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¶

November 8, 2017

Abstract

Many developing countries provide 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 program beneficiaries may lack 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 positively impacted students' academic behaviors (e.g., starting to study early before an exam or catching up on schoolwork missed due to absences). Yet, we find very little evidence that it improved their academic mindsets (e.g., self-beliefs about performance and self-efficacy), perseverance (e.g., grit), or learning strategies (e.g., metacognition). It 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 may be partly because a large number of treatment students were expelled from the program for not meeting its requirements. Finally, we do not find that the program positively impacted 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, Isy Faingold Vigil, Ariel Fiszbein, Verónica Frisancho, Brad Hershbein, Richard Murnane, Ernesto Pais, Jeffrey Puryear, Laura Trucco, Michael Walker, and seminar participants at AEFP, 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. It assesses the impact of the program on immediate (e.g., academic behaviors), intermediate (e.g., academic perseverance and mindsets), and final outcomes (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 very little evidence that the program improved students' academic mindsets (e.g., self-beliefs about performance and self-efficacy), perseverance (e.g., grit), or learning strategies (e.g., metacognition), as measured by self-reports and performance assessments, but we cannot discard the possibility that it had small to moderate effects on these outcomes.

[1] There 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 terms interchangeably to refer to "patterns of thought, feelings, and behavior" (Borghans et al. 2008) other than cognitive ability that lead to student success.

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 number of students were expelled for not complying with the conditions of the program. Fourth, we find no evidence that the program positively impacted students' achievement or personality traits. Importantly, 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. However, we actually found negative impacts on some of these outcomes.

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 (Banerjee et al. 2013). First, the costs of doing so may be too high. These include the *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]

[2]Some 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 most of these programs have not impacted 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 exercise teacher to deliver 32 fully-scripted, 45-minute counseling sessions per week 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.[4] 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,

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).

[3]This 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.

[4]An important source of learning anxiety in this context are competitive high school entrance exams.

while this mentoring program seems appropriate for school systems that place considerable 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 middle-income countries in Latin America.

3 Experiment

3.1 Context

Schooling in Argentina is compulsory and free from age 4 until the end of secondary school. In 12 out 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).[5] 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 runs from February to December.

Education policy in Argentina is shaped by both the national and sub-national (province) governments. According to the National Education Law (LEN) 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 of complementing scholarships with 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).[6] 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).[7] 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).

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

[6] By 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).

[7] Further, 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) is a program that combines a scholarship with non-academic mentoring. It is offered to students entering secondary school (grade 8), and if they comply with program requirements, they can stay in the program until they graduate (grade 12). It is run by one the largest education non-profit in Argentina, which decided to remain anonymous. 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 the country: in 2015, it reached 2,544 students in 16 provinces and the Autonomous City of Buenos Aires.

3.2.1 Scholarship

Each student who participates in the SMP is supposed to receive 10 scholarships from March to December of each year (one per month). Each monthly scholarship is about USD 40, which is on par with other cash transfer programs for secondary school (see Fiszbein et al. 2009). The funds are deposited in a bank account in the name the student's parent or legal guardian and they can be withdrawn at any time and be used for any purpose.

3.2.2 Mentoring

Each student in the SMP is required to participate in 10 meetings with a mentor,[8] from March to December of each year (one per month). Each session lasts 30-45 minutes and is typically held at the school, before or after the school day.[9] 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.

The content of the mentoring sessions is not pre-determined by the foundation, but decided by each mentor based on the needs of each individual student. Yet, all sessions have a common structure: (a) an "icebreaker", in which the mentor seeks to earn the trust of the student, and the student shares his/her schoolwork, as well as a number of reports from teachers and school staff required by the program;[10] (b) a "diagnosis", in which students discuss their experience

[8] Mentors typically have a bachelor's degree in psychology, pedagogical psychology, social work, or education, or they have graduated from a teacher-training program. When they join the foundation, they undergo an induction process and receive a manual. Each year, they have a face-to-face and an online training session. During the year, they also draw resources from and exchange ideas through an online platform.

[9] Whenever it is not possible to meet students before or after school, they are pulled out of their classrooms to attend these meetings.

[10] Students 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).

at school with the mentor as well as their strengths and weaknesses; and (c) an "action plan", in which students and the mentor agree on specific goals (e.g., studying for a math exam).

On the first and last mentoring sessions of each year, mentors assess whether each student should continue in the program. Mentors may suspend or terminate students' participation if they repeat grades, are suspended, or switch schools.[11]

3.2.3 Monitoring

The SMP has several commitment devices to keep students engaged. To join the program, students sign a "commitment contract" in which they agree to: (a) attend regularly to school; (b) behave well; (c) work hard; (d) pass their subjects; and (e) graduate from secondary school. Additionally, during the school year, students are required to attend an annual meeting with other SMP beneficiaries in their area, and to communicate periodically with their sponsors (i.e., individuals, non-profits, or businesses who finance the students' participation in the program) through letters and/or attendance to events.

3.2.4 Theory of change

Table 1 presents the theory of change of the SMP. The program seeks to address two problems that may prevent youth from low-income families from staying and succeeding in secondary school: (a) high costs, low (perceived) benefits, and/or lack of access to credit; and (b) little parental experience with schooling, which limit their capacity to support their children. The scholarship seeks to address the first problem and mentoring sessions the second problem.

[Insert Table 1 here.]

There are three sets of outcomes that the program may impact: (a) immediate (short-term, directly affected by program components); (b) intermediate (medium-term, indirectly affected by the program); and (c) final (medium-term, mediated by immediate and/or intermediate outcomes). In the short-term, the scholarship seeks to increase student attendance and persistence in school and the mentoring sessions to teach students useful academic behaviors (e.g., doing homework, organizing materials, participating in class, and studying). In the medium-term, mentoring is also hypothesized to influence students' academic mindsets (e.g., self-beliefs about performance and efficacy) and perseverance (e.g., grit), as well as their learning strategies (e.g., metacognition). This could occur both through the learning and practice of academic behaviors or directly through students' interaction with their mentors.

[11] Students are dropped from the SMP if they transfer schools for logistical reasons. Mentors cannot continue supporting students who attend schools that may be far from their original school.

Ultimately, the program hopes to improve students' performance in school (i.e., through higher grades and lower grade failure). The program does not try to influence students' learning outcomes or personality traits (e.g., conscientiousness), and the foundation does not perceive these as objectives of its program, but we included these two in the theory of change because we hypothesized that they could result from the immediate and intermediate outcomes and we wanted to use this evaluation to examine whether this was the case.

3.2.5 Costs

We collected the costs of the program using the approach outlined in Dhaliwal et al. (2013). It costs USD 733 per student (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

This study was conducted in the Province of Buenos Aires (PBA). The PBA offers an ideal setting to study policies that could be scaled to the rest of Argentina. First, it 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 similarly to the average student in the country on national exams. 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).

Ten schools from the PBA were invited to participate in the study, based on three criteria: (a) they had to be public schools serving youths from low-income families; (b) they had to have previously participated in the SMP; and (c) they could not have any SMP participants in eighth grade on February 2014. The first criterion was adopted to focus on the most disadvantaged students. The second criterion sought to ensure that schools had familiarity with the SMP and its data collection process. The third criterion prevented having study participants, who were selected by lottery, in the same classroom with regular SMP participants, who were selected through an admissions process. A representative of the foundation met with each school's principal to explain the components of the evaluation. All schools accepted to participate in the study.[12]

[12]Each school was located in a different locality of the PBA: Campana, Ensenada, Gregorio de Laferrere, Guernica, José C. Paz, Merlo, Quilmes, Santos Lugares, Virrey del Pino, and Zárate.

Students who were eligible to participate in the evaluation were selected randomly. First, we randomly selected two grade 7 classrooms per school. Then, students in these classrooms were invited to participate in the evaluation as follows. The foundation sent a note to students' parents through their communications notebooks announcing the date and time of an information session for the program and study. A representative from the foundation held the information session at each school and wrote down the names of the parents of the students who were interested in participating. Finally, the same representative met with all interested parents and their children to collect the baseline data (described in Section 4). All 408 students who participated in this last stage were entered into the lottery roster.

3.4 Lottery

Students in the lottery roster were assigned to the treatment or control groups using a random number generator. The number of available spots in the program was determined by the amount of funds that the foundation had raised for each school, so we randomly assigned students to experimental groups within each school. The lottery resulted in 204 treatment and 204 control students. All treatment students were offered a spot on the SMP.

4 Data

We designed our data collection strategy to track the impact of the program on each step of its hypothesized causal chain, from the immediate to the intermediate and final outcomes outlined in Table 1. 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 the study sample as a whole. As the table shows, the students in our study were academically disadvantaged: almost a third (31%) had

repeated a grade and 5% had dropped out of school. These students were predominantly from low-income families. They had limited assets: only 21% of students had a car, 72% had a refrigerator, and 55% had a computer. They also had substandard living conditions: only 30% had natural gas, 83% had running water, and only 60% of their parents were homeowners.

[Insert Table 3 here.]

The table also checks whether the treatment and control students were comparable at baseline. This is the case for all but five variables. We ran a regression of the treatment dummy on all variables in Table 3, tested the joint significance of all coefficients using an F-test, and could not reject the null that there was no difference between the treatment and control groups.

4.2 Follow-ups

We followed students for three years and collected data on the immediate, intermediate, and final outcomes in the SMP's theory of change outlined in Table 1, including: program participation; academic behaviors; academic mindsets, perseverance, and learning strategies; school performance; student achievement; and personality traits.[13]

4.2.1 Program participation

The program was implemented as intended (Table A.2). In 2014, on average, each participant 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.

Importantly, however, the average treatment student's exposure to the program decreased over time for three main reasons. First, two students never joined the program in 2014. Second, a large percentage of students was suspended at least once from the program for not complying with the conditions discussed in Section 3.2.3: 36% of the total in 2014, 67% of students still in the program in 2015, and 61% of students still in the program in 2016. Third, a large share of students were expelled from the program, either during the school year (6 students in 2014, 10 students in 2015, and 3 students in 2016) or at the end of the year (5 students in 2014, 26 students in 2015, and 34 students in 2016) for not complying with the conditions. By the end of 2016, only 120 of the original 204 treatment students remained in the program.

[13] SMP and school data for the 2016 school year will be available in January and May of 2017, respectively.

