# Tripping at the Finish Line: Experimental Evidence on the Role of Misperceptions on Secondary School Completion

Carolina Lopez*

March 18, 2024

**Abstract**

In many low- and middle-income countries, school completion rates remain low, even among students who have reached secondary school. In Argentina, for example, only half of the teenagers who reach the last year of secondary school complete the coursework necessary to receive a diploma. This paper evaluates a potential driver of this issue: the lack of information about both the probability of graduation and the labor market returns of finishing secondary school. An experiment conducted with senior high school students in Argentina tests the impact of providing information to students on efficiently allocating effort during their senior year and correcting beliefs about their chances of graduating. A second intervention provides information on the returns to education, serving as a benchmark. These interventions increased graduation rates by 10 and 20 percent, respectively, and college enrollment by 38 percent. Notably, improvements in graduation rates were larger for students with a lower initial probability of graduating. Addressing these information gaps in the last mile of secondary schooling has the potential to significantly improve educational outcomes.

**JEL Classifications:** C93, D90, I25, I26

*Development Research Group, World Bank. Email: carolina_lopez@worldbank.org. I am incredibly thankful to Andrew Foster, Pedro Dal Bó, Bryce Steinberg, Anja Sautmann, and Neil Thakral for their support and encouragement. I am specially grateful to Asim Khwaja and Eliana La Ferrara for valuable comments. I am grateful to Alex Eble, Simone Schaner, Jesse Shapiro, Alejandro Ganimian, seminar participants at the Brown Applied Micro Lunch, and the Development Tea group members at Brown University for helpful discussions and feedback. Also, I would like to thank to Angélica Astorga, Felipe Brugués, Amanda Loyola Heufemann, Ignacio Lopez Erazo, Ana Pacheco, Juan Pablo Uribe, Santiago Hermo, Diego Ramos-Toro, Natalia Cantet, and Natalia Sández Pernas for their invaluable support. The field experiment would not have been possible without the authorization and support of the Directorate of Secondary Education of the Ministry of Education of Salta and the authorities of each participant school. I acknowledge the financial support I received to digitized academic records from the Orlando Bravo Center for Economic Research at Brown University and CAF–Development Bank of Latin America. The experiment was approved by the Brown University Institutional Review Board and was registered with the pre-analysis plan at AEA RCT registry as AEARCTR-0004511 (Lopez, 2019). The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent. All errors are my own.

# 1 Introduction

While barriers to education have decreased over time, only 1 out of 3 children complete secondary school in low and middle-income countries (World Bank, 2017) and a large educational achievement gap persists between these children and those in higher-income countries (Glewwe and Muralidharan, 2016). In Argentina, for example, over 95 percent of teenagers in school age are enrolled, but only 50 percent of those who reach their senior year and complete their coursework ultimately receive a diploma. Potential reasons for this gap include lack of information and cognitive biases (such as overconfidence), which may cause students to put less effort than required to complete their degree. Such information gaps and cognitive biases are likely most salient among low-income households, where students may not have exposure to successful graduates or mentors who can provide accurate information. Therefore, a key question is how to induce greater levels of effort, and ultimately higher levels of school completion in these contexts. Completing high school is not only an important signal to the labor market (Spence, 1973), but also has positive impacts on wage prospects (Heckman and LaFontaine, 2010).

In this paper, I study why senior students do not fulfill the requirements to graduate. It is possible that senior students do not take the necessary steps to graduate because they believe that obtaining a diploma is unnecessary or because they are unaware of the effort required to obtain it. While there is evidence on the provision of information about the returns to education,[1] less is known about the latter. At the same time, it is also possible that students believe they will fulfill the requirements to graduate on time, but only because they are overconfident with respect to their chances of graduation. I contribute to the existing literature by generating evidence on the above phenomenon and assessing the relative importance of both aforementioned channels.

In 2019, I conducted a randomized controlled trial in 61 public high schools in the city of Salta, Argentina, to understand whether providing information can improve high school graduation rates. In this setting, senior students fail to graduate because they failed some subjects in previous years and they do not take the make-up exams within the required time frame. Consequently, at the start of the senior year, students are in either good or bad standing, based on the number of pending subjects they have.[2] I estimate the impact of two interventions on the likelihood of graduation for students enrolled as high school seniors. The first intervention (*Production function - PF*) provides information on how to get a high

[1] Jensen (2010) and Nguyen (2008) show positive, Bonilla-Mejía et al. (2019) null, and Loyalka et al. (2013) negative effects of providing information of returns to education on educational outcomes.

[2] In the control group, 55 percent of the students had at least one pending subject at the beginning of their senior year.

school diploma —that is, on the intermediate steps needed to effectively transform effort into educational achievement— and on the probability of high school graduation conditional on academic standing. The second intervention (*Returns to education - RE*), provides information about the economic returns to education. In this setting, a large number of students are at risk of failing to translate their enrollment and attendance into graduation. Failing to obtain a high school diploma can result in drastically lower chances of securing a high-quality job.

The two treatments were implemented through a brief presentation using slides in a single visit to each school and reinforced with reminder messages. In the *PF* arm, the presentation contained statistics on the previous cohort's graduation rates based on their academic standing at the beginning of their senior year, along with information about the intermediate steps necessary to improve academic standing and ensure on-time graduation. The objective of this information session was to establish a link between each potential academic standing (which students knew at the time of the intervention) and the observed graduation rates of students in the same academic condition in 2018, thus allowing students to assess their own probability of graduation. In other words, the *PF* treatment showed them *how* to transform *inputs* (effort) into *output* (high-school diploma). In the *RE* arm, students were shown information containing employment rates and wages by levels of education, using the same delivery format as the other treatment arm. To evaluate the effectiveness of the interventions, I used multiple data sources, including a baseline survey, hard copies of individual academic records obtained from each school, and administrative data. The study sample consisted of nearly 1,800 senior students enrolled in public high schools.

I find that both interventions have a positive and significant impact on graduation rates. *RE* increases the probability of graduation by 10 percentage points (almost 20 percent relative to the control group), while *PF* increases graduation by 5 percentage points (10 percent). The students who experienced the largest increase in the probability of graduation in both intervention arms were those with the worst academic standing at the beginning of their senior year. Furthermore, among these students, I find an increase in observable effort, measured by the probability of attending retake exams and the probability of passing those exams. Both interventions also increased the probability of college enrollment by 5 percentage points compared to the control group.

Analyzing heterogeneous treatment effects, I find that students who were less likely to graduate at baseline (prediction based on past academic performance, observable demographic and socioeconomic characteristics) were the ones with the highest increase in the likelihood of graduation after receiving either one of the information interventions. Furthermore, students who were more likely to graduate at baseline and received the *PF* intervention were more likely to enroll in college in the next academic year.

Why does the $RE$ treatment arm have twice the impact as the $PF$ intervention? The $PF$ arm focuses on helping struggling students to allocate their efforts more efficiently to improve their chances of graduation. In contrast, the $RE$ arm targets inaccurate beliefs about the benefits of education in the labor market and encourages forward-looking behavior. The fact that the $RE$ arm had a larger impact indicates that these inaccurate beliefs were likely more prevalent among all students, regardless of their academic standing.

To investigate whether misperceptions of one's own probability of graduation contribute to low graduation rates, I included a question in the baseline survey asking students to estimate their likelihood of graduation. I create indicators of confidence (under- and overconfidence) by comparing this subjective measure with the estimated probability of graduation at baseline (an objective measure). After receiving the interventions, I asked the students again about their chances of graduation: I find that students' subjective probability of graduation is more accurate after receiving the $PF$ treatment arm. When I estimate heterogeneous effects on graduation based on indicators of confidence, I find that for overconfident students the $RE$ arm has a larger effect than the $PF$ arm, while the effect on underconfident students is similar in magnitude under both treatment arms.

This paper's findings demonstrate the effectiveness of providing high school students with timely and relevant information to improve their decision-making in a high-stakes setting. Unlike previous studies that have focused on the use of monetary or non-monetary incentives to motivate students, or on information on the returns to education, I also examine the impact of information on the educational production function on how students' lack of knowledge on these topics can affect their likelihood of graduating from high school. Education is a key lever for both economic growth and intergenerational mobility (Krueger and Lindahl, 2001; Chetty et al., 2014; Psacharopoulos and Patrinos, 2018) and I contribute to the literature by highlighting the significance of small yet crucial pieces of information that can have a meaningful impact on student achievement.

This paper contributes to the existing literature on how the provision of information can affect educational choices. Existing work has examined the effect of providing information on economic returns to education in contexts with low attendance rates (mainly due to economic constraints), showing an increase in school achievement (Jensen, 2010; Loyalka et al., 2013). However, it has also found that providing information on relatively higher wages for unskilled labor may dissuade students from attending high school (Loyalka et al., 2013), or may not have a significant impact on college enrollment (Bonilla-Mejía et al., 2019). Additional evidence shows that solving information frictions about students' academic standing affects educational decisions (Andrabi et al., 2017; Dizon-Ross, 2019). In addition, the literature on low school achievement has focused mainly on economic constraints such as tuition and other fees, clothes, books, etc. While interventions that reduce these costs do increase attendance, they may not necessarily improve achievement (Ganimian and Murnane, 2016). My paper demonstrates that providing information on the features of the graduation production function or returns to education can increase the accuracy of students' beliefs, resulting in higher graduation rates for senior students.

In addition, I contribute to the literature seeking to understand why people do not use services, infrastructure, or adopt new technologies that can improve their wellbeing when they become available to them. This concern, known as "the last mile problem" is present in many contexts (Mullainathan and Shafir, 2013): individuals forget to submit their taxes on time, low-income students do not use financial aid programs to attend college (Bettinger et al., 2009), farmers do not adopt fertilizer (Duflo et al., 2011), among others. Consequences of these suboptimal decisions are more detrimental in contexts where individuals lack family or other forms of social support (Mullainathan and Shafir, 2013) and may impede those without such resources on their way out of poverty. Education is a key domain in which the "last mile problem" has been understudied. I analyze a setting that allows me to observe this issue among senior high school students who are close to graduating but fail to fulfill all the requirements on time.

Finally, I study whether students ignore or discount new information on finishing high school because of biased beliefs. Empirical evidence shows that individuals tend to overestimate the probability of important outcomes (Feld et al., 2017; Heger and Papageorge, 2018; Machado et al., 2018), leading to suboptimal decisions, especially for unskilled individuals (Choi et al., 2014). In particular, overconfidence in an educational context may lead students to study less (Nowell and Alston, 2007). I contribute to this literature by showing that students may incorrectly assess their chances of graduation, by having the wrong production function of the high school diploma: this biased belief about own performance is detrimental to students' chances of graduation, and I demonstrate that providing accurate information can ameliorate those negative consequences.

My results are relevant for informing policy strategies to increase the demand for high school diplomas among teenagers, especially those disadvantaged and at risk of failing to complete high school on time.[3] I study a vulnerable population in a high-stakes setting

[3]Discussions are currently occurring in many countries and international organizations such as UNICEF (Annual Report 2020 `https://www.unicef.org/reports/unicef-annual-report-2020`) on how to recover from the consequences of the COVID-19 pandemic and the related closure of schools and the impacts on student achievement. Low high school diploma achievement was already a concern before the pandemic in Argentina. UNICEF has reported low school achievement (UNICEF-ARGENTINA, 2017), a referent from the private sector highlighted difficulties in hiring young people with a high school diploma (Diario La Nación, August 6 2021), and civil associations, along with the current National Director of High School

where students' probabilities of failing to obtain high school diplomas are high. Consequently, individuals in this setting have a high chance of being classified as not in education, employment, or training (NEET), which represents an increasing concern in Latin America (Tornarolli, 2016). Although access to the educational system is not restricted in many settings, youths' lack of information can cause them to invest less than the optimum level of effort in education, which can limit their economic opportunities in the medium term by hindering their ability to attend college and work in a job market that places significant importance on high school diplomas as a signal.

The remainder of this paper is divided as follows. In Section 2, I briefly describe the context in which I carried out this randomized controlled trial. In Section 3, I discuss the theoretical framework and predictions for graduation and mechanisms. Section 4 describes the experimental design, randomization, and details of the information interventions of this paper, Section 5 shows the main results, along with their underlying mechanisms. Section 6 presents the main conclusions.

# 2 Context

In Argentina, education is compulsory up to the end of secondary school (5 years, from grade 8 to grade 12); there are free public schools in every district and transportation is sometimes free for students as well. Secondary education is thus accessible for most students. As a result, the share of secondary school-age youth who are attending secondary school is 95.1 percent, with 74.5 percent attending public schools (CEDLAS and World-Bank, 2022). However, high school graduation rates remain low throughout the country. Less than half of the teenagers enrolled in high school actually graduate (UNICEF-ARGENTINA, 2017). Students drop out at different points during high school, but even those who complete the senior year[4] (and attend until the last day of classes) often do not obtain a high school diploma because they fail to fulfill all the mandatory requirements of the system. This is explained in the following subsection.

## 2.1 Educational System and Students' Academic Standing

Students may not graduate because they drop out at different points during high school, mainly owing to "the need to assume adult roles, such as working outside or inside the home, caring for younger or older family members, or taking care of other domestic chores... Other

level, have expressed concerns related to low completion rates (Diario La Nación, August 7 2021).

[4]Throughout this paper I will call grade 12, the last year of secondary school, "senior year."

students drop out because they are not able to deal with school institutional guidelines."[5] But another important explanation, which has attracted less research attention and is not mentioned by the Director of Secondary Education at the national level, is that *students who attend until the last day of high school may still not obtain a high school diploma*. This topic has remained unexplored basically because there are no digitized data at the individual level that allow making conclusions about the magnitude of this issue.

