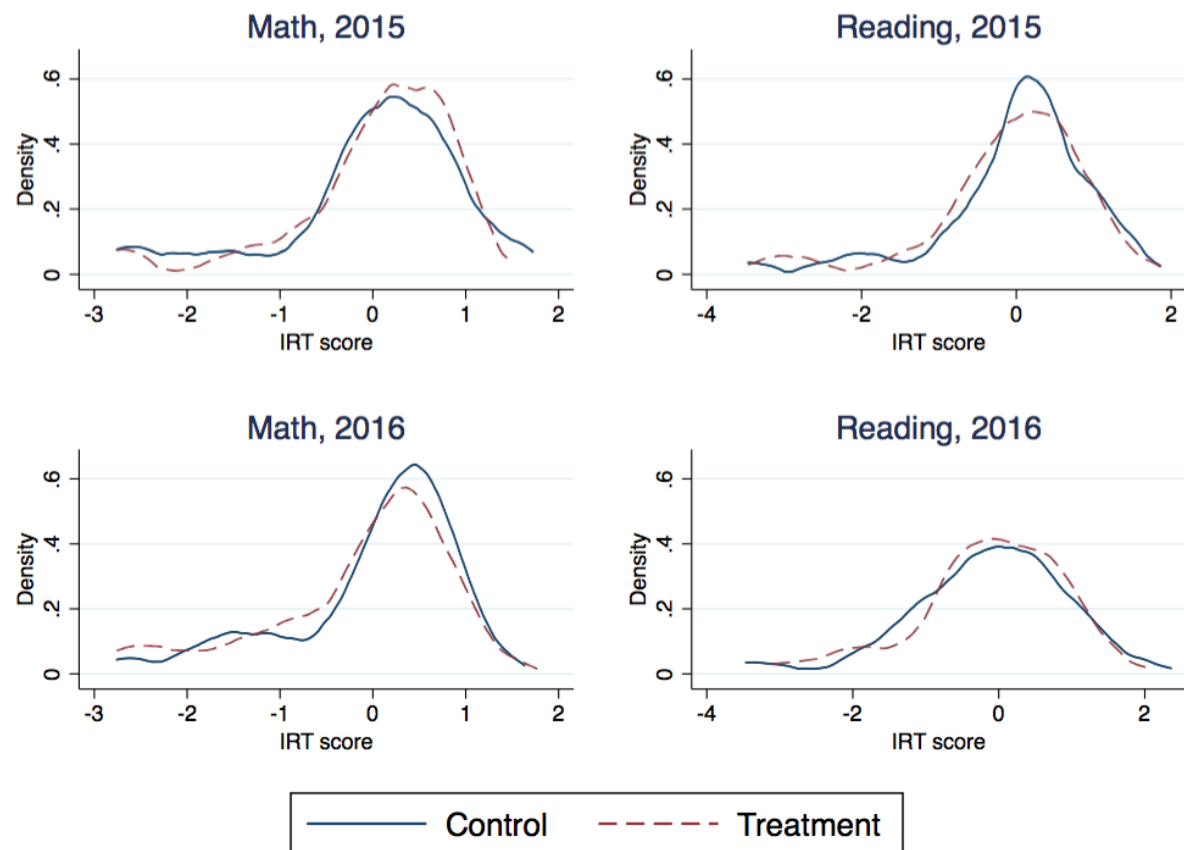
similar ability); b_i , the “difficulty” parameter (i.e., the point of inflection on the θ scale, which reflects where the item functions along the ability scale); and c_i , the “pseudo-guessing” parameter, the asymptotic probability that students will answer the item correctly by chance. The three-parameter model is thus given by:

$$P_i(\theta_p) = c_i + \frac{1 - c_i}{1 + e^{[-1.7a_i(\theta_p - b_i)]}} \quad (\text{D.1})$$

where all parameters are defined as above. The model uses a logistic function to relate student ability and the item parameters to the probability of answering an item correctly.

We generated maximum likelihood estimates of student achievement, which are unbiased individual measures of ability, using the OpenIRT Stata program developed by Tristan Zajonc. Bayesian Markov chain Monte Carlo estimates are similar and available from the authors.

Figure D.1: Distributions of IRT scores on student achievement tests (2015 and 2016)



Note: This 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. The scores have been estimated using a three-parameter Item Response (IRT) model and standardized with respect to the control group in 2015.