4.2.2 Academic behaviors

We collected data on students' academic behaviors in 2015 and 2016. The survey 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 combined instruments that captured students' academic mindsets, perseverance, and learning strategies, as defined by Farrington et al. (2012). Second, we combined instruments that relied on self-reports and performance assessments, in light of ongoing debates in psychology and education about the perils of relying on the former (see Alan et al. 2016; Borghans et al. 2008; Duckworth and Yeager 2015; West et al. 2015). Third, we prioritized 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, which captures their beliefs about performance and self-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 measures students' perseverance and passion for long-term goals; (b) the Domain-Specific Impulsivity Scale for Children (DSIS-C), a survey of students' self-control; (c) *Caras* (Smileys), an assessment of students' self-control; and (d) Labs, an assessment of students' planning skills. Finally, to measure students' learning strategies, we relied on 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 student absences; (b) their math and language grades; (c) whether they passed their

grade; (d) the number of subjects that they failed;[14] (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.[15] They were designed by psychometricians at 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 own standards, including: (a) the *Núcleos de Aprendizaje Prioritario* (NAPs), the contents that the government has prioritized from the national curriculum; and (b) the publicly-released items from the *Operativo Nacional de Evaluación* (ONE), the national student assessment until 2016. The tests were scored 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

Finally, we administered the Big 5 Inventory to measure students' personality traits in 2016. The survey, widely used among psychologists and economists, assesses five different facets: extraversion, agreeableness, conscientiousness, neuroticism, and openness (John et al. 2008; John and Srivastava 1999). The version of the survey 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 distribution of raw scores on each facet.

4.3 Attrition

We examine whether there was differential attrition across experimental groups by running a regression of the baseline student variables on a dummy for treatment assignment, a dummy for attrition status, and an interaction of these dummies and find evidence of differential attrition on some variables and rounds (Tables A.3-A.4). Therefore, as we explain in Section 5, we test the robustness of our impact estimates to the inclusion of the baseline student covariates.

[14] 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 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.

[15] We 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.

5 Empirical strategy

5.1 Intent-to-treat effects

We first estimate the effect of the *offer* of a spot in the SMP (i.e., the intent-to-treat or ITT). This effect is given by:

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

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 each student was offered a spot in the SMP, X_i is an index of family income at baseline,[16] and ϵ_{ijt} is the idiosyncratic error term. All estimations are conducted with clustered standard errors at the school level. 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 the ITT effects separately by year, which are given by:

$$Y_i = \alpha_j + \beta T_i + \gamma X_i + \epsilon_{ij} \tag{2}$$

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.

5.2 Heterogeneous treatment effects

We also explore whether the offer the SMP differentially impacts girls, students who had previously repeated a grade, and students from low-income families.[17] This effect is given by:

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

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.

[16] This index is the first principal component from a principal component analysis of household assets (dummies for students who 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.

[17] We 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.

5.3 Treatment-on-the-treated effects

We also estimate the effect of *receiving* the SMP (i.e., the treatment-on-the-treated or TOT). This effect is given by the two-stage least squares (2SLS) instrumental variables (IV) model:

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

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.[18] Therefore, an estimation of the dose-response relationship between the number of scholarships and an outcome of interest instrumented by the random assignment does not meet the exclusion restriction because the instrument also affects the outcome through the number of 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 mentoring sessions and the outcome of interest.

6 Results

6.1 Academic behaviors

We find clear evidence that the SMP impacted students' academic behaviors. As Table 4 shows, students in the treatment group reported engaging in preventive and corrective behaviors more often that their control group counterparts in 2015 and 2016. The effects were positive, moderate to large (between .15 and .30σ), and remained statistically significant after the inclusion of baseline covariates. All but one of these effects emerged both in 2015

[18] For example, a student may be suspend 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.

and 2016 and were consistent across years in their sign, magnitude, and statistical significance (Table A.5).

[Insert Table 4 here.]

In 2015, the impact of the program on some indexes (e.g., preventive test behavior, corrective test behavior, corrective failing behavior, and corrective absenteeism behavior) was driven partly by a large share of control group students who did not engage in any or engaged in only one of the prompted behaviors (Figure B.1). In 2016, the impact on some indexes (e.g., corrective failing behavior, preventive absenteeism behavior, and corrective absenteeism behavior) was driven partly by an increase in the share of treatment students who engaged in many of these behaviors (Figure B.2). Interestingly, across both experimental groups, the distribution of raw scores shifted to the left between 2015 and 2016 (i.e., on average, students reported engaging in fewer behaviors), which suggests that the multiple administrations of the survey did not bias all students towards increasingly reporting what they believed to be desirable answers. Yet, it is still possible that treatment students increasingly reported to engage in the behaviors that were taught to them during the mentoring sessions, regardless of whether they actually engaged in them.

We find some evidence that the program differentially impacted boys (Table A.6). Boys in the treatment group were more likely to report corrective test, failing, flunking, and absenteeism behavior (by .22 to .31σ). We do not, however, find heterogeneous effects for students who had previously repeated a grade or students from low-income families.

We also find that for each month that a student received both a scholarship and a mentoring session, he/she improved on average between .02 and .04σ on all but two of the indexes of academic behavior (Table A.7). Both the indexes impacted and the magnitude of the effects are consistent with those from the effects of the offer of the program.

6.2 Academic mindsets, perseverance, and learning strategies

We find very little evidence that the impact of the program on academic behaviors translated into 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 about .12σ). Yet, this pooled effect is mostly driven by a positive and statistically significant effect in 2015, which does not emerge in 2014 or 2016 (Table A.8). We find no statistically significant differences across experimental groups in any of the other surveys or performance assessments that we administered in 2014-2016, but we cannot discard moderate small to moderate positive effects.

[Insert Table 5 here.]

Except in the LASSI motivation sub-index, treatment students performed on par with their control group counterparts at all parts of the distribution (Figures C.1-C.3). There were no heterogeneous effects by gender, prior repetition status, or family income. The TOT effects were consistent with the ITT effects in sign, magnitude, and statistical significance (Table A.9).

6.3 School performance

We find almost no evidence that the program impacted school performance when we estimate the effects across all three years of the study. As Table 6 indicates, we detect a marginally statistically significant reduction in the number of student absences (by 2 absences, over a control group mean of 33). Yet, this reduction loses statistically significance once we account for baseline covariates. There were no heterogeneous effects for any sub-groups.

[Insert Table 6 here.]

The pooled effects, however, mask important differences across years (Table A.10). In 2014, the program achieved a statistically significant improvement in students' language grade (by .18 to .2σ) and a reduction in the number of student absences (by 6 to 7 absences), the share of students who failed a grade (by 5 percentage points), and the number of subjects that students failed and carried over to the next year. In 2015, the sign of most effects remained consistent, but the magnitude decreased and was no longer statistically significant. In 2016, the sign of most effects switched and the increase of grade failure was statistically significant (albeit only marginally).

In part, these differences across years may be due to the large share of students who were expelled from the program (Table A.2).[19] It is possible that the program was improving school performance, but that after nearly half of its participants were expelled, the average treatment group student was no longer outperforming his/her counterpart in the control group.

This explanation is consistent with the TOT effects (Table A.11). We find that each month of the combined treatment produced a statistically significant reduction in the number of student absences and the number of failed subjects, even when pooling effects across years. We also find a statistically significant reduction in the share of students who dropped out of school.[20]

[19]This may also explain the differential attrition by treatment group that we observed when we collected school performance data on that year (Table A.4, Panel D).

[20]Two of these three effects are no longer statistically significant once we account for baseline covariates, but this seems mostly due to insufficient statistical power, since relatively minor changes in the point estimates render the effects no longer detectable at conventional levels of statistical significance.

6.4 Student achievement

We find no evidence that the impact that the program might have had on school performance translated into more student learning, as measured by standardized tests of math and reading. As Table 7 indicates, the pooled effect of the program on student achievement across 2015 and 2016 is negative, but statistically insignificant, on both subjects.

[Insert Table 7 here.]

In fact, the program had a negative and marginally statistically significant impact in math in 2016 (of -.17 to -.16σ), which was robust to the inclusion of baseline covariates (Table A.12). The distribution of control and treatment students differed by year: there was a larger share of treatment students performing above average in math in 2015, but a larger share of control students performing at this level in 2016, and the opposite was true for reading (Figure D.1). There were no heterogeneous effects for any sub-groups. The TOT effects were consistent with the ITT effects in sign, magnitude, and statistical significance (Table A.13).

6.5 Personality traits

Finally, the program had no impact on students' personality traits in 2016. In fact, as Table 8 shows, it had a negative and marginally statistically significant impact on conscientiousness (of .22 to .25σ), which remained after the inclusion of baseline covariates. This was partly due to a higher share of control students scoring higher in this facet (Figure A.1). There were no heterogeneous effects for any sub-groups. The TOT effects were largely consistent with the ITT effects in sign, magnitude, and statistical significance (Table A.14).

[Insert Table 8 here.]

7 Discussion

This paper presents the findings from a three-year randomized evaluation of a program that combined scholarships and non-academic mentoring for secondary school students in the Province of Buenos Aires, Argentina. We were interested in evaluating this program not only because it is one of the longest-standing and largest education initiatives run by a non-profit in the country, but also because it offered an opportunity to shed light on the question of why cash transfers and scholarships have rarely impacted student learning in developing countries. The program did not aim to improve student achievement. However,

its mentoring component addressed a potential reason why cash transfers may have been insufficient to improve learning: the lack of experience of low-income parents with schooling, which may limit their capacity to transfer productive socio-emotional skills to their children. Therefore, by evaluating the combined effect of scholarships and mentoring, we hoped to gain some insights into the importance of this aspect of the lives of low-income students.

We designed our data collection strategy purposefully to measure the impact of the program at every step of our hypothesized theory of chain, from immediate (e.g., academic behaviors), to intermediate (e.g., academic perseverance and mindsets), and final outcomes (e.g., school performance and student achievement). We sought to understand whether the program worked as we expected and to identify any breakdowns in our proposed causal pathway.

We found that the program was implemented as intended, providing (roughly) one scholarship and one mentoring session for each month of the school year. The program increased students' propensity to engage in plausibly productive academic behaviors (e.g., starting to study early for an exam or catching up on schoolwork missed due to absences). However, we found very little evidence that it improved academic mindsets (e.g., self-beliefs about performance and self-efficacy), perseverance (e.g., grit), and learning strategies (e.g., metacognition). It improved some metrics of school performance (e.g., language grades, student absenteeism, grade failure, and the number of failed subjects) on its first year. Yet, nearly half of its beneficiaries were expelled by the third year of the study and the remaining students did not see similar improvements in school performance in subsequent years. We found no evidence that the program positively impacted math or reading achievement or personality traits.