To graduate from high school, students must pass a fixed number of subjects per year (usually 10-12).[6] The academic year begins in March and classes finish by December, but the year officially ends in February. In December and February there are examination dates which allow students who failed subjects during the academic year to remedy their academic standing. Students who receive a score higher than 5 (the exams are graded on a 10-point scale) pass subjects which they previously failed. If a student does not remedy their standing in all subjects by the beginning of the next academic year, they can still be promoted with at most two failed subjects —with a grade lower than 6 (if a student has three or more failed subjects, they must repeat the year). Those failed subjects must be passed at some point during the students' following years of high school to receive a diploma; I refer to these failed subjects as *pending subjects* going forward. All high schools have three examination dates on which students can pass pending subjects each year (July, December, and February). At any given time while in high school, students can have at most two accumulated pending subjects (for example, they can have one from grade 10 and another from grade 11 or 2 from grade 10).

Each student is fully aware of the number of pending subjects they have.[7] I use this concept throughout this paper to categorize students by academic standing at the beginning of their senior year. They can be considered as "in good standing" (zero pending subjects) or "in bad standing" (one or two pending subjects). During phone interviews, school administrators said that the main driver of low graduation rates is the prevalence of pending

[5] Interview with the Director of Secondary Education of the National Ministry of Education about low graduation rates (Diario La Nación, August 7 2021).

[6] There are no national or provincial exams to determine minimum levels of proficiency or to enroll to post public secondary education. According to a national law (`https://www.argentina.gob.ar/normativa/nacional/ley-24521-25394/actualizacion`) "All persons who pass secondary education can freely and unrestrictedly enter at the higher education level."

[7] In the grade reports that students receive at the end of the academic year, failed subjects are highlighted and pending subjects from previous years are noted in a dedicated section. During the academic year, these reports are sent (via students) to the parents/guardians to be signed every quarter. Although it is possible for students to forge signatures, parents are aware of the dates they should receive a report. To verify parents'/guardians' knowledge of their high school senior students' academic status, interviews were conducted prior to the design of the intervention. The adults reported that they were fully aware of their children's academic status and pushed them to improve their situation, but "they are not able to enforce rules."

subjects; the administrators report that students either fail the examinations that would allow them to pass pending subjects or do not attend them at all. They also stressed the importance of timely graduation, as once students are out of the formal system, it is less likely that they will return to school to inquire about the steps needed to obtain a diploma. Those who are more likely to return are individuals who have found employment and are asked by their employers to provide proof of their high school diploma.

## 2.2 Educational Situation in Salta

The intervention was carried out in the city of Salta, the capital of the Argentinian province bearing the same name. In this setting, education and transportation are free for all students enrolled in formal schooling. In 2018, the province of Salta had the eighth-largest subnational secondary school system in Argentina (among 24 provinces), but it was one of the country's worst-performing school systems (Ganimian, 2020): in 2017, only 28.7 percent of students in their senior year of high school performed at a "satisfactory" level in math.

According to self-reported data from an anonymous national survey of students collected at the end of the 2017 academic year (Aprender, 2017), almost 40 percent of senior students were in bad standing (had at least one pending subject). In Figure 1, Panel A, I show data from the control group (cohort 2019), and I observe that at the beginning of their senior year, more than 55 percent of the students had at least one pending subject. These findings indicate that the chances of timely graduation for these cohorts were low, and at the same time, it reveals how common it is for students to have pending subjects at the beginning of the academic year. In Panel B, I show how in fact, having pending subjects is detrimental to obtaining the high school diploma on time: on average, for those with 2 pending subjects, the graduation rate is 12 percent, while for those with 0 pending subjects, the graduation rate is 87 percent.

Table 1, Panel A, shows that in the control group, the graduation rate is 50 percent among students who achieved the senior year and that students had higher expectations about their chances of *timely* graduation when I asked them about their perceived likelihood of graduation in the baseline survey.

At the onset of this study, qualitative field work was conducted to understand why students who had already invested at least 5 years of their lives attending high school were failing to obtain a diploma in their last year. Principals, other school authorities, and teachers were in accord in reporting that students do not exert enough effort to pass pending subjects and often do not attend the examination periods to remedy their standing. They also noted that these issues become worse during students' senior year.[8] [9]

Students in bad standing stated that they did not use the examination exam dates because they had other "important" matters but they would use the next available one "for sure," pass the exam, and receive a diploma on time (by the end of the senior academic year). Procrastination can be a feasible explanation for this behavior; these students face several dates to remedy their standing and do not use them. Evidence indicates that the use of deadlines has no effect on educational outcomes (Gershoni and Stryjan, 2023) because students face other constraints. I consider a potential behavioral bias in this setting: overconfidence. By using a definition of confidence described in detail in Subsection 4.6, I classified students as over- and underconfident. At baseline, more than 80 percent of the students are overconfident. In Table 1, Panel B and C, I show that the performance of overconfident students from the control group is lower than the performance of underconfident students. The confidence of most students, mainly in bad academic standing, in getting the diploma suggests cognitive dissonance regarding what they believe about their actions and effective effort to obtain the certification. I use this insight in the next section to develop a theoretical framework that relates beliefs to effort.

# 3 Theoretical Framework

Previous literature in economics and psychology indicates that performance in education is inversely correlated with overconfidence. Those with better performance "know more about what they do not know" (Machado et al., 2018; Banks et al., 2019). This indicates that unskilled students are more confident than the skilled ones.

But what happens if they learn the true probability of the outcome they are confident about? How will students' beliefs and therefore their subsequent behavior change if they

[8]The last year of secondary education is an eventful year for the students owing to several institutional and non-institutional activities, with students beginning to make arrangements in 11th grade. Some of these activities are the *último primer día* (last first day of classes in the secondary level), *presentación de la promo* (every year each class's members pick colors and a name that represent them, and design t-shirts and hoodies personalized for each student; they introduce their colors, name, and clothing to the rest of the school using music and a performance, inviting all their relatives), commencement ceremony (regardless of whether they obtain a diploma, all senior students participate in a ceremony organized by the school where non-official diplomas are delivered to each student; this ceremony celebrates their presence in the school after at least 5 years), *prom night* (a dinner organized and hosted by students, with the participation of school authorities, teachers, and students' relatives), and other private events hosted by students.

[9]In addition, I conducted qualitative interviews with the main employment agencies that medium and large firms hire to recruit employees in Salta. Recruiters stated that even for jobs that require minimum skills, such as cashiers and shelf stockers, employers require completion of secondary school. Employers are also starting to prefer young people attending any level of education beyond high school to compensate for their lack of experience and as a "signal of responsibility and commitment."

are informed about their true probabilities of graduation? The answer is not obvious. Some overconfident students will realize that there are things they do not know and will respond with more effort, while others could learn that they are too far away from the goal and become discouraged. Some underconfident students may become motivated and work harder to achieve their goal, while others may obtain confirmation of what they already believe and will not change their effort.

I formalize these insights in a model that relates effort to probability of graduation and beliefs. I show how the provision of information affects beliefs, then effort and consequently affects the probability of graduation. This is not the only possible model that could explain the insights that motivated this experiment, but it helps to produce a simple way to think about the impact of the treatments on effort and graduation.

### Assumptions

*Preferences and Beliefs.—* In this model, a student in her senior year decides how much effort $e$ to exert to graduate. Graduation provides a reward in terms of utility, $g(.)$ times the value of getting the diploma $V$ (the *returns to education*), but exerting effort is costly. I assume $g(.)$ is a concave production function and the main primitives of the model are described below.

How effort translates into probability of graduation (production function $g(.)$) and its cost of the depends on student's type $i$. There are two possible types: type (1) students with high return to effort in senior year $\beta_h$; type (2) students with low return to effort in senior year $\beta_l$. In addition, even if students do not exert effort there exists a positive probability to obtain the diploma given by $\alpha$ which captures students' ability and past effort, and also there are two types $\alpha_h$ and $\alpha_l$. Given these assumptions, the production function of the high school diploma is expressed as follows: $g(\beta_i e + \alpha_i)$.

Costs linearly depend on effort and I assume there are two types of cost, depending on students' type: a student with high ability and as a consequence better performance will have a lower cost than a student with less ability. The cost function is then $\delta_i e$ where $i = l, h$.

*States of the World.—* Students may be uncertain about the shape of the production function in their senior year and their abilities. For the sake of simplicity in explaining my main arguments, I will assume that there are only two potential states that combine these beliefs: the first one has a probability $p$ and the second one $(1 - p)$. There are four potential combinations of $\beta_i$ and $\alpha_i$. A student could think that the return to effort is low to get the diploma but it could be compensated with high ability; or the student could think that they own ability is low, so to get the diploma a high return to effort is perceived; and so on.

*Assumptions on Parameters.—* Under uncertainty of the returns to effort, and to illustrate the point of the $PF$ treatment, I make the following assumptions:

- State 1 occurs with probability $p$ this state is represented by $\beta_l$ and $\alpha_h$.
- State 2 occurs with probability $(1-p)$ this state is represented by $\beta_h$ and $\alpha_l$.

I assume that the perceived cost of effort is negatively correlated with the academic standing of students (which could be correlated with ability, Spence (1973)). Importantly, I assume that the $PF$ treatment modifies the perception of $\hat{p}$, and the $RE$ only modifies the perception of $V$, which is represented by $\hat{V}$.

Following my notation, I formalize the concept of self-perception of own probability of graduation:

**Definition 1** *For student $i$, the perceived returns to effort is defined as $\hat{\beta}_i$ and the perceived ability $\hat{\alpha}_i$, then if a student believes that $\beta_i e + \alpha_i < \hat{\beta}_i e + \hat{\alpha}_i$, the student is classified as overconfident; if the student believes that $\beta_i e + \alpha_i > \hat{\beta}_i e + \hat{\alpha}_i$, the student is underconfident.*

The low graduation rate at the end of the academic year may reflects the lack of knowledge of students on several dimensions. The misinformation could be about the translation of effort into graduation or in ability, or the misinformation could also be about economic returns to education. Now, beliefs will play a crucial role in graduation. I assume that uncertainty about the returns to effort is summarized in the perceived probability in which state of the world the student is in $\hat{p}$. Then, the expected probability of graduation is given by:

$$E(\tilde{g}) = \left[\hat{p} g\left(\hat{\beta}_l e + \hat{\alpha}_h\right) + (1-\hat{p})\, g\left(\hat{\beta}_h e + \hat{\alpha}_l\right)\right]$$

The maximization problem is the following:

$$\max_{e} \quad E(\tilde{g})\, \hat{V} - \delta_i e$$

Given the assumptions about the functional forms, this problem has a unique solution given by $e^* = e(\hat{p}, \hat{V})$.

### Role of the Treatment Arms

I consider the effect of two separate treatments. The $PF$ treatment consists of a shock to the students' beliefs about what state of the world they are in. The $RE$ treatment consists of a change in the perceived returns to graduation. I organize the results in two propositions.

**Proposition 1** *(PF) Changes in the belief of the states of the world have an ambiguous effect on the optimal effort. Formally,*

$$\frac{de^*}{d\hat{p}} \lesseqgtr 0$$

**Proof.** See Appendix C.3 for a full derivation. ■

The result of this derivative is *undetermined*, and it depends on the curvature of the $g(.)$ function and the values of its parameters. This formalizes the fact that without further information about students, the direction of the change in behavior (how much effort they are going to exert) is not obvious. Some students will realize that they are in a better state of the world than previously thought and will respond with more effort. Other students have accurate perceptions about the state of the world they are in; for these students, the treatment will only confirm their existing beliefs, and thus might produce no change in exerted effort. Other students could learn they are in the bad state of the world, they could either become discouraged (and exert less effort) or motivated (and exert more effort) upon treatment.

**Proposition 2** *(RE) Optimal effort is increasing in the perceived returns:*

$$\frac{de^*}{d\hat{V}} > 0$$

**Proof.** See Appendix C.3 for a full derivation. ■

This result does not depend on the type of student, and it will be the same regardless of a student being under- or overconfident. An increase in perceived returns to education should lead to an increase in effort.

### 3.0.1 Summary of Mechanisms

The chain of causality in my model is explained as follows. First, students receive one of the two pieces of information, and then, depending on the information received, there are two different mechanisms that explain a change in graduation due to a change in effort:

- *PF:* Students update their beliefs about the right state of the world they are in, and they correct the level of effort they exert to obtain a high school diploma.
- *RE:* Students receive truthful information and update their priors on perceived returns to education, which motivates students to achieve a diploma.

In the next section, I show the experimental design I use to estimate the effect of two different pieces of information on high school graduation.

# 4 Experimental Design

To answer my research questions, I conducted an RCT in the city of Salta,[10] Argentina, from August 2019 to November 2019. The population details and the experiment's design are discussed below.

## 4.1 Ethical considerations

This research project required IRB approval. Given that some minors (according to the Argentinian law, individuals aged less than 18 years old) are included in the sample, consent from parents and students was sought following the instructions of the IRB office at Brown University, the school principals, and authorities from the Ministry of Education of the Province of Salta. In addition, the material prepared for students (contents for the online platform, survey instrument, and presentations) was approved by the Ministry of Education; officials at the Ministry of Education were not informed in advance which information treatment arm was randomly assigned to each school.

## 4.2 Sample

The eligible population for this study is students attending their senior year at public high schools in Salta. While some schools can have more than one shift, I only considered the morning and afternoon shifts due to logistic/budget constraints. Power calculations were conducted using information from the 2018 academic year (see Appendix C, section C.1). In 2018, there were 2933 enrolled students in the senior class across 63 school-shifts. The unit of randomization is at the school-shift level given that randomization at the individual or class level would be more likely to contaminate the control group.

## 4.3 Timeline

At the beginning of this project, in mid-October 2018, I contacted authorities of the Ministry of Education of Salta. The office in charge of supervising my intervention was the Directorate of Secondary Education. They have overseen all the stages of the intervention. In addition to having their approval, I needed the direct approval of each school's principal and vice-principals, who were more aware of the specifics of each shift: school festivities, exams, and trips.[11]

[10] From hereon, Salta refers to the capital city and not the province.

[11] Each school has one principal and if the school has more that one shift there is a vice principal per each shift. From hereon, I use the term "school" to refer to "school-shift."