Our study contributes to the literature on the impact of cash transfers. We make two main advances over existing work. First, we are among the first to evaluate the impact of this type of program on socio-emotional skills.[21] Second, we examine the complementarity between scholarships, which seek to alleviate financial constraints to schooling, and non-academic mentoring, which target other constraints affecting students from low-income families.

Our results also build on the rapidly growing literature on socio-emotional skills in developing countries. Although there have recently been several evaluations of interventions targeting specific skills (see Alan et al. 2016; Outes et al. 2017), ours is one of the first studies to shed light on how these skills relate to one another.[22] Specifically, we document that impacts on narrowly-defined academic behaviors may not necessarily translate into broader changes in academic mindsets, perseverance, or learning strategies, let alone personality traits.

Finally, we also contribute to the ongoing debate on the disconnect between schooling and learning in the developing world. The bulk of the evidence on this question has focused on

[21] To our knowledge, there is only one other such study in a developing country (Barrera-Osorio et al. 2017).

[22] This question has received far more attention in developed countries (see, for example, Heller et al. 2017; Jackson 2017; Kraft 2017; Kraft and Blazar 2017; West et al. 2015).

lower-middle-income countries, where this disconnect stems from students lacking the requisite basic skills to keep up with curricular expectations and classroom instruction caters to the top of the achievement distribution (see, for example, Muralidharan et al. 2016; Muralidharan and Zieleniak 2014; Pritchett and Beatty 2015). We show that, in upper-middle-income countries like Argentina, students do not learn even when they meet grade-level expectations. This suggests that, in these contexts, curricular expectations may be too low and/or instruction may focus on aspects that do not contribute to academic skills.

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 Desarollo (BID).

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

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

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

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

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., D. Filmer, and A. de Barros (2017). From schooling to young adulthood: Evidence from a 9-year follow-up in rural Cambodia. *Unpublished manuscript*. Cambridge, MA: Harvard Graduate School of Education.

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.

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

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 Unviersity 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., 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.

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.

Jackson, C. K. (2017). What do test scores miss? The importance of teacher effects on non-test score outcomes. *Journal of Political Economy*.

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.

Kraft, M. A. (2017). Teacher effects on complex cognitive skills and social-emotional competencies. *Journal of Human Resources*.

Kraft, M. A. and D. Blazar (2017). Teacher and teaching effects on students' attitudes and behaviors. *Educational Evaluation and Policy Analysis 39*(1), 146–170.

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

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.

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. *Unpublished manuscript*. Delhi, India: Abdul Latif Jameel Poverty Action Lab (J-PAL).

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.

Table 1: SMP's theory of change

Problem	Inputs/Activities	Outputs	Immediate outcomes	Intermediate outcomes	Final 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	Participants	Location
2014			
• School year starts	Feb		
• Student survey	May 14-26	100% sample	School
• Household survey		81% sample	School
		19% sample	Phone
• Lottery is conducted	Jun		
• Surveys of academic mindsets,	Nov 10-Dec 4	80% sample	School
perseverance, and learning strategies	Dec 18-Jan 16	17% sample	Home
2015			
• SMP data for 2014	Jan	100% treatment group	
• School year starts	Feb		
• School performance data for 2014	May	100% sample	
• Math and reading tests	Jun 22-Jul 6	63% sample	School
	Jul 13-Aug 12	26% sample	Home
• Surveys of academic mindsets,	Oct 14-Nov 6	66% sample	School
perseverance, and learning strategies	Nov 3-Dec 1	24% sample	Home
• Survey of academic behaviors			
2016			
• SMP data for 2015	Jan	94% treatment group	
• School year starts	Feb		
• School performance data for 2015	May	86% sample	
• Math and reading tests	May 9-21	64% sample	School
	May 30-Jun 21	28% sample	Home
• Surveys of academic mindsets,	Sep 19-30	61% sample	School
perseverance, and learning strategies	Oct 7-29	23% sample	Home
• Survey of academic behaviors			
• Survey of personality traits			
2017			
• SMP data for 2016	Jan	75% treatment group	
• School year starts	Feb		
• School performance data for 2016	May	90% sample	

Notes: (1) The surveys of academic mindsets, perseverance, and learning strategies and the survey of academic behaviors were administered on the same date and location in 2015. (2) The surveys of academic mindsets, perseverance, and learning strategies, the survey of academic behaviors and the survey of personality traits were administered on the same date and location in 2016. (3) SMP and school data are collected directly from the foundation and schools, respectively, so the table does not specify the location of data collection. (4) SMP data are only available for treatment students (control students did not participate in the program).

Table 3: Balancing checks (baseline)

Variable	All (1)	Control (2)	Treatment (3)	Difference (4)	N (5)
Panel A. Student survey					
Argentine	.977	.98	.975	-.005	397
	(.149)	(.141)	(.157)	(.02)	
Female	.52	.544	.495	-.049	408
	(.5)	(.499)	(.501)	(.051)	
Age	12.435	12.502	12.368	-.131	407
	(1.062)	(1.153)	(.961)	(.11)	
Attends morning shift	.578	.583	.574	-.008	408
	(.494)	(.494)	(.496)	(.045)	
Previously repeated grade(s)	.309	.322	.297	-.024	404
	(.463)	(.468)	(.458)	(.044)	
Previously dropped out of school	.05	.073	.027	-.047*	360
	(.218)	(.262)	(.163)	(.022)	
Panel B. Household survey					
Has car	.21	.163	.256	.096***	405
	(.408)	(.371)	(.438)	(.026)	
Has natural gas	.298	.269	.327	.064*	403
	(.458)	(.444)	(.47)	(.034)	
Has running water	.825	.805	.846	.051	401
	(.38)	(.397)	(.362)	(.047)	
Has in-house bathroom	.824	.809	.838	.03	408
	(.382)	(.394)	(.369)	(.045)	
Has solid floor	.985	.98	.99	.01	408
	(.121)	(.139)	(.099)	(.006)	
Has fridge	.72	.677	.764	.087**	404
	(.449)	(.469)	(.426)	(.028)	
Has computer	.545	.547	.542	-.002	404
	(.499)	(.499)	(.499)	(.026)	
Has Internet	.386	.383	.389	.01	404
	(.487)	(.487)	(.489)	(.036)	
Has cell phone	.913	.891	.936	.045	404
	(.282)	(.313)	(.245)	(.029)	
Parent is homeowner	.598	.564	.632	.07*	408
	(.491)	(.497)	(.483)	(.033)	
F-statistic	1.455				
p-value	.115				

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 4: ITT effects on academic behaviors (2015-2016)

	Control	Effect size	
	(1)	(2)	(3)
Proactive school behavior (std.)	0	.158	.146
	(1)	(.091)	(.087)
	180	707	707
Preventive homework behavior (std.)	0	.222**	.207**
	(1)	(.088)	(.077)
	180	707	707
Corrective homework behavior (std.)	0	.223**	.224**
	(1)	(.078)	(.077)
	180	707	707
Preventive test behavior (std.)	0	.193***	.195***
	(1)	(.04)	(.036)
	180	707	707
Corrective test behavior (std.)	0	.204***	.207***
	(1)	(.054)	(.053)
	180	707	707
Corrective failing behavior (std.)	0	.299***	.297***
	(1)	(.051)	(.047)
	180	707	707
Corrective flunking behavior (std.)	0	.154**	.159**
	(1)	(.057)	(.054)
	180	707	707
Preventive absenteeism behavior (std.)	0	.251***	.256***
	(1)	(.063)	(.067)
	180	707	707
Corrective absenteeism behavior (std.)	0	.279***	.278***
	(1)	(.063)	(.058)
	180	707	707
Corrective free period behavior (std.)	0	.021	.015
	(1)	(.093)	(.089)
	180	707	707
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.

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

	Control	Effect size	
	(1)	(2)	(3)
Grit (std.)	0	.052	.065
	(1)	(.053)	(.056)
	193	1102	1102
DSISC (std.)	0	.093	.1
	(1)	(.073)	(.075)
	193	1102	1102
Self-beliefs (std.)	0	.073	.079
	(1)	(.089)	(.087)
	193	1102	1102
LASSI - Organization and Planning (std.)	0	.028	.049
	(1)	(.083)	(.084)
	193	1102	1102
LASSI - Motivation (std.)	0	.124*	.132*
	(1)	(.065)	(.065)
	193	1102	1102
CARAS - Reflexivity Index (std.)	0	.002	.01
	(1)	(.074)	(.076)
	193	1094	1094
LABS (std.)	0	-.054	-.054
	(1)	(.044)	(.047)
	193	1102	1102
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.

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

	Control	Effect size	
	(1)	(2)	(3)
Language grade (std.)	0	.028	.019
	(1)	(.086)	(.093)
	203	1012	1012
Math grade (std.)	0	.009	.014
	(1)	(.054)	(.059)
	203	1011	1011
No. of pending subjects	1.516	-.332	-.319
	(2.511)	(.182)	(.199)
	190	980	980
No. of absences	33.4	-2.643*	-2.292
	(28.056)	(1.427)	(1.527)
	173	944	944
Failed grade	.148	-.015	-.016
	(.356)	(.018)	(.02)
	203	1057	1057
Transferred schools	.025	-.019	-.017
	(.155)	(.013)	(.014)
	203	1057	1057
Dropped out of school	.054	-.025	-.025
	(.227)	(.014)	(.014)
	203	1057	1057
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.

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

	Control	Effect size	
	(1)	(2)	(3)
Math achievement (std.)	0	-.055	-.081
	(1)	(.093)	(.091)
	164	683	683
Reading achievement (std.)	0	-.008	-.03
	(1)	(.066)	(.073)
	173	706	706
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.

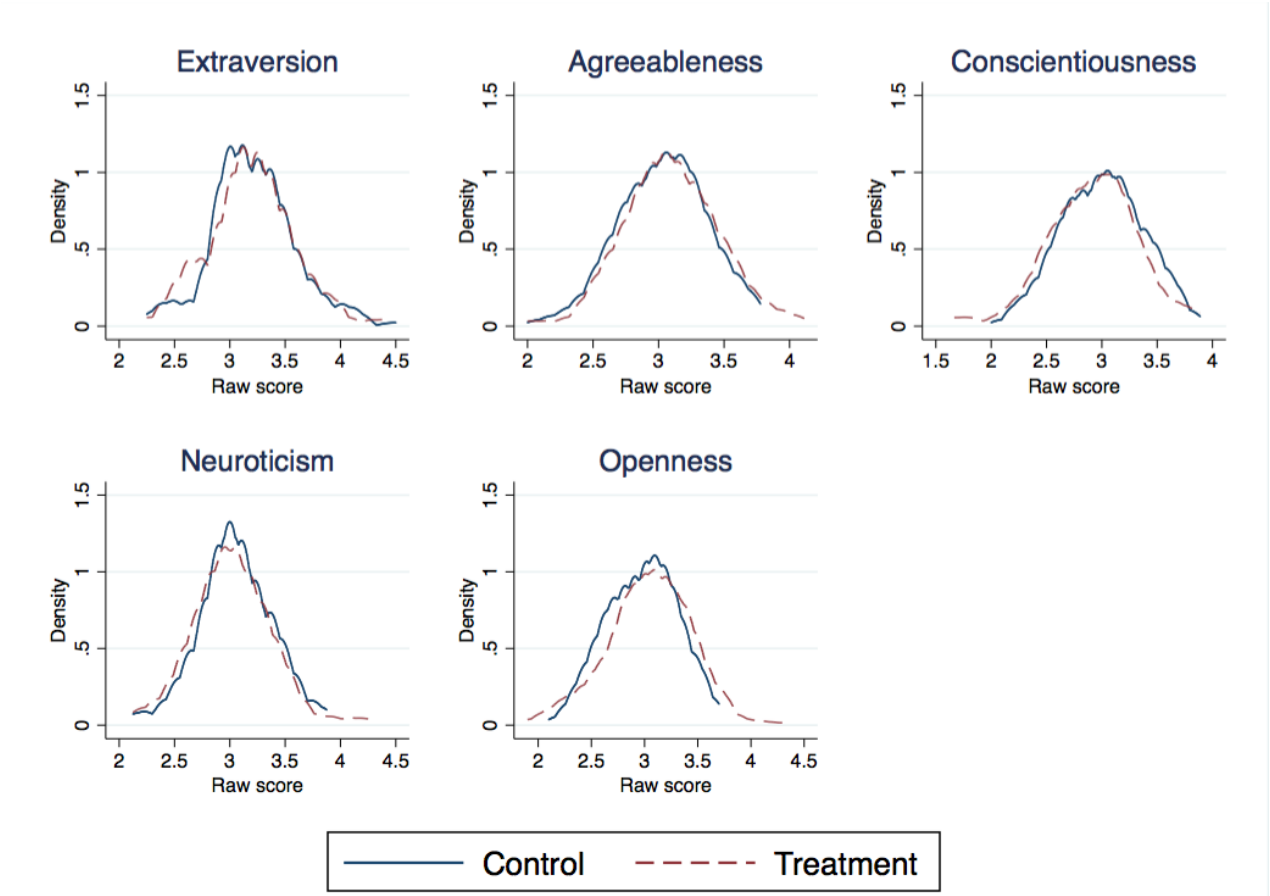
Table 8: ITT effects on personality traits (2016)

Index	Control	Effect size	
	(1)	(2)	(3)
Extraversion (std.)	0	-.055	-.069
	(1)	(.09)	(.089)
	165	341	341
Agreeableness (std.)	0	.182	.168
	(1)	(.116)	(.126)
	165	341	341
Conscientiousness (std.)	0	-.245*	-.217*
	(1)	(.111)	(.112)
	165	341	341
Neuroticism (std.)	0	-.104	-.106
	(1)	(.087)	(.084)
	165	341	341
Openness (std.)	0	.186	.17
	(1)	(.106)	(.102)
	165	341	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.

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)

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

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

Table A.2: Program participation (2014-2016)

Variable	2014 Treatment (1)	N (2)	2015 Treatment (3)	N (4)	2016 Treatment (5)	N (6)
Scholarships received	7.51	204	7.817	191	6.516	155
	(3.023)		(3.347)		(3.454)	
Intended mentoring sessions	9.093	204	8.77	191	7.387	155
	(1.025)		(2.902)		(2.964)	
Actual mentoring sessions	7.819	204	7.487	191	6.348	155
	(1.782)		(3.291)		(3.38)	
Individual mentoring sessions	7.245	204	8.152	191	6.613	155
	(1.912)		(2.723)		(2.688)	
Group mentoring sessions	1.848	204	.618	191	.774	155
	(1.503)		(.707)		(.865)	
Sessions that had to be rescheduled	.24	204	.565	191	.477	155
	(.558)		(1.069)		(.907)	
Sessions to which parent was invited	5.858	204	7.157	191	5.123	155
	(2.295)		(2.56)		(2.642)	
Sessions to which parent attended	5.49	204	4.738	191	2.877	155
	(2.412)		(2.758)		(2.142)	
Sessions that used required materials	6.26	204	5.665	191	2.8	155
	(2.342)		(3.136)		(3.105)	
Mentors per student	1.191	204	1.099	191	1.077	155
	(.394)		(.3)		(.268)	
Student never joined	.01	204	0	191	0	155
	(.099)		(0)		(0)	
Student was suspended	.363	204	.67	191	.606	155
	(.773)		(1.21)		(1.198)	
Student was expelled during the year	.029	204	.052	191	.019	155
	(.169)		(.223)		(.138)	
Student was expelled at the end of the year	.025	204	.136	191	.219	155
	(.155)		(.344)		(.415)	

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

Table A.3: Attrition checks, by round of data collection

	Constant (1)	Treatment (2)	Attritor (3)	Interaction (4)	N
Panel A. Round 1, 2014					
Argentine	.98***	0	-.074	-.899***	397
	(.01)	(.016)	(.079)	(.079)	
Female	.544***	-.063	-.273	.214	408
	(.03)	(.053)	(.155)	(.162)	
Age	12.707***	-.112	.522	.465	407
	(.06)	(.113)	(.415)	(.411)	
Attends morning shift	.425***	-.014	-.116	.037	408
	(.023)	(.044)	(.121)	(.119)	
Previously repeated grade(s)	.441***	-.027	.053	.573***	404
	(.023)	(.045)	(.128)	(.128)	
Previously dropped out of school	.056***	-.033	.227	-.347*	360
	(.016)	(.022)	(.183)	(.181)	
Panel B. Round 2, 2015					
Argentine	.973***	0	.021*	.002	369
	(.009)	(.017)	(.01)	(.018)	
Female	.604***	-.069	-.359	-.061	380
	(.029)	(.057)	(.228)	(.238)	
Age	12.491***	-.097	1.749	-.889	379
	(.071)	(.12)	(1.027)	(1.269)	
Attends morning shift	.434***	-.012	.239	-.406	380
	(.025)	(.049)	(.218)	(.251)	
Previously repeated grade(s)	.42***	-.037	.104	.198	377
	(.023)	(.044)	(.23)	(.371)	
Previously dropped out of school	.023*	-.042*	-.073*	.152*	337
	(.012)	(.021)	(.034)	(.08)	
Panel C. Round 3, 2015					
Argentine	.98***	0	-.024	-.057	397
	(.013)	(.018)	(.05)	(.112)	
Female	.593***	-.094	-.428***	.36**	408
	(.032)	(.054)	(.07)	(.127)	
Age	12.662***	-.092	.536*	-.291	407
	(.057)	(.115)	(.241)	(.372)	
Attends morning shift	.441***	-.019	-.167	.071	408
	(.03)	(.052)	(.113)	(.141)	
Previously repeated grade(s)	.439***	-.033	.065	.121	404
	(.022)	(.042)	(.093)	(.196)	
Previously dropped out of school	.051***	-.033*	.128	-.112	360
	(.012)	(.015)	(.077)	(.135)	

Notes: (1) The table examines whether there is differential attrition by experimental group. Column 4 shows the interaction between the treatment and attrition dummies for each variable collected at baseline. (2) * significant at 10%; ** significant at 5%; *** significant at 1%. (3) Standard errors in column 4 are clustered at the school level.

Table A.4: Attrition checks, by round of data collection

	Constant (1)	Treatment (2)	Attritor (3)	Interaction (4)	N
Panel D. Round 4, 2016					
Argentine	.969***	.014	.035**	-.143*	397
	(.01)	(.017)	(.015)	(.076)	
Female	.536***	-.058	-.071	.053	408
	(.034)	(.059)	(.075)	(.093)	
Age	12.576***	-.011	1.084***	-.669	407
	(.06)	(.131)	(.208)	(.397)	
Attends morning shift	.463***	-.046	-.297***	.222*	408
	(.027)	(.056)	(.075)	(.119)	
Previously repeated grade(s)	.395***	.005	.36***	-.142	404
	(.025)	(.048)	(.091)	(.192)	
Previously dropped out of school	.023	-.001	.282**	-.353**	360
	(.023)	(.026)	(.122)	(.143)	
Panel E. Round 5, 2016					
Argentine	.974***	0	.03***	0	369
	(.009)	(.017)	(.009)	(.017)	
Female	.587***	-.072	.346***	-.928***	380
	(.032)	(.058)	(.029)	(.058)	
Age	12.553***	-.098	.707***	1.098***	379
	(.063)	(.115)	(.059)	(.115)	
Attends morning shift	.442***	-.024	-.424***	.024	380
	(.028)	(.05)	(.025)	(.05)	
Previously repeated grade(s)	.425***	-.028	.751***	.028	377
	(.023)	(.042)	(.021)	(.042)	
Previously dropped out of school	.02	-.037*	.02	0	337
	(.011)	(.02)	(.011)	(0)	
Panel F. Round 6, 2016					
Argentine	.977***	.002	-.003	-.049	397
	(.012)	(.017)	(.027)	(.089)	
Female	.573***	-.074	-.219***	.096	408
	(.031)	(.055)	(.061)	(.067)	
Age	12.632***	-.065	.466*	-.296	407
	(.063)	(.109)	(.222)	(.25)	
Attends morning shift	.417***	.003	-.02	-.089	408
	(.04)	(.058)	(.119)	(.184)	
Previously repeated grade(s)	.433***	-.034	.072	.095	404
	(.028)	(.046)	(.091)	(.116)	
Previously dropped out of school	.065***	-.04**	.037	-.035	360
	(.01)	(.015)	(.049)	(.063)	