This process finished in the first quarter of 2019. At the same time, I requested from the directorate access to five "representative" schools to collect individual data about school performance and graduation. This administrative data was not available, so I followed their recommendation to collect data that was stored in secured rooms at each school building to protect student privacy. The main intentions were to compute statistics at the individual level for use in the *PF* treatment arm and to confirm that the graduation rate is in fact approximately 50 percent, in large part owing to the pending subjects issue (see more details in Appendix B, section B.1).

In two out of those five schools, I tested the survey instruments on groups of 11th graders to assess the time they required and to reword questions if necessary to facilitate students' understanding. Several edits were made to the survey instruments at this point. Revision was crucial because school principals allotted just one hour at each school to avoid disruptions to the schools' usual schedules. Visit day was coordinated with the vice principal at each school. The visits were conducted between August and November 2019 (see Figure 2), before the beginning of the final exams date. During the visits I collected the baseline survey data and I conducted the interventions with the help of research assistants from the Department of Economics at Universidad Nacional de Salta (UNSa). I planned to collect the school academic records by the end of February 2020, after the end of the formal academic year. However, the COVID-19 pandemic hit Argentina by March 2020 and the national government imposed a strict lockdown that included the closure of schools. The government's decision halted the data collection process until March 2021.

## 4.4 Data

### Baseline Survey

A description of the baseline data collection process follows. At least 2 days before the intervention date, the research team visited and delivered to the school administrators envelopes containing consent forms for parents of senior students. At a date and time agreed on with the school administrators, the team met with all students of the school in a single room.[12] A description of the activities conducted during each visit day is shown in Figure 3.

To get access to all schools to collect baseline questionnaire data (including those in the control group) and to implement the interventions, the research team visited all schools in the sample to demonstrate how to access a free online platform with math content (designed for this study along with professors at UNSa). This aspect of the intervention serves as a "placebo" for the schools in the control group. Before the presentation of the online plat-

[12] No authority knew beforehand which treatment was randomly assigned to each school.

form, all students took a baseline survey. The questionnaire included the following sections: demographic characteristics, past academic performance, household characteristics, perceptions about labor market outcomes (employment and earnings) by level of education, and expectations about each student's future. In addition, a question about the self-perception of timely graduation was included in the survey (*subjective* measure of confidence in the probability of graduation).

At the meeting with students, school administrators introduced the research team. Then, tablets were given to students, a short presentation (containing slides with pictures) was shown to instruct students on their use, and the students were asked to fill out the questionnaire. At the same time, a brief explanation of the questionnaire was provided.[13] Afterward, the research team showed a presentation introducing the online platform. If applicable, the information treatments were then conducted. After the presentation, the research team asked students to answer an additional question about their perceptions of their own graduation (the same question as in the beginning of the questionnaire). This question was intended to test for any changes in students' perceptions after hearing the information presented, and is the only experimental outcome included in the survey.

Given that a single presentation, including statistics and unknown facts for the students, could not have been enough to change the students effort, I sent an SMS and/or email two weeks before the December (2019) and February (2020) examination periods (senior students' chances to pass pending subjects and failed subjects) to briefly reinforce the information treatment received (excluding students attending schools in the control group).[14] As was shown in previous papers, reminders can help to boost information interventions (Damgaard and Nielsen, 2018).

### School Academic Records

I collected information about academic performance after the end of the 2019 academic year, in February 2020. As shown in Figure 2, this process was heavily delayed by almost one calendar year because of the closure of schools in response to the COVID-19 pandemic. Those individual records contained data on performance during the entire school year and graduation, as well as information about students' pending subjects (if any) and attendance at examination dates for senior students' pending and failed subjects. An example of an individual record is shown in Figure A1, Appendix A.

[13]In schools where a high attendance of students (more than 80) was expected, questionnaires were delivered in paper format.

[14]Cellphone numbers and email addresses were collected during the baseline survey. See the reminders in Appendix B, section B.2.

**Administrative Records**

I also collected information on college enrollment and formal employment. I obtained university enrollment information for the 2020 academic year —the academic year immediately after the graduation of my treated cohort— from the main universities of Salta (UNSa and Universidad Católica de Salta, UCASAL) and formal employment information from SIPA (Sistema Integrado Previsional Argentino), which is an integrated database set up jointly by the social security administration, ANSES (Administración Nacional de Seguridad Social), and the national tax authority, AFIP (Administración Federal de Ingresos Públicos).

## 4.5 Experimental treatments

The treatment assignment was randomly determined at the school level stratifying by the number of students and geographic area of Salta. Information interventions considered in this study are described below (a complete description of the treatments is in Appendix B, section B.1).

> *Production Function - PF*: Using data from a subset of students of the previous cohort (2018), I computed the rate of on-time graduation (by December 2018, after the December examination period) for students with and without pending subjects at the beginning of the 2018 cohort's senior year. The overall on-time completion rate for this subsample was 50 percent. Having pending subjects is not necessarily the main cause of failure to obtain a diploma—students can fail to pass additional subjects in their senior year—but providing this information would highlight the role of pending subjects in getting a diploma and the importance of using examination periods. The provision of this information should highlight aspects of the production function of high school graduation that students do not fully know or understand, such as how much effort should be devoted to passing pending subjects and subjects taken during students' senior year.[15]
>
> Suggestions about *how* to improve academic standing were provided to all students (because at the time of the visit the status of each student was unknown to the research team). All of these suggestions were *intermediate steps* to effectively transform inputs into outputs. The information provided included the following: request mock exams (*modelos de examen*) from teachers,[16] ask for study material

[15]See a discussion about this specific piece of information in Appendix B, section B.3.

[16]These exams should be available for every subject and all years, as was requested by the Directorate of Secondary Education for all public high schools since 2018. Given that compliance of all the teachers could not be verified before the intervention, this information was included in the presentation, highlighting the

from classmates or students from younger cohorts (given that teachers and required academic material can change over time), talk with teachers in advance to ask them for studying recommendations, or ask which teachers will be a part of the committee in each subject.[17]

*Returns to Education - RE*: Students might not be aware of the disadvantages of not finishing high school and the impacts on their labor market prospects. The provision of information about the formal employment rate and average earnings by level of education should incentivize students to obtain a diploma on time (to attend college or find a job in the formal sector). This piece of information is akin to Jensen (2010). In my case, I use data from the National Household Survey (second semester of 2018), restricting the sample to employed individuals aged 18-30 who reside in Salta and are not currently enrolled in any form of school. I computed Mincer equations considering, in addition to the maximum level of education achieved, age, gender, and marital status to compute average monthly wages and formal employment.

No information was provided in the control group. As in the other arms, this group received the presentation about the free online platform and its use is not part of this analysis.[18]

Only one school principal with two shifts (out of 64 schools) refused to participate, even though I had authorization from the Directorate of Secondary Education. After several conversations, the reasons were not disclosed and authorities of the Ministry of Education preferred not to force the school principal to participate. Another school was excluded from the analysis due to administrative complications in the implementation.

Students' participation differed between the intervention treatment arms (see column 1 in Appendix Table A1). A higher percentage of students and parents decided not to participate in the *PF* treatment. This selection into participation could have had detrimental impacts on the analysis of this treatment arm, but the protocol of the visits to the schools allow me to discard selection in participation: no school authorities knew beforehand which treatment was assigned to their school and the research team itself only knew which treatment should be implemented 30 minutes before the arrival at each school. To test for the reason of participation differences, Figure A2 in Appendix A shows that the difference is driven by a

fact that it was mandatory for teachers to prepare that material.

[17] Usually, the committee for each subject/year is formed by three to five teachers depending on the number of students enrolled for that particular exam period. Also, exams are mostly written exams to have proof of the performance of the student in case any dispute arises with parents.

[18] As specified in the AEA registry of this project, two other interventions were conducted. One could not be implemented as planned due to logistics constraints, and the other had a null impact due to issues with the timing of the implementation. Appendix B, section B.4 explains the details of both interventions.

single school with low participation rate, as it can be observed in Panel B. By excluding the observations from that school the significant effect disappears (column 2 in Appendix Table A1). The main results of this paper are robust to the exclusion of that school (see Table A3 in Appendix A).

### 4.6 Measuring Students' Confidence in Graduation

To measure students' self-confidence about graduation, I use two sources of data: the baseline questionnaire and administrative data that provide information about the graduation of each student. I use a question that asks about the self-estimated probability of graduation as a *subjective measure* (see Figure A3, which was used in the questionnaire) and a set of observable characteristics of the students and their households to predict the probabilities of graduation as an *objective measure*. For this step, I first only consider observations in the control group and then extrapolate the predictions to the entire sample.

Given the graduation difference that I observed at baseline for students with zero pending subjects versus those with one or two pending subjects, I estimate different predictions for each group. I use a lasso approach to select the covariates in each regression and avoid searching. The candidate variables selected were individual and household characteristics; area of the city dummies; student age; student gender; if the student has children or is pregnant; average grades during the first two quarters of the senior year; if the student has a job or takes care of a family member; if the student repeated at least one year in secondary school; if their parent/guardian has some post-high school education; if the student does not live in an overcrowded dwelling; if the household has a computer, a washing machine, air-conditioning, or heating; and pairwise interactions between all previously listed students' characteristics. Missing values were recoded to the sample mean and separately dummied out. These missing dummies are also used to construct pairwise interactions. In addition, I added graduation from the 2018 cohort at the school level, along with strata fixed effects.

Figure 4 shows the distribution of the estimated probabilities on the left and the distribution of the difference with respect to the self-estimation of students' graduation probabilities on the right, in Panel A for students with 0 pending subjects and Panel B for students with at least one pending subject. According to my definition of confidence (see Section 3, **Definition 1**), students with a positive difference are classified as underconfident (the objective measure is higher than the subjective one) and those with a negative difference as overconfident. Figure 5 shows that there are no differences across treatment arms. In Appendix Figure A4, I show the proportion of over- and underconfident students by treatment arm and gender.

# 5 Results

In this section, I first show balance checks across treatment arms, then I discuss my main empirical strategy, my results on high school graduation, and the mechanisms that could explain the impacts. In addition, I study some heterogeneous impacts by socioeconomic status and gender and the effects of the treatment arms on college enrollment and formal employment. I close this section by discussing a potential driver of the results on high school graduation and college enrollment.

## 5.1 Balance Checks

Table 2 shows the general characteristics of the students included in my sample and verifies the randomization balance by using the baseline survey and administrative records. The first column of the table displays means and standard deviations of baseline characteristics in the control group (students who attended classes the day of the visit of the research team and gave consent for participation). Columns 2 and 3 present coefficients from the following regression specification:

$$y_{is} = \beta_0 + \beta_{PF}\ \textit{Production Function}_s + \beta_{RE}\ \textit{Returns to Education}_s + \delta_s + \epsilon_{is} \tag{1}$$

where $y_{is}$ is the outcome of interest for student $i$ who attends school-shift $s$, the dummy variables $\textit{Production Function}_s$ and $\textit{Returns to Education}_s$ indicate which information treatment school $s$ received, $\delta_s$ indicates the strata fixed effects (Bruhn and McKenzie, 2009). Errors are clustered at the school level. To control for previous differences in graduation, I add graduation rates at the school level from the previous cohort (senior students in 2018). I could not collect this information before the randomization procedure (to capture differences in school quality) so I add this variable as a control. Each row shows results from a separate regression. Columns 4 and 5 show p-values of the tests of *PF=RE* and *PF=RE=0*, given that the comparison of the two information treatments is of special interest.

Table 2 Panel A shows that the average number of students that participate in each school visit is almost 31 and there are no significant differences between treatment arms. Panel B shows students' characteristics. On average they are 18 years old. Sixty percent of participants are female, and 6 percent have children (all students) or are pregnant (if female). At the time of the visit, 73 percent of the students had an email address and 86 percent reported having access to a cellphone. Eighty-seven percent of the students live with their mother and 58 percent live with their father.

Panel C shows some household characteristics. Seventy-six percent of the students report having a computer (desktop or laptop), and 85 percent state that they have some internet access (via their household, cellphones, school, or public places). On average, students' households have 1.74 persons per room. Thirty-three percent of the students have at least one parent or guardian with at least some college education. Forty-five percent of the students state that they are working—either for a family business or independently—and 20 percent state that they take care of a family member. There are no statistically significant differences in these measures between the two treatment arms.

Panel D includes information about past academic performance of the participants in high school (self-reported). Thirty-eight percent of the students state that they have repeated at least one year during high school, and 55 percent had at least one pending subject at the time of the visit.

Panel E shows the variables that indicate expectations. Ninety-five percent of the participants stated that they want to attend college the next academic year and 87 percent are interested in looking for a job after the end of the school year. At the time of the school visit, students perceived that their chances of on-time graduation were 78 percent. None of these variables exhibit statistically significant differences between information treatment arms.

## 5.2 Empirical Strategy and Main Results

To estimate the effect of the information treatments, I use the following specification:

$$y_{is} = \beta_0 + \beta_{PF}\ \textit{Production Function}_s + \beta_{RE}\ \textit{Returns to Education}_s + \delta_s + x'_{is}\omega + \eta_{is} \quad (2)$$

This equation is the same as equation (1) but is augmented to control for additional individual characteristics given by $x'_{is}$. To avoid specification searching of covariates, they were selected using double lasso (Belloni et al., 2014). Also notice that $y_{is}$ here represents the main outcome of interest: graduation. I interpret the results through the lens of the model described in Section 3.

Table 3, column 1, shows that graduation for all students who were selected to participate in either treatments arm increases, and the effects are statistically significant: (1) students in the *PF* treatment arm are 5 percentage points more likely to graduate (10 percent with respect to the control group) and (2) those in the *RE* are 10 percentage points more likely to graduate (20 percent with respect to the control group). I find that the difference between these treatment effects is statistically significant. Results with no controls are shown in Appendix Table A2.

Jensen (2010) finds an increase of 5 percentage points in the likelihood of graduation of

Dominican students using an intervention closer to *RE*. A potential explanation for the higher impacts found in the current study could be related to the fact that my target population was comprised mainly of students who were closer to receiving their high school diplomas. Additionally, my setting has fewer economic barriers: enrollment and transportation to school are free.

**Proposition 1** in the model stated in Section 3, shows that impacts of *PF* on effort (and then graduation) are undetermined. Results show an increase in the likelihood of graduation, indicating a higher effort among students or a better allocation of it.

In Table 3, columns 2 and 3 show the treatment effects by academic standing at the beginning of senior year: in good standing (zero pending subjects) or in bad standing (at least one pending subject). I do not observe a significant effect for students in good standing for students in the *PF* arm, and the magnitude of the effect is close to zero. A likely reason for this finding is that these students already know how much effort they should devote to study to succeed (because of that they are in good standing). This is not the case for those students in bad standing. Although they received bad news via the *PF* (being in bad standing is correlated with a low probability of graduation), the information provided should help them to realize *how* to allocate their effort effectively in order to obtain a diploma. For this subset of students, I observe an increase of 7 percentage points (more than 30 percent with respect to the control group).

**Proposition 2** stated that the expected sign of the *RE* arm is increasing in the perceived returns. I found higher positive impacts of this treatment arm for the entire sample, and as columns 2 and 3 of Table 3 show, when I separate the analysis by pending subject condition, I observe positive impacts for both students in good and bad academic standing.

In the next subsection, I discuss with more detail potential channels that could explain my main results on graduation.

## 5.3 Mechanisms for Production Function and Returns to Education

### Perceptions on Graduation

To understand the drivers of these results, I study the role of self-perception of graduation on actual graduation (Table 4) by using the answers to the questions about students' own chances of graduation before and after the interventions. An important part of the *PF* treatment was to make students aware of the correct shape of the production function of the high school diploma based on their academic standing at the beginning of the senior year. As previously mentioned, at the time of the intervention, the standing of the students was their private information and the goal was to allow students to create a *mapping* of their situation with regard to graduation rates of similar students from the previous year. I computed the difference of the *subjective* probabilities of timely graduation (*after-before*) to check for the direction of the updates.

Under the theoretical framework presented above, perceptions of graduation should only change if students update their beliefs about the level of effort needed to obtain their diploma. This is only possible if they receive information about the actual probabilities, the effort that is required, and all the intermediate steps needed to successfully transform that effort into graduation. Table A4 shows how students update their *subjective* probability of graduation by comparing self-reported probabilities measured before and after the provision of the information treatments. Individuals who received the *PF* treatment became more accurate with respect to their own chances of graduation: the variable decreases by 2 points compared to the baseline response (statistically significant at the 5 percent level).[19] Overconfident students drive this result (column 3, Table A4).

I analyze graduation by academic standing and its relationship to my definition of confidence in Table 4. I interact the treatment received with dummy variables that indicate the level of confidence (under- or overconfident, see Figure 4), and I show impacts on graduation for the entire sample. Importantly, the results show that none of the treatment arms caused a discouragement effect. Among the students who received the *PF* arm, underconfident students were 8 percentage points more likely to graduate, although the difference with overconfident students (5 percentage points) is not statistically significant. Additionally, the *RE* arm had a larger effect on overconfident students compared to the *PF* arm, and the difference in graduation is statistically significant at the 5 percent level.

The improvement in overconfident students' perception of their graduation probability, observed in Table A4 as a result of the *PF* intervention, did not translate into a statistically significant increase in graduation for this subgroup of students by the end of the academic year.

### Performance During Senior Year

To understand how the information treatments impact students' performance during the academic year, I separate the analysis by considering what happens with the mandatory senior subjects and pending subjects by February 2020 (the end of the academic year). Both of these variables determine if a student receives a high school diploma: if they pass *all* the senior subjects and have no pending subjects, then they graduate. To understand my results

[19]Notice that the students in the control became less accurate (more optimistic about their chances of graduation). A reasonable explanation for this result is that a visit to the school by a student at an American university and students at UNSa could have generated an optimistic response among students given that there is almost no formal connection between secondary and post-secondary levels.

better, I split the sample considering the number of pending subjects, but these results should be considered with caution because of the small sample sizes.

Table 5, Panel A column 1, shows the impact of the information treatments on a dummy variable that indicates if the student passed all the senior subjects for the entire sample. The *RE* treatment increases the probability of passing all the senior subjects by 5 percentage points (7 percent, statistically significant at the 5 percent level). The *PF* arm has a small and non-significant impact. In Panel B, column 1, I study the impact on the probability of passing all senior subjects by the level of confidence at baseline for the entire sample. Results indicate that the positive impacts on the probability of passing all senior subjects of the *RE* are driven by those underconfident students; they are 6 percent more likely to pass all the senior subjects with respect to the overconfident ones, although the difference is not statistically significant.

To analyze the performance of those with pending subjects, Table 5, Panel A column 2, shows the impact of the treatments on a dummy variable that indicates if the student has at least one pending subject left by the end of the academic year. Both treatments decrease the probability of having at least one pending subject: those students who receive the *PF* arm are 7 percentage points less likely to have pending subjects (8 percent, significant at the 5 percent level) and those who receive the *RE* are 12 percentage points less likely to have pending subjects (15 percent, significant at the 1 percent level). In Panel B, I include the interaction by level of confidence: underconfident at baseline are those who respond most to both treatment arms. Column 2 shows that among those who receive the *PF* arm, underconfident students are 34 percentage points less likely to have pending subjects left with respect to the overconfident ones (difference is statistically significant at the 5 percent level), and among those who receive the *RE* underconfident are 21 percentage points less likely to have a pending subject left with respect to the overconfident ones (difference significant at the 10 percent level).

#### Observable Effort of Students in Bad Academic Standing

I analyze the effect of the information treatments on variables that indicate direct measures of effort to pass pending subjects: (1) enrollment in the examination period (December 2019 and/or February 2020) and (2) attendance at the examination period. The first variable indicates effort because, according to high school requirements, only students who explicitly register for the examination date are allowed to take the exam.[20] Therefore, enrollment can be seen as a direct measure of effort. The second variable indicates whether students actually

[20] The exam committee is formed by teachers who are going to be in charge of preparing the exam. If no student is enrolled, the committee is not formed.

attended the examination, which can be considered as an additional indicator of effort. For this variable, I did not restrict the analysis to enrolled students.

Table 6, Panel A, shows positive impacts of the information treatments on these outcomes, but only the variable attendance to examination periods for those who received the *RE* treatment is statically significant (at the 1 percent level, column 2). Panel B shows the effect of the information treatments by confidence level at baseline. Underconfident students respond more to the *PF* by increasing their attendance more than overconfident students; the difference between the two types of students is more than 40 percentage points (significantly different at the 1 percent level). The *RE* treatment arm also has differences in favor of the underconfident students (27 percentage points more likely to attend the examination periods with respect to overconfident students in this treatment arm).

## 5.4 Heterogeneous Effects

### Time Preferences

The *RE* treatment implies a forward-looking behavior on the students' side, given that they have to wait a considerable amount of time to enjoy their labor market outcomes.

Following this argument, I consider the role of time preferences on timely graduation. By using a set of questions in the baseline questionnaire following a standard Becker-DeGroot-Marschak procedure (Bursztyn and Coffman, 2012), I computed the discount factor for each student. I then took the median and separated students based on whether they were above or below the median. Results are shown in Table 7. As expected, the effect in the *RE* treatment arm is greater and statistically significant for students above the median. Although the difference with respect to students under the median value is not statistically significant, it highlights the importance of considering this relevant individual characteristic when providing information like this to teenagers.

Furthermore, it can also be observed that the magnitudes for both types of students (below and above the median discount factor) that received the *PF* are lower and nonsignificant. This finding is consistent with the information provided, as this treatment does not encourage forward-looking behavior.

### Socioeconomic Status and Gender

In the baseline questionnaire, I did not include a question about family income due to that question's low response rate in the pilot survey. To generate a proxy for economic status, I use an index constructed by using variables indicating the ownership of goods including air conditioning, heating, a washing machine, and a personal computer, whether the student's family lives in an overcrowded dwelling,[21] and whether at least one parent or guardian has some post-secondary education. If the index is less than or equal to 3, I classified the student as "poor" and otherwise, as "least poor."[22]

Table 8 shows that in the control group, students classified as poor have a lower graduation rate (45 percent), which is 14 percentage points lower than the least poor students. In column 1, I show that poorer students are positively affected by both treatments: students in the $PF$ treatment arm are 8 percentage points more likely to graduate than the control group, and those in the $RE$ treatment arm are 14 percentage points more likely to graduate than the control group. Both results are statistically significant at the 1 percent level, and the difference of the magnitudes is also statistically significant at the 5 percent level.

Table 8 also shows the impacts by gender. Columns 3 and 4 show that female students are more likely to graduate than male students in the control group. I observe higher impacts for male students but the differences are not statistically significant.

## 5.5 Other outcomes

One of the objectives of this paper was to analyze the effects of information treatments beyond secondary school. Given certain data limitations (explained below), I only consider whether the student is enrolled in a university in the academic year after my interventions were conducted (2020) or enters formal employment from the last quarter of 2020 to the first quarter of 2021.

### College enrollment

College enrollment indicates that a student wants to invest more in their human capital, so exploring the effects of my information treatments on enrollment is key to determining their medium-run effects. To construct this variable, I requested individual enrollment data for the 2020 academic year from UNSa and UCASAL. These are the most important universities in Salta; the first one is public and free, and the second one is private.

An important fact to highlight is that enrollment in UNSa is open and unrestricted by law, meaning that there are no general barriers to access. There are no entrance examinations or quotas, and students' performance during high school does not affect their selected degree. It is important to stress that the only requirement is a high school diploma, although students with pending subjects can enroll provisionally. It was not possible to obtain information on other tertiary educational centers, so my measure only includes universities.

[21] This variable indicates that on average students live in a household with more than two people per room.

[22] For the control group, the median value of this variable is 3 and the mean is 3.12.

In addition, it is not very likely that students from Salta (attending a public high school) would move to another province to attend college. Even if they were to attend a public university in a different location, they would have to consider the cost of moving and housing, which are expensive compared to UCASAL. There are no available data at the national level that would allow me to test the percentage of students who move to another province to study at the post-secondary level. Given these facts, my results represent a lower bound of the effect of the information treatments on tertiary education.

Table 9 column 1 shows that only 13 percent of the students in the control group are enrolled in university, and both treatment arms increase the probability of enrollment by 5 percentage points (almost 40 percent). These effects are statistically significant at the 10 percent level in the case of *PF* arm and 5 percent in the case of *RE*. Bonilla-Mejía et al. (2019) present an experiment aimed to improve college enrollment in Colombia by providing information on returns to education for senior students and no effects were found. A potential explanation for my results is that the settings are different regarding access to post-secondary education: in Argentina there are no examination entrance exams for colleges, public post-secondary institutions are free, and in many districts public transportation for all students is free.

**Formal Employment**

Formal employment is an outcome of interest after high school completion. To construct this variable, I use administrative records of the students by using their national IDs. This is not public information, but participating students (and parents/guardians, if the student was a minor) gave me consent to check their employment status.

The system only allows access to information from the 6 previous months at the time of the inquiry.[23] Given the strict quarantine imposed by the government in Argentina in response to the COVID-19 pandemic, I decided to include information from the last quarter of 2020 (when some restrictions were lifted) to the first quarter of 2021. The output *formal employment* is a dummy variable equal to 1 if the participant was registered as a formal employee for at least one month out of those 6 months.

Column 2 of Table 9 shows the results for both treatment arms. As expected, the level of formal employment for the control group is small; only 3 percent of the students in that group have a formal job at the considered time. However, both treatment arms generate a negative and statistically significant impact on formal employment. A potential, but not conclusive, explanation is that students' reservation wage increased after receiving the treatments.

One key caveat is that the sample size in this analysis is lower than the original sample

[23]See "Administrative records" in Subsection 4.4.

because I did not find information for all students in the administrative data—there were errors in IDs in the data I received from the high schools. To test for potential attrition issues, I created a dummy variable equal to 1 if a student was not found and 0 otherwise. Then I run the main specification and I do not find differences across treatment arms (see Table A5 in Appendix A).

## 5.6 Discussion about Results on High School Graduation and College Enrollment

Besides observing the positive impacts of both treatment arms on high school graduation, I have shown that one arm is more effective than the other. These results indicate that the *RE* addresses inaccurate beliefs that were likely to be widespread among all students, regardless of their academic performance. On the other hand, the *PF* arm seems to focus on students with poor academic standing at the baseline (55 percent of the students in the control group had at least one pending subject at the beginning of their senior year).

At the same time, both arms increase the probability of college enrollment by the same magnitude (5 percentage points). To shed more light on this puzzle, I analyze the differential treatment effects by considering the likelihood of graduation at baseline. The construction of this prediction is described in Subsection 4.6, and I used it to create a dummy variable that indicates whether students were more or less likely to graduate from high school based on observable characteristics at baseline.

Table 10 presents the impacts of the interaction between the dummy variable indicating likelihood of graduation at baseline and both treatment arms on high school graduation (column 1). The results show that among those who received the *PF* arm, students who were less likely to graduate at baseline increased their chances of graduation by an additional of 7 percentage points (statistically significant at the 5 percent level), compared to those who were more likely to graduate. However, the difference between the coefficients for the two groups is not statistically significant. For students who received the *RE* arm, those who were less likely to graduate at baseline increased their chances of high school graduation by an additional of 4 percentage points compared to those who were more likely to graduate, but the difference between the coefficients is not statistically significant. These results suggest that students who are more disadvantaged, are the ones who benefit the most from receiving the treatment. Despite reaching their senior year of high school, these students required additional information to overcome the obstacles hindering their path to obtaining their diploma. These misperceptions were particularly pronounced among those with a lower likelihood of graduating at baseline.