Notes: (1) The table examines whether there is differential attrition by experimental group. Column 4 shows the interaction between the treatment and attrition dummies for each variable collected at baseline. (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: ITT effects on academic behaviors, by year (2015-2016)

	2015			2016		
	Control	Effect size		Control	Effect size	
	(1)	(2)	(3)	(4)	(5)	(6)
Proactive school behavior (std.)	0	.115	.11	-.217	.208*	.187*
	(1)	(.161)	(.159)	(.816)	(.1)	(.094)
	180	366	366	165	341	341
Preventive homework behavior (std.)	0	.235*	.228*	-.468	.216**	.191**
	(1)	(.125)	(.119)	(.867)	(.087)	(.084)
	180	366	366	165	341	341
Corrective homework behavior (std.)	0	.205**	.212**	-.347	.253**	.247**
	(1)	(.081)	(.082)	(.81)	(.102)	(.107)
	180	366	366	165	341	341
Preventive test behavior (std.)	0	.209**	.22**	-.413	.186***	.177***
	(1)	(.07)	(.071)	(.768)	(.048)	(.048)
	180	366	366	165	341	341
Corrective test behavior (std.)	0	.216**	.22**	-.37	.198**	.199**
	(1)	(.081)	(.08)	(.749)	(.068)	(.07)
	180	366	366	165	341	341
Corrective failing behavior (std.)	0	.263**	.265***	-.292	.348***	.343***
	(1)	(.084)	(.078)	(.695)	(.046)	(.041)
	180	366	366	165	341	341
Corrective flunking behavior (std.)	0	.12	.132	-.286	.2***	.199***
	(1)	(.087)	(.081)	(.729)	(.047)	(.048)
	180	366	366	165	341	341
Preventive absenteeism behavior (std.)	0	.176*	.189*	-.418	.339***	.337***
	(1)	(.096)	(.098)	(.862)	(.061)	(.069)
	180	366	366	165	341	341
Corrective absenteeism behavior (std.)	0	.255**	.26**	-.342	.313***	.307***
	(1)	(.089)	(.082)	(.74)	(.062)	(.058)
	180	366	366	165	341	341
Corrective free period behavior (std.)	0	.039	.036	-.277	.005	-.003
	(1)	(.11)	(.112)	(.889)	(.121)	(.111)
	180	366	366	165	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.6: ITT effects on academic behaviors by gender (2015-2016)

	Constant (1)	Treatment (2)	Female (3)	Interaction (4)	N (5)
Proactive school behavior (std.)	-.188	.118	-.014	.082	707
	(.128)	(.133)	(.076)	(.104)	
Preventive homework behavior (std.)	-.497***	.206	.033	.053	707
	(.089)	(.122)	(.087)	(.135)	
Corrective homework behavior (std.)	-.309***	.25*	-.05	-.042	707
	(.093)	(.111)	(.073)	(.107)	
Preventive test behavior (std.)	-.45***	.23**	.045	-.052	707
	(.083)	(.085)	(.094)	(.121)	
Corrective test behavior (std.)	-.425***	.331***	.084	-.216**	707
	(.078)	(.069)	(.091)	(.08)	
Corrective failing behavior (std.)	-.349***	.443***	.131	-.246***	707
	(.09)	(.082)	(.135)	(.075)	
Corrective flunking behavior (std.)	-.331***	.326***	.111	-.305**	707
	(.08)	(.083)	(.117)	(.107)	
Preventive absenteeism behavior (std.)	-.537***	.408***	.273**	-.251	707
	(.106)	(.109)	(.09)	(.151)	
Corrective absenteeism behavior (std.)	-.45***	.441***	.207*	-.276**	707
	(.093)	(.063)	(.092)	(.087)	
Corrective free period behavior (std.)	-.288**	.16	.007	-.271	707
	(.089)	(.093)	(.161)	(.212)	

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.7: TOT effect on academic behaviors (2015-2016)

	Control	Scholarships		Mentoring		Combined	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Proactive school behavior (std.)	0	.023*	.022*	.024*	.022*	.025*	.023*
	(1)	(.012)	(.012)	(.013)	(.012)	(.013)	(.013)
	180	707	707	707	707	707	707
Preventive homework behavior (std.)	0	.033***	.031***	.034***	.032***	.035***	.033***
	(1)	(.012)	(.01)	(.012)	(.011)	(.013)	(.011)
	180	707	707	707	707	707	707
Corrective homework behavior (std.)	0	.033***	.033***	.034***	.034***	.035***	.036***
	(1)	(.01)	(.01)	(.01)	(.01)	(.011)	(.011)
	180	707	707	707	707	707	707
Preventive test behavior (std.)	0	.028***	.029***	.029***	.03***	.031***	.031***
	(1)	(.006)	(.005)	(.006)	(.005)	(.006)	(.005)
	180	707	707	707	707	707	707
Corrective test behavior (std.)	0	.03***	.031***	.031***	.032***	.032***	.033***
	(1)	(.008)	(.008)	(.008)	(.008)	(.008)	(.008)
	180	707	707	707	707	707	707
Corrective failing behavior (std.)	0	.044***	.044***	.045***	.045***	.047***	.047***
	(1)	(.007)	(.006)	(.007)	(.006)	(.008)	(.007)
	180	707	707	707	707	707	707
Corrective flunking behavior (std.)	0	.023***	.024***	.023***	.024***	.024***	.025***
	(1)	(.008)	(.008)	(.008)	(.008)	(.009)	(.008)
	180	707	707	707	707	707	707
Preventive absenteeism behavior (std.)	0	.037***	.038***	.038***	.039***	.04***	.041***
	(1)	(.008)	(.009)	(.009)	(.01)	(.009)	(.01)
	180	707	707	707	707	707	707
Corrective absenteeism behavior (std.)	0	.041***	.041***	.042***	.043***	.044***	.044***
	(1)	(.008)	(.007)	(.008)	(.008)	(.008)	(.008)
	180	707	707	707	707	707	707
Corrective free period behavior (std.)	0	.003	.002	.003	.002	.003	.002
	(1)	(.013)	(.012)	(.013)	(.013)	(.014)	(.013)
	180	707	707	707	707	707	707
School FE?	Y	Y	Y	Y	Y	Y	Y
Year FE?	Y	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 in 2014 (column 1) and the dose-response relationship or TOT effect without (columns 2, 4 and 6) and with covariates (columns 3, 5 and 7), pooled across 2014-2016. (2) * significant at 10%; ** significant at 5%; *** significant at 1%. (3) Standard errors in columns 2-7 are clustered at the school level.

Table A.8: ITT effects on academic mindsets, perseverance, and learning strategies, by year (2014-2016)

	2014			2015			2016		
	Control	Effect size		Control	Effect size		Control	Effect size	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Grit (std.)	0	.079	.091	0	.115	.133	0	-.053	-.044
	(1)	(.073)	(.072)	(1)	(.088)	(.089)	(1)	(.093)	(.098)
	193	395	395	180	366	366	165	341	341
DSISC (std.)	0	.099	.123	0	.133	.136*	0	.04	.027
	(1)	(.099)	(.106)	(1)	(.076)	(.073)	(1)	(.082)	(.085)
	193	395	395	180	366	366	165	341	341
Self-beliefs (std.)	0	.042	.055	0	.114	.121	0	.061	.052
	(1)	(.102)	(.104)	(1)	(.085)	(.085)	(1)	(.104)	(.1)
	193	395	395	180	366	366	165	341	341
LASSI - Organization and Planning (std.)	0	.027	.051	0	.03	.06	0	.024	.029
	(1)	(.101)	(.099)	(1)	(.079)	(.085)	(1)	(.112)	(.113)
	193	395	395	180	366	366	165	341	341
LASSI - Motivation (std.)	0	.153	.157	0	.173**	.19**	0	.034	.037
	(1)	(.128)	(.128)	(1)	(.074)	(.077)	(1)	(.056)	(.055)
	193	395	395	180	366	366	165	341	341
CARAS - Reflexivity Index (std.)	0	-.009	-.001	0	.044	.055	0	-.036	-.031
	(1)	(.082)	(.078)	(1)	(.082)	(.089)	(1)	(.1)	(.104)
	193	394	394	176	360	360	164	340	340
LABS (std.)	0	-.014	-.012	0	-.114	-.113	0	-.024	-.028
	(1)	(.066)	(.074)	(1)	(.109)	(.121)	(1)	(.055)	(.049)
	193	395	395	180	366	366	165	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.9: TOT effect on academic mindsets, perseverance, and learning strategies (2014-2016)

	Control	Scholarships		Mentoring		Combined	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Grit (std.)	0	.007	.009	.007	.009	.008	.01
	(1)	(.007)	(.007)	(.007)	(.008)	(.008)	(.008)
	193	1102	1102	1102	1102	1102	1102
DSISC (std.)	0	.013	.014	.013	.014	.014	.015
	(1)	(.009)	(.01)	(.01)	(.01)	(.01)	(.011)
	193	1102	1102	1102	1102	1102	1102
Self-beliefs (std.)	0	.01	.011	.01	.011	.011	.012
	(1)	(.011)	(.011)	(.012)	(.012)	(.013)	(.012)
	193	1102	1102	1102	1102	1102	1102
LASSI - Organization and Planning (std.)	0	.004	.007	.004	.007	.004	.007
	(1)	(.011)	(.011)	(.011)	(.011)	(.012)	(.012)
	193	1102	1102	1102	1102	1102	1102
LASSI - Motivation (std.)	0	.017*	.018**	.018**	.019**	.018**	.019**
	(1)	(.009)	(.009)	(.009)	(.009)	(.009)	(.009)
	193	1102	1102	1102	1102	1102	1102
CARAS - Reflexivity Index (std.)	0	0	.001	0	.001	0	.002
	(1)	(.009)	(.01)	(.01)	(.01)	(.01)	(.011)
	193	1094	1094	1094	1094	1094	1094
LABS (std.)	0	-.007	-.007	-.008	-.008	-.008	-.008
	(1)	(.006)	(.006)	(.006)	(.006)	(.006)	(.007)
	193	1102	1102	1102	1102	1102	1102
School FE?	Y	Y	Y	Y	Y	Y	Y
Year FE?	Y	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 in 2014 (column 1) and the dose-response relationship or TOT effect without (columns 2, 4 and 6) and with covariates (columns 3, 5 and 7), pooled across 2014-2016. (2) * significant at 10%; ** significant at 5%; *** significant at 1%. (3) Standard errors in columns 2-7 are clustered at the school level.

Table A.10: ITT effects on school performance, by year (2014-2016)

	2014			2015			2016		
	Control	Effect size		Control	Effect size		Control	Effect size	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Language grade (std.)	0	.195*	.196*	0	-.039	-.062	0	-.156	-.168
	(1)	(.099)	(.099)	(1)	(.092)	(.104)	(1)	(.108)	(.121)
	203	406	406	159	331	331	127	275	275
Math grade (std.)	0	.103	.112	0	-.005	.004	0	-.108	-.116
	(1)	(.097)	(.1)	(1)	(.077)	(.082)	(1)	(.101)	(.111)
	203	406	406	158	331	331	127	274	274
No. of pending subjects	1.516	-.494**	-.492**	2.688	-.307	-.273	.912	-.145	-.146
	(2.511)	(.191)	(.191)	(3.717)	(.433)	(.485)	(.907)	(.095)	(.097)
	190	386	386	157	325	325	125	269	269
No. of absences	33.4	-7.098**	-6.92**	35.582	-.443	.143	35.607	.395	.832
	(28.056)	(2.337)	(2.334)	(26.657)	(2.047)	(2.197)	(23.622)	(2.232)	(2.54)
	173	346	346	152	321	321	128	277	277
Failed grade	.148	-.06**	-.06**	.253	-.033	-.034	.162	.062*	.059
	(.356)	(.022)	(.022)	(.436)	(.035)	(.037)	(.37)	(.032)	(.039)
	203	406	406	170	348	348	148	303	303
Transferred schools	.025	-.01	-.01	.088	-.012	-.014	.108	-.038	-.028
	(.155)	(.012)	(.014)	(.284)	(.039)	(.042)	(.312)	(.027)	(.028)
	203	406	406	170	348	348	148	303	303
Dropped out of school	.054	-.026	-.026	.024	-.013	-.011	.074	-.038	-.04
	(.227)	(.021)	(.021)	(.152)	(.019)	(.019)	(.263)	(.026)	(.026)
	203	406	406	170	348	348	148	303	303
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, 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.11: TOT effect on school performance (2014-2016)

	Control	Scholarships		Mentoring		Combined	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Language grade (std.)	0	.003	.002	.004	.003	.004	.003
	(1)	(.01)	(.011)	(.011)	(.012)	(.011)	(.012)
	203	1012	1012	1012	1012	1012	1012
Math grade (std.)	0	.001	.002	.001	.002	.001	.002
	(1)	(.007)	(.007)	(.007)	(.008)	(.007)	(.008)
	203	1011	1011	1011	1011	1011	1011
No. of pending subjects	1.516	-.042*	-.04*	-.044*	-.043*	-.045*	-.044*
	(2.511)	(.022)	(.024)	(.023)	(.025)	(.024)	(.026)
	190	980	980	980	980	980	980
No. of absences	33.4	-.335**	-.292	-.357**	-.311	-.367**	-.321
	(28.056)	(.17)	(.182)	(.178)	(.191)	(.184)	(.197)
	173	944	944	944	944	944	944
Failed grade	.148	-.002	-.002	-.002	-.002	-.002	-.002
	(.356)	(.002)	(.002)	(.002)	(.003)	(.002)	(.003)
	203	1057	1057	1057	1057	1057	1057
Transferred schools	.025	-.003	-.002	-.003	-.002	-.003	-.002
	(.155)	(.002)	(.002)	(.002)	(.002)	(.002)	(.002)
	203	1057	1057	1057	1057	1057	1057
Dropped out of school	.054	-.003	-.003	-.004**	-.003	-.004**	-.004**
	(.227)	(.002)	(.002)	(.002)	(.002)	(.002)	(.002)
	203	1057	1057	1057	1057	1057	1057
School FE?	Y	Y	Y	Y	Y	Y	Y
Year FE?	Y	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 in 2014 (column 1) and the dose-response relationship or TOT effect without (columns 2, 4 and 6) and with covariates (columns 3, 5 and 7), pooled across 2014-2016. (2) * significant at 10%; ** significant at 5%; *** significant at 1%. (3) Standard errors in columns 2-7 are clustered at the school level.

Table A.12: ITT effects on student achievement, by year (2015-2016)

	2015			2016		
	Control	Effect size		Control	Effect size	
	(1)	(2)	(3)	(4)	(5)	(6)
Math achievement (std.)	0	.054	.012	.007	-.161*	-.174*
	(1)	(.14)	(.13)	(.927)	(.081)	(.086)
	164	335	335	168	348	348
Reading achievement (std.)	0	-.067	-.076	-.158	.05	.017
	(1)	(.085)	(.092)	(1.094)	(.099)	(.1)
	173	349	349	173	357	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.13: TOT effect on student achievement (2015-2016)

	Control	Scholarships		Mentoring		Combined	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Math achievement (std.)	0	-.008	-.012	-.008	-.012	-.009	-.013
	(1)	(.013)	(.013)	(.013)	(.013)	(.014)	(.014)
	164	683	683	683	683	683	683
Reading achievement (std.)	0	-.001	-.004	-.001	-.005	-.001	-.005
	(1)	(.009)	(.01)	(.01)	(.01)	(.01)	(.011)
	173	706	706	706	706	706	706
School FE?	Y	Y	Y	Y	Y	Y	Y
Year FE?	Y	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 in 2014 (column 1) and the dose-response relationship or TOT effect without (columns 2, 4 and 6) and with covariates (columns 3, 5 and 7), pooled across 2014-2016. (2) * significant at 10%; ** significant at 5%; *** significant at 1%. (3) Standard errors in columns 2-7 are clustered at the school level.

Table A.14: TOT effect on personality traits (2016)

	Control	Scholarships		Mentoring		Combined	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Extraversion (std.)	0	-.01	-.012	-.01	-.013	-.01	-.013
	(1)	(.015)	(.015)	(.015)	(.015)	(.016)	(.016)
	165	341	341	341	341	341	341
Agreeableness (std.)	0	.032*	.03	.033*	.031	.034*	.032
	(1)	(.018)	(.02)	(.019)	(.021)	(.019)	(.022)
	165	341	341	341	341	341	341
Conscientiousness (std.)	0	-.043**	-.038**	-.044**	-.039**	-.046**	-.041**
	(1)	(.018)	(.018)	(.019)	(.019)	(.019)	(.02)
	165	341	341	341	341	341	341
Neuroticism (std.)	0	-.018	-.019	-.019	-.019	-.02	-.02
	(1)	(.014)	(.014)	(.014)	(.014)	(.015)	(.015)
	165	341	341	341	341	341	341
Openness (std.)	0	.033*	.03*	.033*	.031*	.035*	.032*
	(1)	(.019)	(.018)	(.019)	(.018)	(.02)	(.019)
	165	341	341	341	341	341	341
School FE?	Y	Y	Y	Y	Y	Y	Y
Year FE?	Y	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 in 2014 (column 1) and the dose-response relationship or TOT effect without (columns 2, 4 and 6) and with covariates (columns 3, 5 and 7), pooled across 2014-2016. (2) * significant at 10%; ** significant at 5%; *** significant at 1%. (3) Standard errors in columns 2-7 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;[23] (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;[24] (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"[25] 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

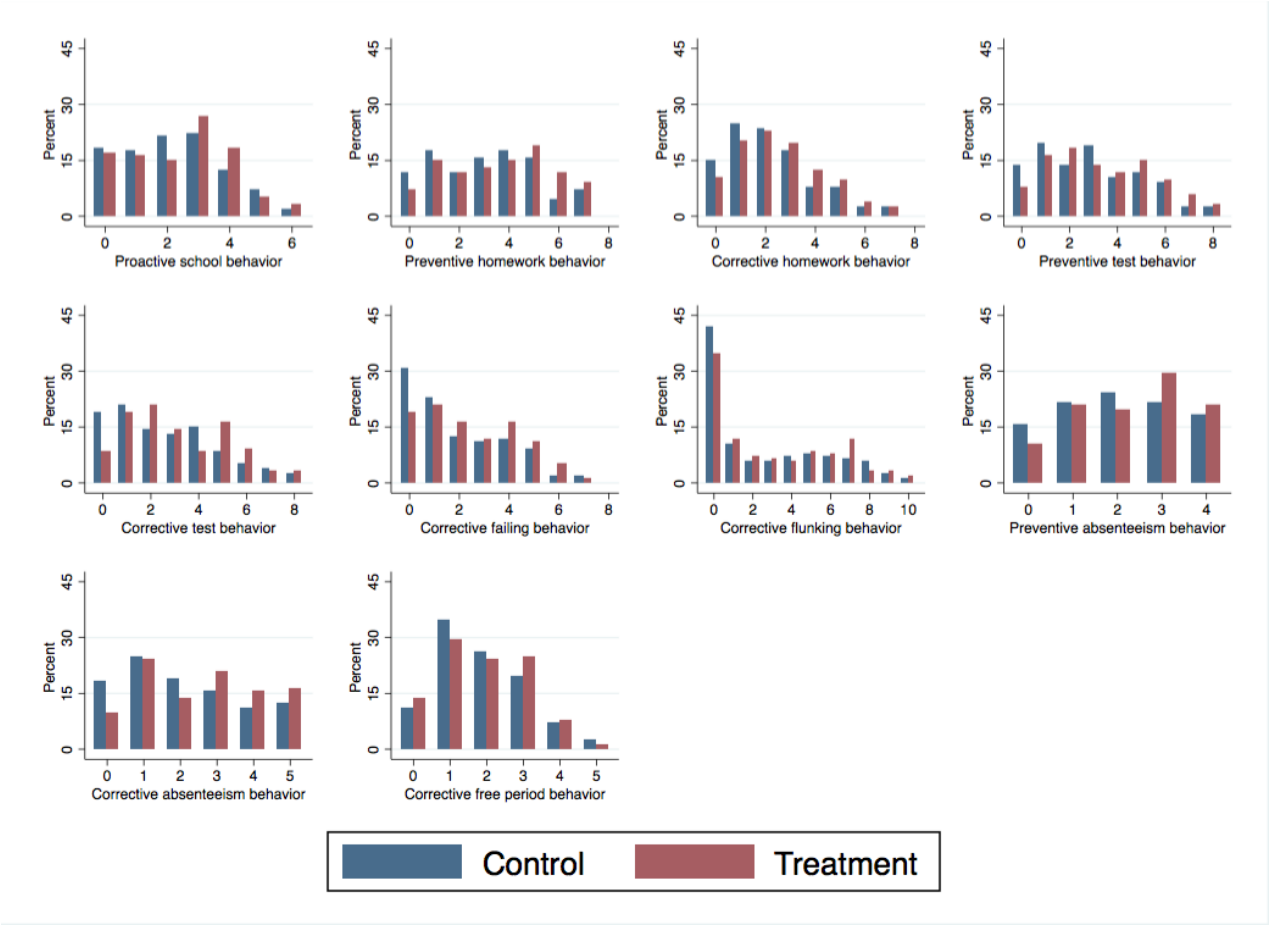
[23] As we mentioned in a previous footnote, in Argentina, when students fail a subject, they need to take an exam to pass it in December. If they fail this exam, they need to take another exam in March. They can fail up to two subjects in March. If they fail more, they can take these exams once again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