Table 10, column 2, shows the result on college enrollment. These results should be considered with caution, given the small proportion of students enrolled in college (13 percent in the control group), and also it seems there is a small correlation between actual college enrollment and predicted probability of high school graduation (see Appendix Figure A5). Among those who received the *PF* arm, those more likely to get the high school diploma at baseline are 7 percentage points more likely to be enrolled in college during the next academic year with respect to the less likely to graduate. There are no striking differences for those who received the *RE* arm. To understand these results better, I analyze the correlation of students' earnings perceptions reported at baseline with the likelihood of graduation from high school at baseline.[24]

In the baseline survey, I asked students to form a perception of expected earnings (by the level of education). They could have a positive misperception (meaning they overestimate the returns to education, relative to the true values) or a negative one (underestimation of returns to education). I was not able to collect the same information after the intervention (to check for updates in perceptions) because this section was very time-consuming for the students and I had limited time to conduct the interventions. In Figure 6, Panel A, I observe that, on average, students were accurate in predicting levels of earnings for incomplete high school and incomplete college, but they overestimated the returns for complete high school and complete college. Notice that students were inaccurate in predicting the strictly positive correlation between the level of education and average earnings that is present in this setting. To explore whether students more likely to graduate at baseline had higher perceptions of earnings, in Panel B, I show the relationship between a dummy variable that indicate that for each level of education a given student estimated earnings above the median value and the dummy variable that indicates if they are more likely to graduate at baseline. Those more likely to graduate at baseline are more likely to provide higher levels of earnings for incomplete college, with respect to those less likely to graduate.

Taken together these results, there is suggestive evidence that the *RE* arm had a stronger impact on the most disadvantaged students, who learned more about the value of investing in education. On the other hand, the *PF* arm had a greater impact on college enrollment for students who were more likely to graduate at baseline and received information about the importance of having a diploma, while also helping those less likely to graduate obtain their high school diploma. Overall, both arms were effective in increasing high school graduation rates and college enrollment, but the impact differed depending on the student's baseline

[24] The baseline questionnaire also collected data on levels of employment by achieved educational level. However, when all those variables were added up, the total did not add up to 100 percent as was required in the survey.

characteristics.

# 6 Conclusions

This paper analyzes the effect of information interventions on improving high school graduation rates by correcting students' mistaken perceptions through two interventions: a novel intervention and a traditional one. The first intervention aims to make students aware of their chances of graduating based on their academic standing at the beginning of senior year, teaching them how to effectively transform inputs into outputs ($PF$). The second intervention provides information about the returns to education based on the achieved educational level ($RE$).

Students' perceptions about their probabilities of graduation and the returns to education could be modified by providing the correct information that targets each mistaken belief. I observed positive and significant effects in both treatment arms on timely graduation, and the magnitudes are more significant than those found in other studies. I also found positive and significant impacts on college enrollment, while previous studies aimed at driving demand for post-secondary education did not find this effect. I provide evidence that indicates that after the provision of information students exert more effort, are more likely to pass senior subjects, and those from low socioeconomic status benefit the most from these interventions.

I show that the likelihood of graduation measured at baseline explains the positive impacts on graduation and college enrollment. The $RE$ intervention was particularly effective for the most disadvantaged students, as it helped them understand the value of investing in education. The $PF$ helped students who were less likely to graduate to obtain their high school diploma but it had a stronger impact on college enrollment for students who were more likely to graduate at baseline.

The findings of this study have substantive policy importance: graduation rates can be improved in low-income settings using an inexpensive intervention that corrects inaccurate beliefs that are more likely to be present in low-income households. Small bureaucratic hurdles, which those with substantial parental or other forms of social support can easily negotiate, may trip up those without such resources. In these contexts, the provision of small pieces of information offers an excellent opportunity to improve graduation rates, as shown in this paper. Students who are positively affected by this intervention now have a previously unavailable chance to achieve economic mobility. Questions for future research are when it is more effective to provide this type of information to students and whether combining both types of information interventions could further boost the positive impacts found in this study.

# References


Andrabi, T., J. Das, and A. I. Khwaja (2017). Report cards: The impact of providing school and child test scores on educational markets. *American Economic Review 107*(6), 1535–1563.

Aprender (2017). *Aprender 2017: Informe de resultados, Salta, 5to año de secundaria.* Secretaría de Evaluación Educativa, Ministerio de Educación y Deportes de la Nación.

Banks, J., L. S. Carvalho, and F. Perez-Arce (2019). Education, Decision Making, and Economic Rationality. *The Review of Economics and Statistics 101*(3), 428–441.

Belloni, A., V. Chernozhukov, and C. Hansen (2014). High-Dimensional Methods and Inference on Structural and Treatment Effects. *Journal of Economic Perspectives 28*(2), 29–50.

Bettinger, E., B. T. Long, P. Oreopoulos, and L. Sanbonmatsu (2009). The Role of Simplification and Information in College Decisions: Results from the H&R Block FAFSA Experiment. Technical report.

Bonilla-Mejía, L., N. L. Bottan, and A. Ham (2019). Information Policies and Higher Education Choices Experimental Evidence from Colombia. *Journal of Behavioral and Experimental Economics 83*, 101468.

Bruhn, M. and D. McKenzie (2009). In Pursuit of Balance: Randomization in Practice in Development Field Experiments. *American Economic Journal: Applied Economics 1*(4), 200–232.

Bursztyn, L. and L. C. Coffman (2012). The Schooling Decision: Family Preferences, Intergenerational Conflict, and Moral Hazard in the Brazilian Favelas. *Journal of Political Economy 120*(3), 359–397.

CEDLAS and World-Bank (2022). Socio-economic database for latin america and the caribbean. https://www.cedlas.econo.unlp.edu.ar/wp/en/estadisticas/sedlac/.

Chetty, R., N. Hendren, P. Kline, E. Saez, and N. Turner (2014). Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility. *American Economic Review 104*(5), 141–147.

Choi, S., S. Kariv, W. Müller, and D. Silverman (2014). Who is (More) Rational? *American Economic Review 104*(6), 1518–1550.

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

Dizon-Ross, R. (2019). Parents' beliefs about their children's academic ability: Implications for educational investments. *American Economic Review 109*(8), 2728–2765.

Duflo, E., M. Kremer, and J. Robinson (2011). Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya. *American Economic Review 101*(6), 2350–2390.

Feld, J., J. Sauermann, and A. de Grip (2017). Estimating the Relationship between Skill and Overconfidence. Research Memorandum 005, Maastricht University, Graduate School of Business and Economics (GSBE).

Ganimian, A. J. (2020). Growth-Mindset Interventions at Scale: Experimental Evidence From Argentina. *Educational Evaluation and Policy Analysis 42*(3), 417–438.

Ganimian, A. J. and R. J. Murnane (2016). Improving Education in Developing Countries. *Review of Educational Research 86*(3), 719–755.

Gershoni, N. and M. Stryjan (2023). Do deadlines affect project completion? experimental evidence from israeli vocational colleges. *Journal of Economic Behavior and Organization 205*, 359–375.

Glewwe, P. and K. Muralidharan (2016). Improving Education Outcomes in Developing Countries. In *Handbook of the Economics of Education*, pp. 653–743. Elsevier.

Heckman, J. J. and P. A. LaFontaine (2010). The american high school graduation rate: Trends and levels. *Review of Economics and Statistics 92*(2), 244–262.

Heger, S. A. and N. W. Papageorge (2018). We Should Totally Open a Restaurant: How Optimism and Overconfidence Affect Beliefs. *Journal of Economic Psychology 67*, 177–190.

Jensen, R. (2010). The (Perceived) Returns to Education and the Demand for Schooling. *The Quarterly Journal of Economics 125*(2), 515–548.

Krueger, A. B. and M. Lindahl (2001). Education for Growth: Why and for Whom? *Journal of Economic Literature 39*(4), 1101–1136.

Lopez, C. (2019). Understanding low completion rates of high school in a developing country.

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.

Machado, L., , and C. Y. and (2018). Self-Assessment Accuracy, Overconfidence and Student Performance. *The New Educational Review 53*(3), 85–94.

Mullainathan, S. and E. Shafir (2013). *Scarcity: Why Having Too Little Means So Much*. New York: Times Books, Henry Holt and Company.

Nguyen, T. (2008). *Information, Role Models and Perceived Returns to Education: Experimental Evidence from Madagascar*. Ph. D. thesis, MIT.

Nowell, C. and R. M. Alston (2007). I Thought I Got an A! Overconfidence Across the Economics Curriculum. *The Journal of Economic Education 38*(2), 131–142.

Psacharopoulos, G. and H. A. Patrinos (2018). Returns to Investment in Education: A Decennial Review of the Global Literature. *Education Economics 26*(5), 445–458.

Spence, M. (1973). Job Market Signaling. *The Quarterly Journal of Economics 87*(3), 355–374.

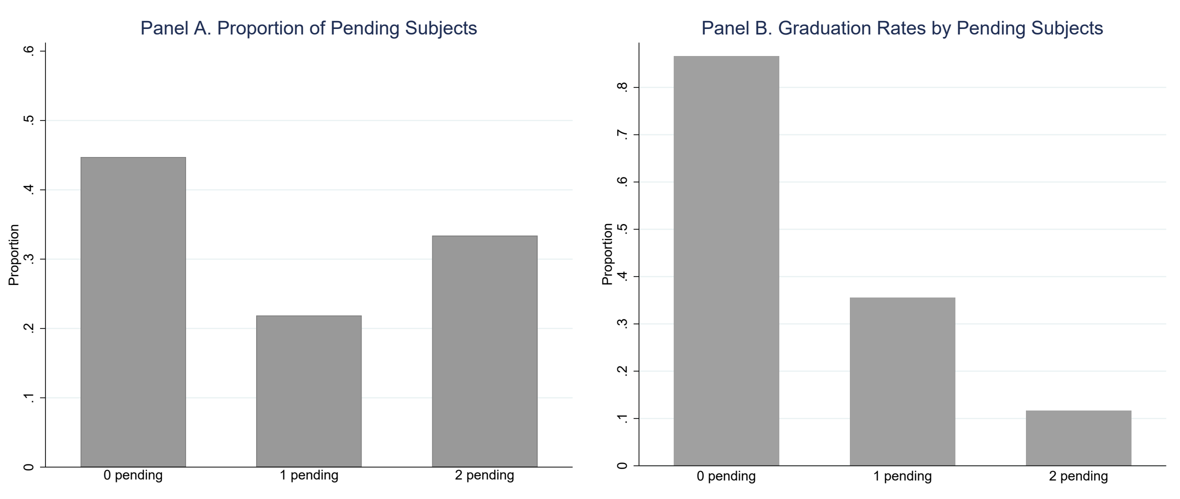
Tornarolli, L. (2016). El Fenómeno de los NiNis en América Latina. Technical Report 2016/18, CAF.

UNICEF-ARGENTINA (2017). *Para Cada Adolescente una Oportunidad: Posicionamiento sobre Adolescencia*. https://www.unicef.org/argentina/media/1396/file/Posicionamiento

World Bank (2017). *World Development Report 2018: Learning to Realize Education's Promise*. Washington, DC: World Bank.