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

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

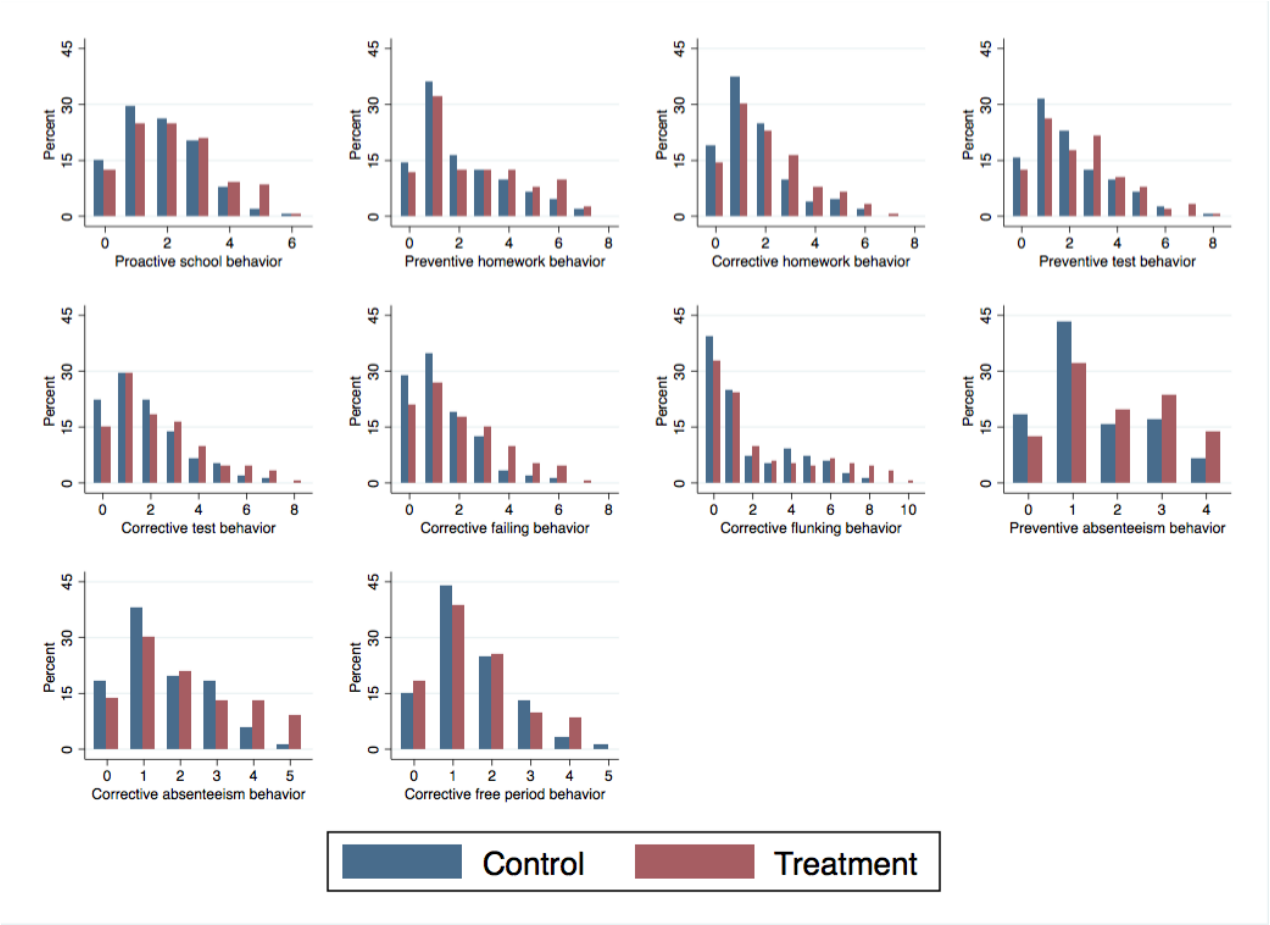
1 if the student engaged in the prompted behavior and 0 otherwise. The score for the index ranges from 0 to 5.

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).