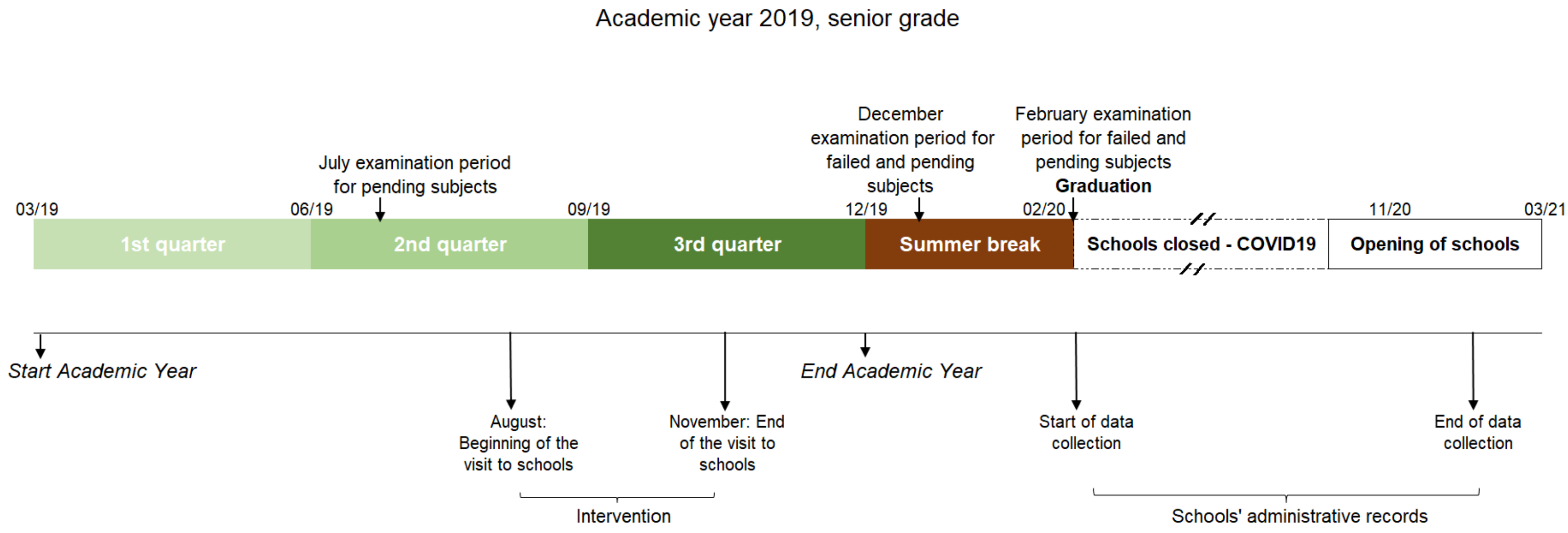
# Figures

Figure 1: Senior Students and Pending Subjects, Control Group



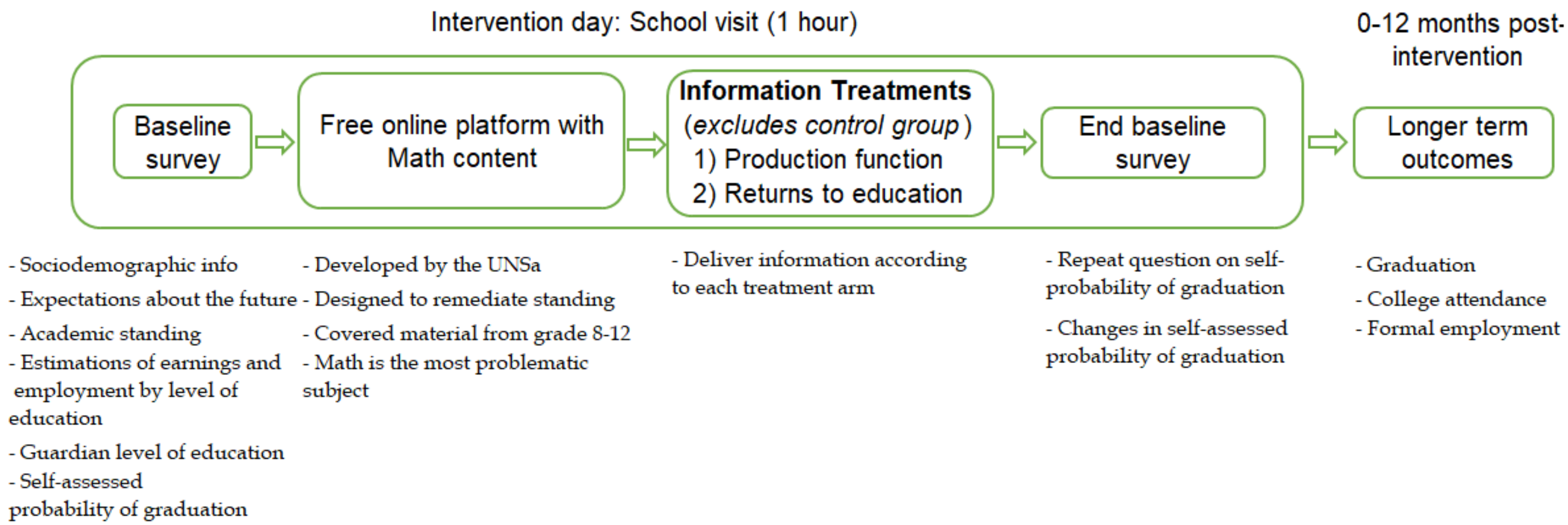

*Notes:* Sample limited to the control group. The horizontal axis displays the number of pending subjects at the beginning of the senior year. Panel A indicates the proportion of students with 0, 1, and 2 pending subjects at the start of their senior year. Panel B displays the average graduation rate for students based on the number of pending subjects at the beginning of their senior year. The data was obtained from schools' administrative records.

Figure 2: Timeline, Intervention and Data Collection



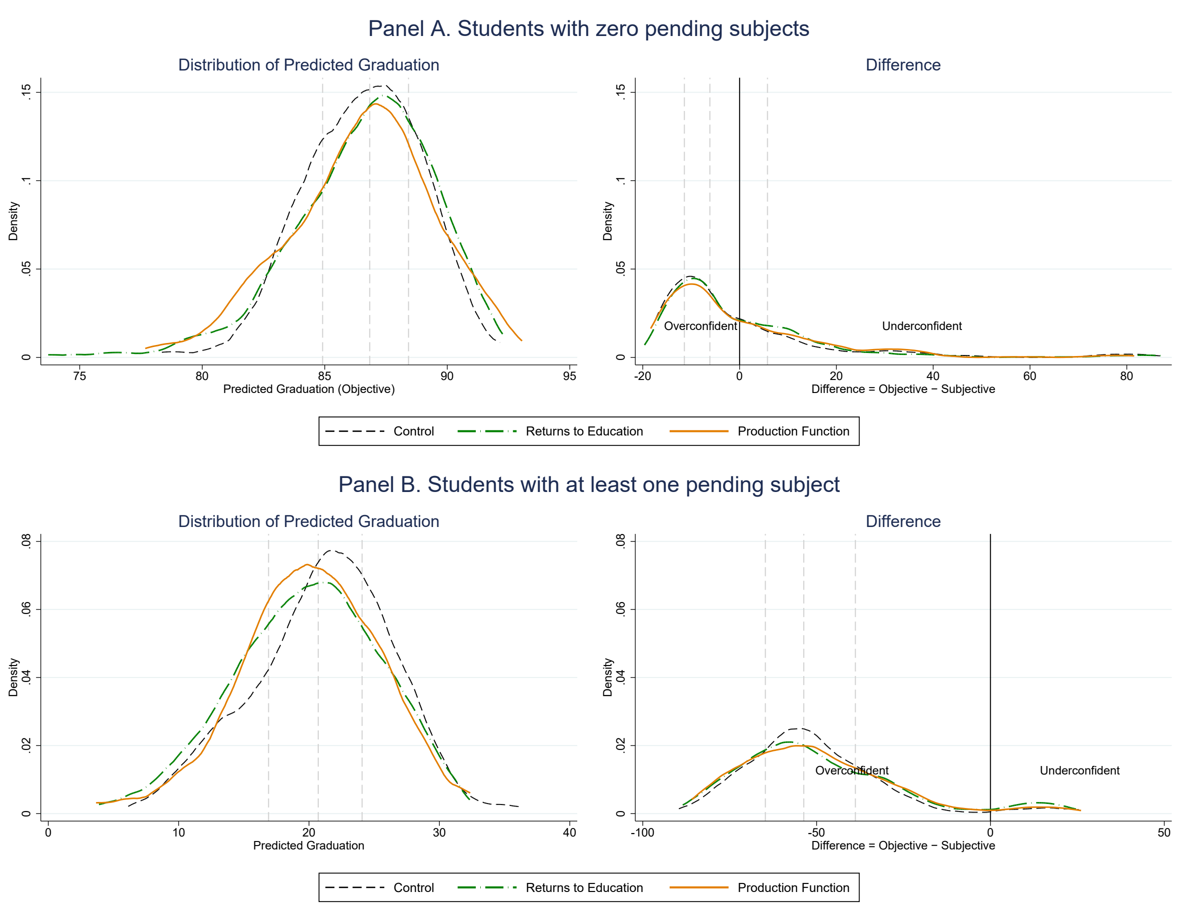

*Notes:* The intervention was designed for senior high school students in 2019. In 2018, discussions began with the Ministry of Education of Salta to define the scope of the intervention. The main survey instrument was tested during the first quarter of 2019. Subsequently, meetings were held with school authorities to obtain additional permissions. Visits to the schools began in August and concluded at the beginning of November. The intervention was conducted via one visit to each school, and the baseline questionnaire was administered at the beginning of each visit. The main outcome, graduation, was registered for each student in administrative records located in safe rooms in each school building. Data collection started in February 2020, after the last examination period to obtain the high school diploma on time, but it was interrupted due to the COVID-19 lockdown imposed in Argentina. Data collection concluded in March 2021.

Figure 3: The Intervention



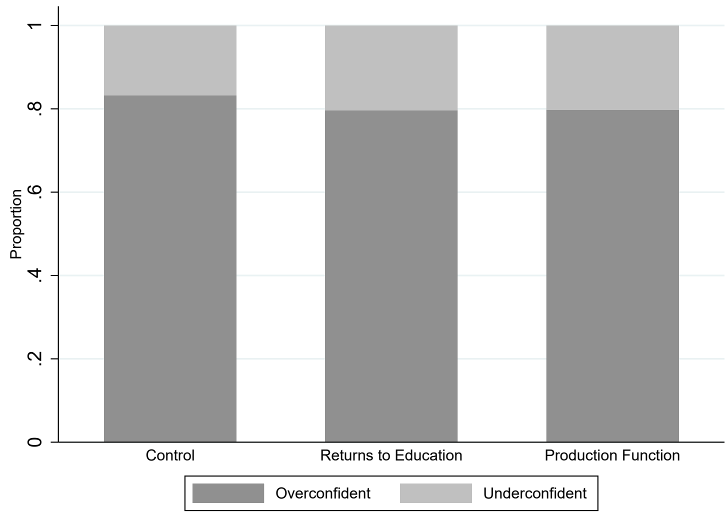

*Notes:* The intervention was conducted via one visit to each school and lasted no longer than one hour, as advised by school authorities. All senior students were gathered in one room. Activities included collecting a baseline survey from students at the beginning of the visit. Next, the research team demonstrated how to get access to a free online platform with Math content (including those students in the control group). Information interventions were delivered using slides to all students in schools randomly selected to receive each treatment arm. At the end of the visit, the question about students' perceptions of graduation was repeated to check for any updates after they received the information. The questionnaire underwent multiple rounds of testing at the start of the intervention, and several changes were made to the wording of the final question; a higher variability in responses was found using the format shown in Figure A3 in Appendix C.

Figure 4: Distribution of Predicted Graduation and Difference with Self-estimation by Treatment Group



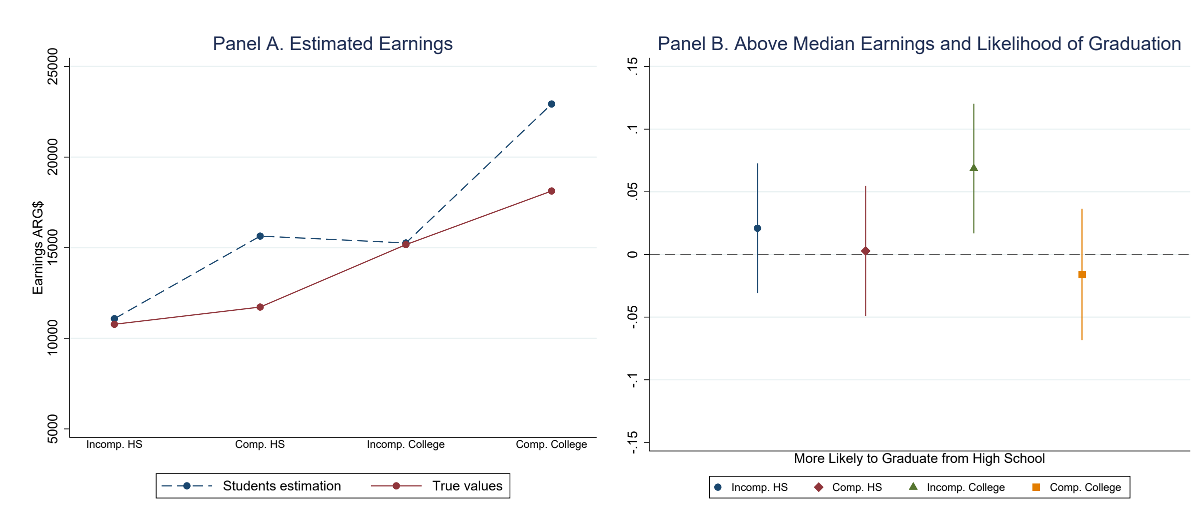

*Notes:* Graphs on the left show kernel density estimates of the distribution of predicted probability of graduation by treatment arm. Vertical dashed lines indicate 25th, 50th, and 75th percentiles of overall distribution, respectively. Graphs on the right show the difference between the predicted probabilities of Panel A and the self-reported beliefs of students about their own probabilities of graduation: a positive difference indicates that students underestimated their chances of graduation and a negative one that they overestimated their probabilities of graduation. Vertical dashed lines indicate 25th, 50th, and 75th percentiles of overall difference, respectively.

Figure 5: Overconfidence by Treatment Arm. All Students.




*Notes:* This graph shows proportions of overconfident-underconfident students computed according the classification shown in Figure 4, by treatment arm.

Figure 6: Perception of Earnings and Correlation with Likelihood to Graduate from High School




*Notes:* Panel A: Estimated earnings by students in the baseline survey. For each level of education, I coded as missing values responses below 5 percent and above 95 percent. Panel B: Separated regression coefficients of a dummy variable equal to 1 if the earnings' perception is above the median value on a dummy variable equal to 1 if the students are more likely to graduate from high school (above the median predicted value for high school graduation). Confidence intervals at the 95 percent level.

# Tables

Table 1: Descriptive Statistics from Control Group

| | (1) Mean | (2) N |
|---|---|---|
| *Panel A. All students* | | |
| Graduation (by February 2020) | 0.504 | 617 |
| Students' Graduation estimation at baseline | 0.784 | 615 |
| Students' Graduation estimation at endline | 0.842 | 601 |
| Number of pending subjects at the beginning of the senior year | 0.887 | 617 |
| *Panel B. Underconfident students* | | |
| Graduation (by February 2020) | 0.612 | 103 |
| Students' Graduation estimation at baseline | 0.569 | 101 |
| Students' Graduation estimation at endline | 0.740 | 101 |
| Number of pending subjects at the beginning of the senior year | 0.272 | 103 |
| *Panel C. Overconfident students* | | |
| Graduation (by February 2020) | 0.482 | 514 |
| Students' Graduation estimation at baseline | 0.826 | 514 |
| Students' Graduation estimation at endline | 0.863 | 500 |
| Number of pending subjects at the beginning of the senior year | 1.010 | 514 |

*Notes*: Sample of students in the control group. This table shows the performance under the status quo and the perceptions about own probability of graduation. Panel A shows the result for all students in the control group, Panel B restricts the sample to the students classified as underconfident, and Panel C shows the results for overconfident students. Students are classified as under- or overconfident following the definition shown in Subsection 4.6.