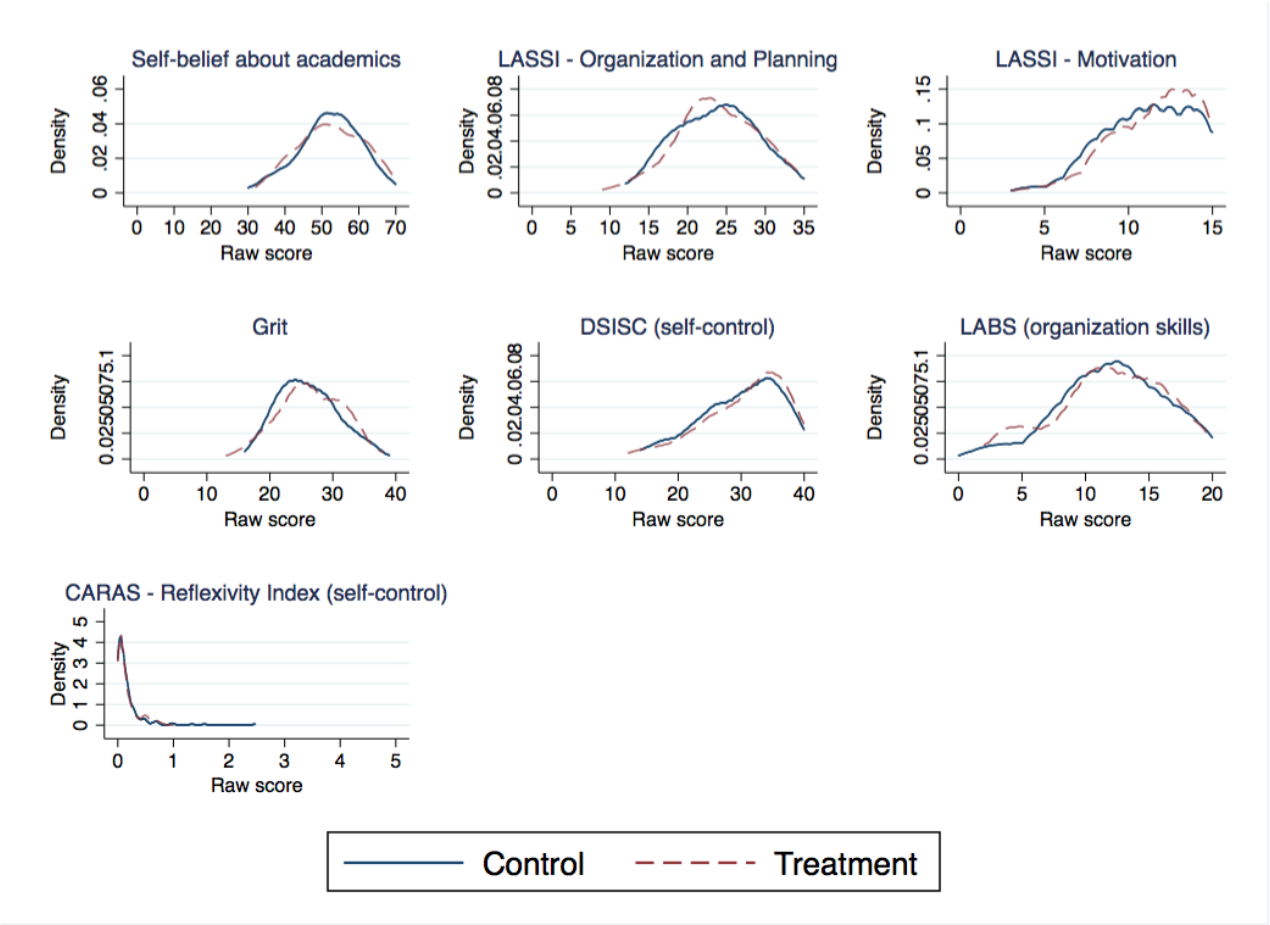
C.5 Labs

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*

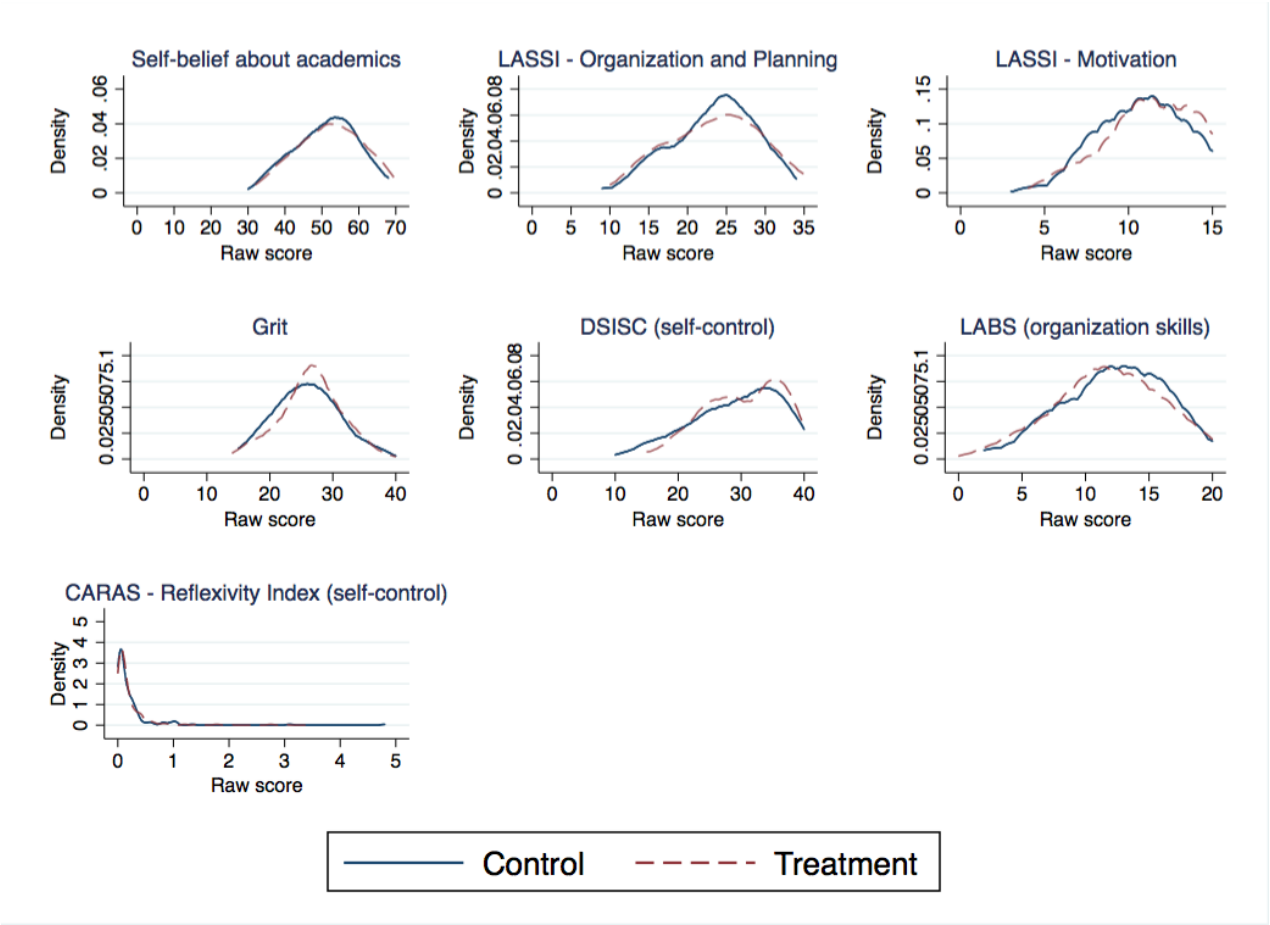
The *Caras* (Smileys) assessment shows students 45 sets of three smileys and asks them to cross out the smiley that is not like the others. The metric of interest is the "reflexivity index": the number of net (i.e., correct minus incorrect) correct answers, divided by the number of errors. This assessment 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)



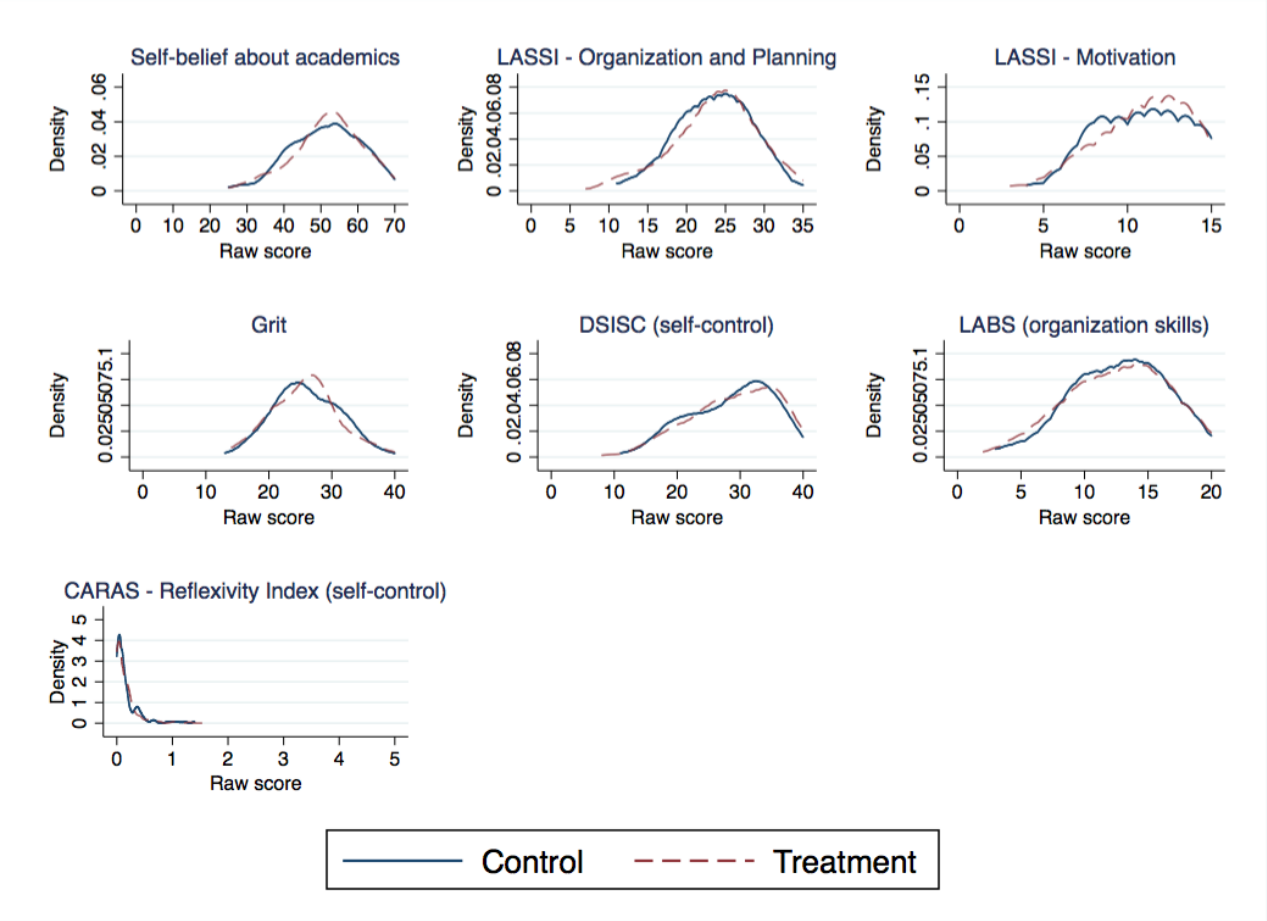
Note: This figure shows the distribution of raw scores for each survey and assessment of academic mindsets, perseverance, and learning strategies in 2014, with a superimposed normal distribution.

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

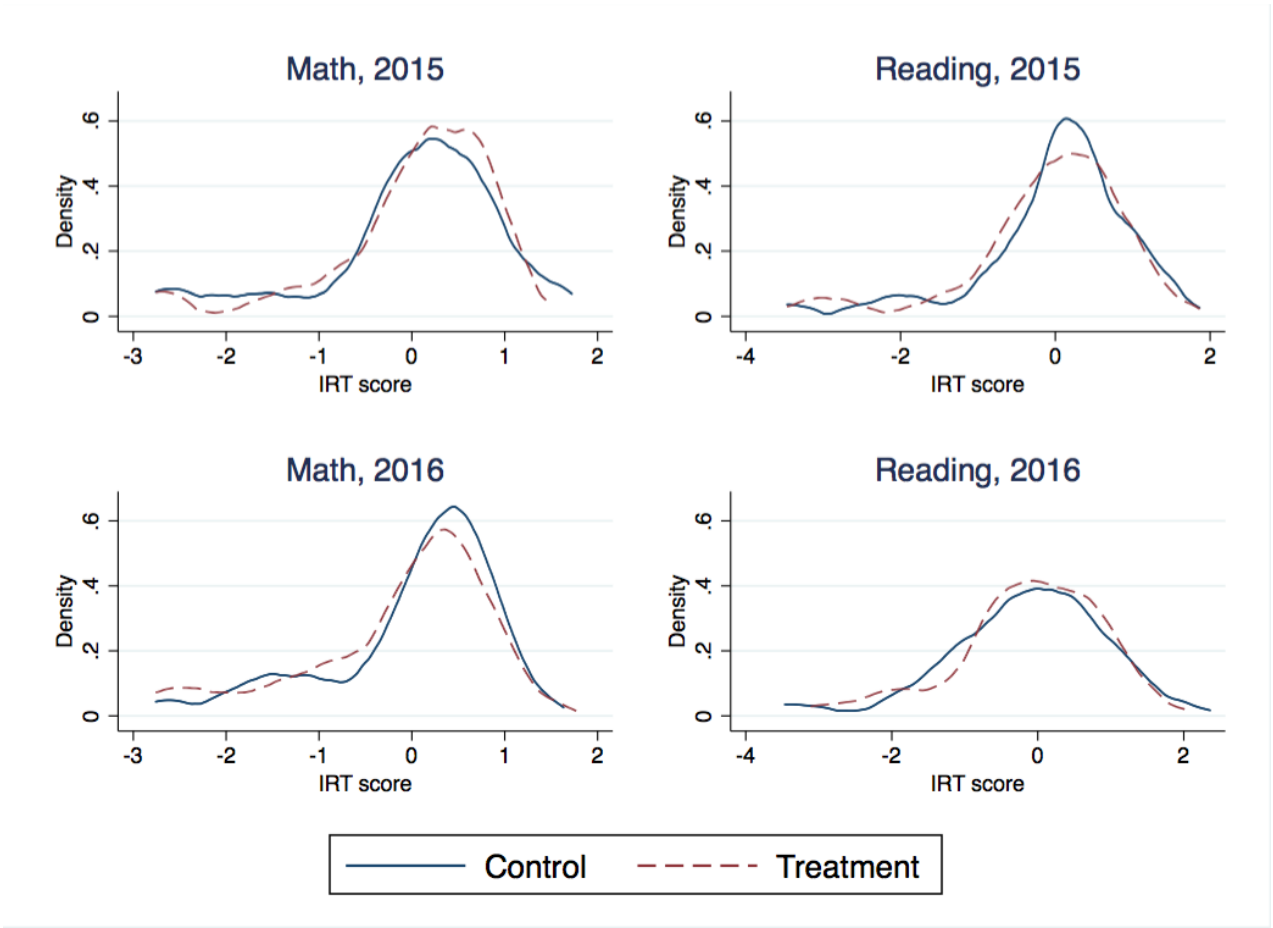
similar ability); b_i, the "difficulty" parameter (i.e., the point of inflection on the θ scale, which reflects where the item functions along the ability scale); and c_i, the "pseudo-guessing" parameter, the asymptotic probability that students will answer the item correctly by chance. The three-parameter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

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

where all parameters are defined 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.