Table 2: Randomization Verification

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | | Regression Coefficients | | P-Value | | |
| | Control Mean | Returns to Education | Production Function | Joint test RE=PF | Joint test RE=PF=0 | N |
| *A. Sample Frame (School-shift)* | | | | | | |
| Number of Students | 30.9 | 0.1 | -4.66 | 0.296 | 0.441 | 61 |
| | [16.8] | (5.31) | (4.53) | | | |
| *B. Students Characteristics* | | | | | | |
| Age | 18 | -.028 | 0.022 | 0.69 | 0.921 | 1776 |
| | [0.968] | (0.145) | (0.12) | | | |
| Gender | 0.598 | -.001 | 0.016 | 0.611 | 0.861 | 1786 |
| | [0.491] | (0.029) | (0.034) | | | |
| Pregnancy/Has children | 0.06 | -.002 | -.002 | 0.975 | 0.987 | 1700 |
| | [0.237] | (0.013) | (0.013) | | | |
| Has email | 0.725 | 0.003 | 0.036 | 0.282 | 0.387 | 1767 |
| | [0.447] | (0.04) | (0.033) | | | |
| Has cellphone | 0.857 | -.006 | -.015 | 0.705 | 0.753 | 1771 |
| | [0.35] | (0.025) | (0.02) | | | |
| Lives with mother | 0.87 | -.007 | -.024 | 0.38 | 0.458 | 1786 |
| | [0.336] | (0.02) | (0.02) | | | |
| Lives with father | 0.58 | -.003 | -.037 | 0.094 | 0.132 | 1786 |
| | [0.494] | (0.021) | (0.021) | | | |
| *C. Households Characteristics* | | | | | | |
| Has computer | 0.761 | 0.027 | 0.011 | 0.505 | 0.585 | 1777 |
| | [0.427] | (0.026) | (0.025) | | | |
| Has internet access | 0.845 | -.006 | 0.019 | 0.211 | 0.384 | 1777 |
| | [0.362] | (0.024) | (0.02) | | | |
| Persons per room | 1.74 | -.069 | -.025 | 0.386 | 0.381 | 1759 |
| | [0.919] | (0.05) | (0.05) | | | |
| Parent has some superior educ. | 0.335 | -.01 | -.023 | 0.705 | 0.776 | 1786 |
| | [0.473] | (0.048) | (0.036) | | | |
| Student works or helps in the family business | 0.454 | -.009 | -.012 | 0.917 | 0.882 | 1786 |
| | [0.498] | (0.026) | (0.025) | | | |
| Student takes care of family members | 0.196 | 0.048 | 0.009 | 0.122 | 0.151 | 1786 |
| | [0.397] | (0.025) | (0.022) | | | |
| *D. Students Academic Performance* | | | | | | |
| Has repeated a year in high school | 0.384 | -.057 | -.064 | 0.893 | 0.401 | 1786 |
| | [0.487] | (0.061) | (0.047) | | | |
| At least one pending subject from previous years | 0.553 | -.037 | -.058 | 0.529 | 0.305 | 1786 |
| | [0.498] | (0.035) | (0.037) | | | |
| *E. Expectations* | | | | | | |
| Wants to attend college | 0.951 | -.028 | -.024 | 0.789 | 0.11 | 1786 |
| | [0.215] | (0.016) | (0.012) | | | |
| Wants to work after school | 0.874 | -.03 | -.034 | 0.792 | 0.158 | 1786 |
| | [0.333] | (0.019) | (0.018) | | | |
| Perceived probability of obtaining the diploma | 0.784 | 0.003 | 0.009 | 0.597 | 0.77 | 1783 |
| | [0.22] | (0.012) | (0.013) | | | |

*Notes*: Column 1 reports the number of non-missing observations of variables among all students in the control group. All regressions include graduation from the cohort 2018 at the school-shift level, and strata fixed effects.

Table 3: Impacts of Information on Graduation by Pending Subjects

| | (1) | (2) | (3) |
|---|---|---|---|
| | | Graduation | |
| | All | Zero Pending | At least One Pending |
| Production Function | 0.0528 | -0.0136 | 0.0730 |
| | (0.0241) | (0.0271) | (0.0271) |
| Returns to Education | 0.103 | 0.0422 | 0.125 |
| | (0.0255) | (0.0224) | (0.0319) |
| P-value: PF = RE | 0.038 | 0.010 | 0.124 |
| P-value: PF = RE = 0 | 0.000 | 0.016 | 0.000 |
| Mean (Control) | 0.50 | 0.87 | 0.21 |
| N | 1786 | 833 | 953 |

*Notes*: Robust standard errors clustered at the school-shift level in parentheses. All regressions include graduation from the cohort 2018 at the school-shift level, and strata fixed effects. Eligible controls include area of the city dummies, student age, student gender, if the student has children or is pregnant, average grades of classes during the first 2 quarters of the senior year, if the student has a job or takes care of a family member dummy, if the student repeated at least one year in secondary school, if her/his parent/guardian has some superior education, if the student does not live in a crowded dwelling, if in the household there is a computer, a washing machine, an AC, heating, and pairwise interactions between all previously-listed controls. Missing values are recoded to the sample mean and separately dummied out. These missing dummies are also used to construct pairwise interactions.

Table 4: Impacts of Information on Graduation by Confidence on Graduation

| | (1)<br>Graduation |
|---|---|
| Production Function × Overconfidence | 0.0300 |
| | (0.0287) |
| Production Function × Underconfidence | 0.0820 |
| | (0.0450) |
| Returns to Education × Overconfidence | 0.0920 |
| | (0.0298) |
| Returns to Education × Underconfidence | 0.115 |
| | (0.0461) |
| Overconfidence | -0.109 |
| | (0.0478) |
| P-value: PF × Overconfident = PF × Underconfident | 0.381 |
| P-value: RE × Overconfident = RE × Underconfident | 0.696 |
| P-value: PF × Overconfident = RE × Overconfident | 0.020 |
| P-value: PF × Underconfident = RE × Underconfident | 0.406 |
| Mean (Control, Underconfident) | 0.61 |
| N | 1786 |

*Notes*: Robust standard errors clustered at the school-shift level in parentheses. All regressions include graduation from the cohort 2018 at the school-shift level, and strata fixed effects. See notes in Table 3 for a list of potential controls.

Table 5: Impacts of Information on Performance

| | (1) Passed all senior subjects | (2) At least one pending subject left |
|---|---|---|
| *Panel A. No Interactions* | | |
| Production Function | 0.013 | -0.067 |
| | (0.024) | (0.028) |
| Returns to Education | 0.049 | -0.12 |
| | (0.022) | (0.032) |
| P-value: PF = RE | 0.152 | 0.095 |
| P-value: PF = RE = 0 | 0.074 | 0.000 |
| Mean (Control) | 0.65 | 0.79 |
| *Panel B. Interactions with Students' Confidence* | | |
| Production Function × Overconfidence | -0.0055 | -0.051 |
| | (0.030) | (0.029) |
| Production Function × Underconfidence | 0.050 | -0.39 |
| | (0.050) | (0.14) |
| Returns to Education × Overconfidence | 0.035 | -0.11 |
| | (0.024) | (0.036) |
| Returns to Education × Underconfidence | 0.093 | -0.32 |
| | (0.046) | (0.100) |
| Overconfidence | 0.0028 | -0.21 |
| | (0.038) | (0.055) |
| P-value: PF × Overconfident = PF × Underconfident | 0.378 | 0.017 |
| P-value: RE × Overconfident = RE × Underconfident | 0.257 | 0.078 |
| P-value: PF × Overconfident = RE × Overconfident | 0.183 | 0.090 |
| P-value: PF × Underconfident = RE × Underconfident | 0.405 | 0.620 |
| Mean (Control, Underconfident) | 0.64 | 1 |
| N | 1786 | 853 |

*Notes*: Robust standard errors clustered at the school-shift level in parentheses. All regressions include graduation from the cohort 2018 at the school-shift level, and strata fixed effects. See notes in Table 3 for a list of potential controls.

Table 6: Impacts of Information on Observable Effort

| | (1) Enrollment for Examination Period | (2) Attendance to Examination Period |
|---|---|---|
| *Panel A. No Interactions* | | |
| Production Function | 0.030 | 0.055 |
| | (0.065) | (0.036) |
| Returns to Education | 0.042 | 0.13 |
| | (0.074) | (0.039) |
| | | |
| P-value: PF = RE | 0.859 | 0.048 |
| P-value: PF = RE = 0 | 0.832 | 0.005 |
| | | |
| Mean (Control) | 0.62 | 0.44 |
| | | |
| *Panel B. Interactions with Students' Confidence* | | |
| Production Function × Overconfidence | 0.027 | 0.034 |
| | (0.066) | (0.038) |
| Production Function × Underconfidence | 0.020 | 0.46 |
| | (0.12) | (0.13) |
| Returns to Education × Overconfidence | 0.033 | 0.11 |
| | (0.072) | (0.041) |
| Returns to Education × Underconfidence | 0.11 | 0.38 |
| | (0.12) | (0.13) |
| Overconfidence | -0.087 | 0.21 |
| | (0.066) | (0.11) |
| | | |
| P-value: PF × Overconfident = PF × Underconfident | 0.958 | 0.002 |
| P-value: RE × Overconfident = RE × Underconfident | 0.449 | 0.058 |
| P-value: PF × Overconfident = RE × Overconfident | 0.931 | 0.031 |
| P-value: PF × Underconfident = RE × Underconfident | 0.514 | 0.518 |
| | | |
| Mean (Control, Underconfident) | 0.71 | 0.21 |
| N | 853 | 853 |

*Notes*: Robust standard errors clustered at the school-shift level in parentheses. All regressions include graduation from the cohort 2018 at the school-shift level, and strata fixed effects. See notes in Table 3 for a list of potential controls.

Table 7: Impacts on Graduation by Time Preferences

| | (1) Graduation |
|---|---|
| Production Function × Above Median | 0.0349 |
| | (0.0364) |
| Production Function × Below Median | 0.0394 |
| | (0.0371) |
| Returns to Education × Above Median | 0.117 |
| | (0.0347) |
| Returns to Education × Below Median | 0.0438 |
| | (0.0487) |
| Above Median Discount Factor | -0.0208 |
| | (0.0402) |
| P-value: RE × Above Median = RE × Below Median | 0.238 |
| P-value: PF × Above Median = PF × Below Median | 0.928 |
| Mean (Control, Below Median) | 0.56 |
| N | 1562 |

*Notes*: Robust standard errors clustered at the school-shift level in parentheses. All regressions include graduation from the cohort 2018 at the school-shift level, and strata fixed effects. To compute the dummy variable *Above Median Discount Factor* I classified the students under that category if the discount factor is higher than the median value of the variable discount factor today vs. one week. See notes in Table 3 for a list of potential controls.

Table 8: Impacts of Information on Graduation by Poverty Level and Gender

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Graduation | | | |
| | Poor students | Less poor students | Female students | Male students |
| Production Function | 0.0787 | 0.0421 | 0.0522 | 0.0747 |
| | (0.0289) | (0.0302) | (0.0323) | (0.0299) |
| Returns to Education | 0.144 | 0.0523 | 0.0982 | 0.112 |
| | (0.0303) | (0.0390) | (0.0352) | (0.0284) |
| P-value: PF = RE | 0.020 | 0.726 | 0.112 | 0.238 |
| P-value: PF = RE = 0 | 0.000 | 0.327 | 0.020 | 0.000 |
| Mean (Control) | 0.45 | 0.59 | 0.57 | 0.40 |
| N | 1109 | 677 | 1061 | 725 |

*Notes*: Robust standard errors clustered at the school-shift level in parentheses. All regressions include graduation from the cohort 2018 at the school-shift level, and strata fixed effects. To classify students as *Poor* or *Less Poor* I created an index variable that includes ownership of household items and a variable that indicates if at least one parent or guard has some college education. In total the index includes 6 dummy variables, if the score is lower or equal to 3 the student is classified as poor. See notes in Table 3 for a list of potential controls.

Table 9: Impacts of Information on Other Outcomes

| | (1) College Enroll-ment | (2) Formal Employ-ment |
|---|---|---|
| Production Function | 0.0518 | -0.0144 |
| | (0.0273) | (0.00868) |
| Returns to Education | 0.0543 | -0.0224 |
| | (0.0235) | (0.00764) |
| P-value: PF = RE | 0.909 | 0.227 |
| P-value: PF = RE = 0 | 0.059 | 0.012 |
| Mean (Control) | 0.13 | 0.032 |
| N | 1786 | 1348 |

*Notes*: Robust standard errors clustered at the school-shift level in parentheses. All regressions include graduation from the cohort 2018 at the school-shift level, and strata fixed effects. College is a dummy variable equal to 1 that indicates if the student is formally enrolled in at least one college of Salta during 2020 (Universidad Nacional de Salta or Universidad Catolica de Salta). Formal employment is a dummy variable equal to one if the student was employed in the formal sector at least one month during the last quarter of 2020 and the first quarter of 2021. See notes in Table 3 for a list of potential controls.

Table 10: Impacts of Information on Graduation and College Enrollment by Likelihood of High School Graduation Based on Observable Characteristics at Baseline

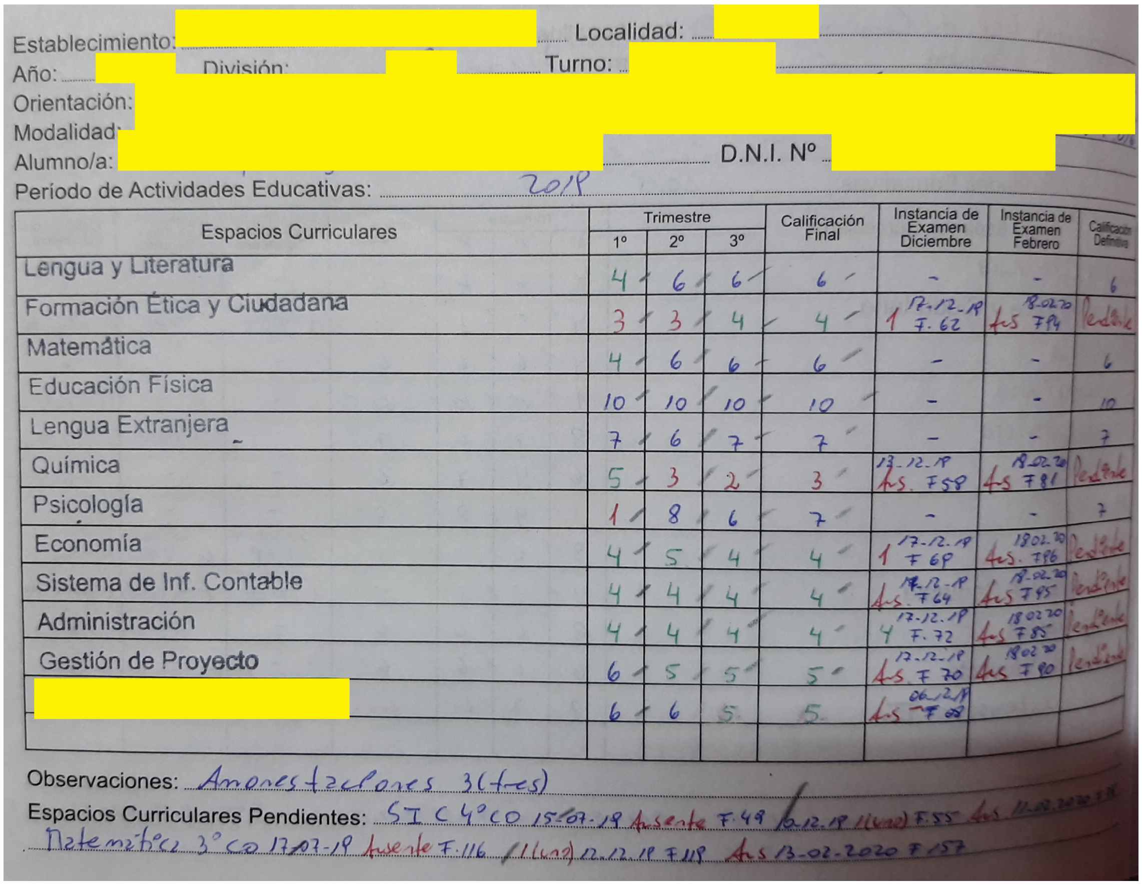
| | (1) Graduation | (2) College Enrollment |
|---|---|---|
| Production Function × More Likely | -0.00420 | 0.0865 |
| | (0.0298) | (0.0505) |
| Production Function × Less Likely | 0.0708 | 0.0153 |
| | (0.0316) | (0.0195) |
| Returns to Education × More Likely | 0.0581 | 0.0453 |
| | (0.0292) | (0.0426) |
| Returns to Education × Less Likely | 0.105 | 0.0607 |
| | (0.0328) | (0.0222) |
| More Likely to Graduate | 0.434 | 0.0229 |
| | (0.0379) | (0.0334) |
| P-value: PF × More Likely = PF × Less Likely | 0.115 | 0.176 |
| P-value: RE × More Likely = RE × Less Likely | 0.298 | 0.726 |
| P-value: PF × More Likely = RE × More Likely | 0.009 | 0.289 |
| P-value: PF × Less Likely = RE × Less Likely | 0.374 | 0.039 |
| Mean (Control, Less Likely) | 0.19 | 0.059 |
| N | 1786 | 1786 |

*Notes*: Robust standard errors clustered at the school-shift level in parentheses. See notes in Table 3 for a list of potential controls. College is a dummy variable equal to 1 that indicates if the student is formally enrolled in at least one college of Salta during 2020 (Universidad Nacional de Salta or Universidad Catolica de Salta). *More likely to Graduate* is a dummy variable, I classified the students under that category if the prediction of likelihood to graduate from high school is higher than its median value, see Subsection 4.6.

# A Appendix: Additional Information

## Figures

Figure A1: Example of Student Academic Report.



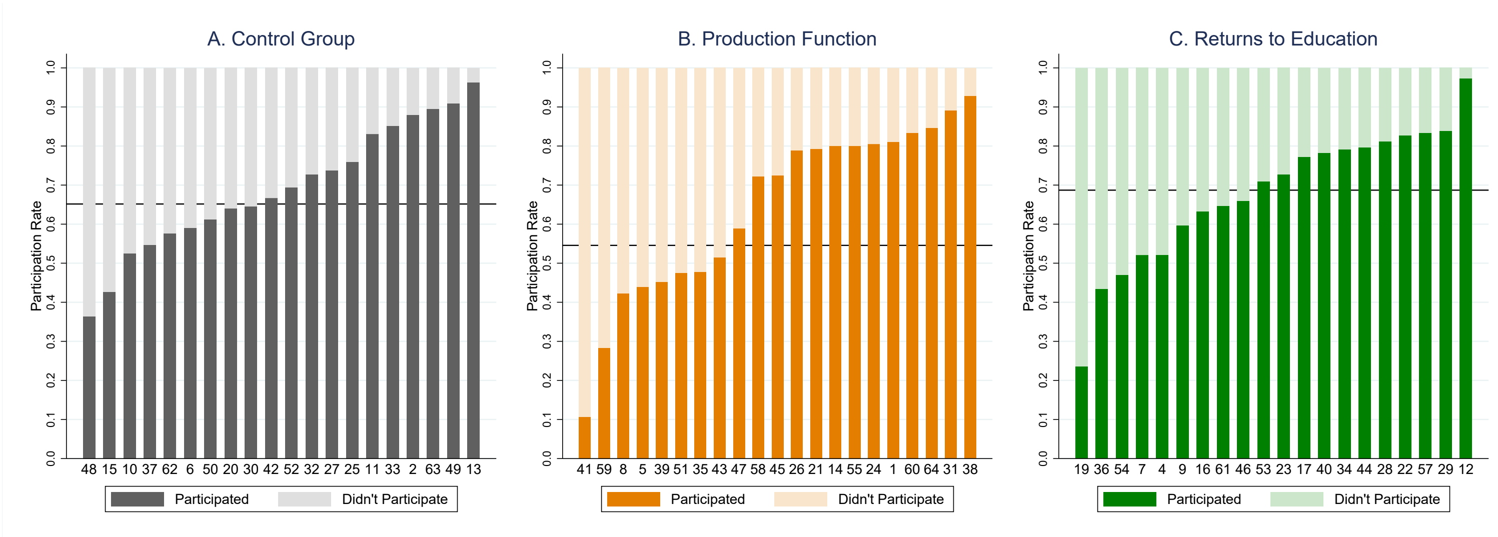
Establecimiento: [redacted] Localidad: [redacted]
Año: [redacted] División: [redacted] Turno: [redacted]
Orientación: [redacted]
Modalidad: [redacted]
Alumno/a: [redacted] D.N.I. Nº [redacted]
Período de Actividades Educativas: 2018

| Espacios Curriculares | Trimestre 1º | Trimestre 2º | Trimestre 3º | Calificación Final | Instancia de Examen Diciembre | Instancia de Examen Febrero | Calificación Definitiva |
|---|---|---|---|---|---|---|---|
| Lengua y Literatura | 4 | 6 | 6 | 6 | - | - | 6 |
| Formación Ética y Ciudadana | 3 | 3 | 4 | 4 | 1 17-12-18 F. 62 | Aus 18-02-20 F94 | Pendiente |
| Matemática | 4 | 6 | 6 | 6 | - | - | 6 |
| Educación Física | 10 | 10 | 10 | 10 | - | - | 10 |
| Lengua Extranjera | 7 | 6 | 7 | 7 | - | - | 7 |
| Química | 5 | 3 | 2 | 3 | 13-12-18 Aus. F58 | 18-02-20 Aus F81 | Pendiente |
| Psicología | 1 | 8 | 6 | 7 | - | - | 7 |
| Economía | 4 | 5 | 4 | 4 | 1 17-12-18 F 69 | 18-02-20 Aus. F96 | Pendiente |
| Sistema de Inf. Contable | 4 | 4 | 4 | 4 | 17-12-18 Aus. F64 | 18-02-20 Aus F95 | Pendiente |
| Administración | 4 | 4 | 4 | 4 | 4 17-12-18 F. 72 | 18 02 20 Aus F85 | Pendiente |
| Gestión de Proyecto | 6 | 5 | 5 | 5 | 17-12-18 Aus F 70 | 18 02 20 Aus F 90 | Pendiente |
| [redacted] | 6 | 6 | 5 | 5 | 06-12-18 Aus F 08 | | |
| | | | | | | | |

Observaciones: Amonestaciones 3 (tres)
Espacios Curriculares Pendientes: SIC 4º CO 15-07-18 Ausente F. 49 10-12-18 1 (uno) F. 55 Aus 11-02-2020 F 1[illegible]
Matemática 3º CO 17-07-18 Ausente F. 116 1 (uno) 12-12-18 F 118 Aus 13-02-2020 F 157

*Notes:* The format is similar in all secondary schools. The top of each record registers information about the school, shift, academic year, and student's personal information. The middle section lists all the mandatory subjects during senior year. Next to each name, the 3 following columns show the grades for quarters 1, 2, and 3, then the final grade (notice it is not an average of the quarters). If the student didn't pass a subject during the academic year, the next two columns are used to register attendance to the examination periods of December 2019 and February 2020, and the last columns indicate the definitive grade. At the bottom of the record, there is a space for general observations and a dedicated space to register existing pending subjects (if any) and enrollment to examination periods (with dates), attendance, and grades.

## Figure A2: Participation Rates at the School Level

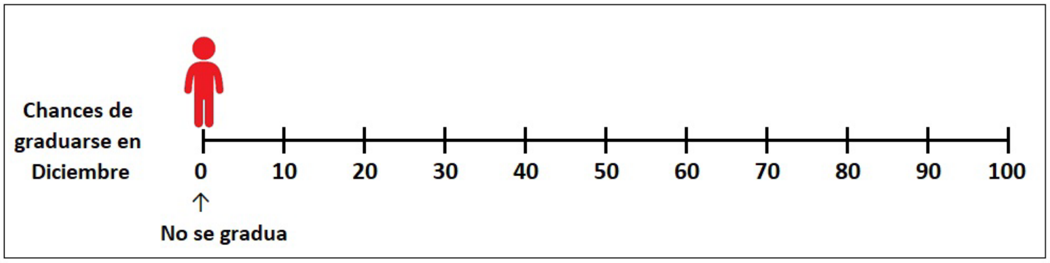



*Notes:* Horizontal axis shows random numbers assigned to each school. In each panel, the horizontal black lines indicates the participation rate for the entire treatment arm.

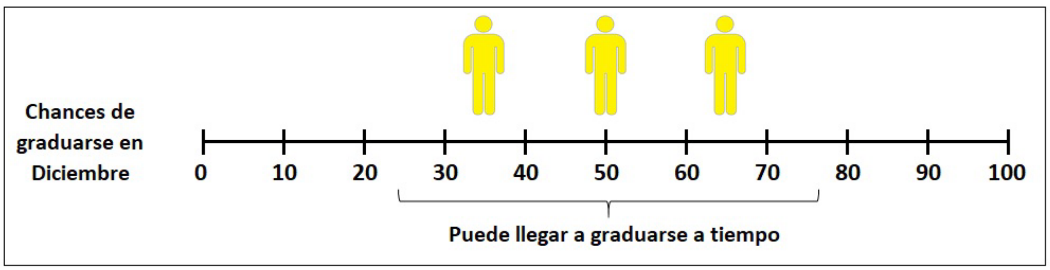
Figure A3: Prompts used to ask own probability of graduation

***Probability***: *It is a number that indicates how likely an event is to occur, in general it is expressed as a percentage of 0 to 100. For example, what do you think is the probability that a 5th year student receives his or her high school in December? After the exam dates of that month.*

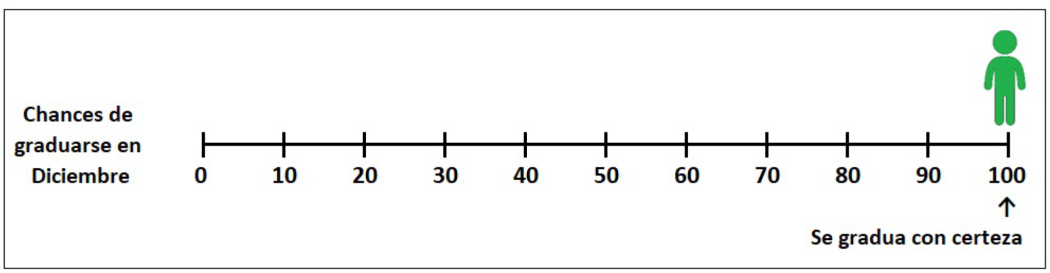
*Example 1*: A student who does not study, frequently skips classes. Has pending subjects and does not attend the exam periods, who does not pass all the subjects this year, has a 0% probability of receiving the diploma in December.




*Example 2*: A student who studies sometimes, sometimes skips classes, has some pending subjects, has a chance to receive the diploma on time.



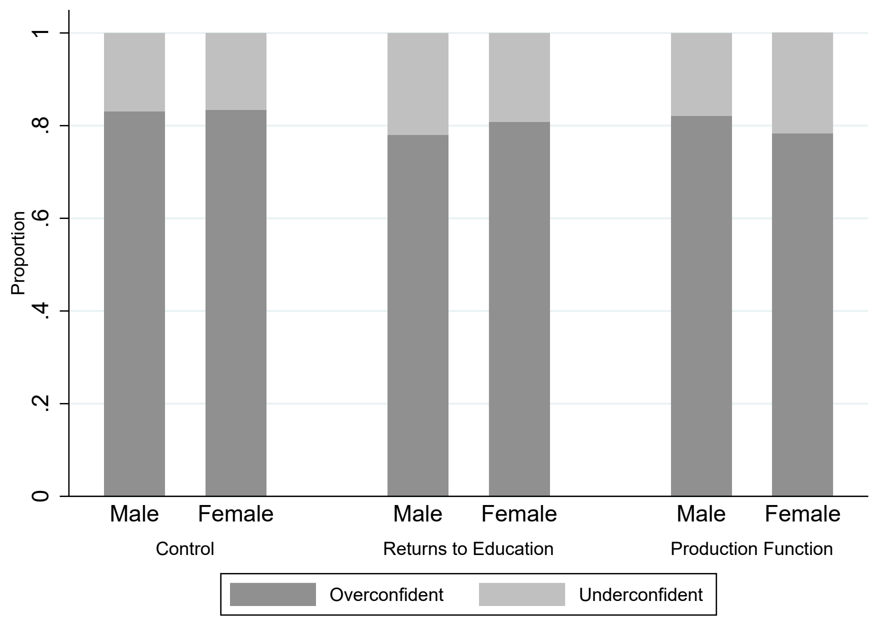

*Example 3*: A student who always studies, never skips classes, does not have pending subjets, with grade 10 in all subjects this year, has a 100% probability of receiving the diploma.




***What are your chances of receiving the high school diploma in December? Insert a value from 0 to 100: ______***

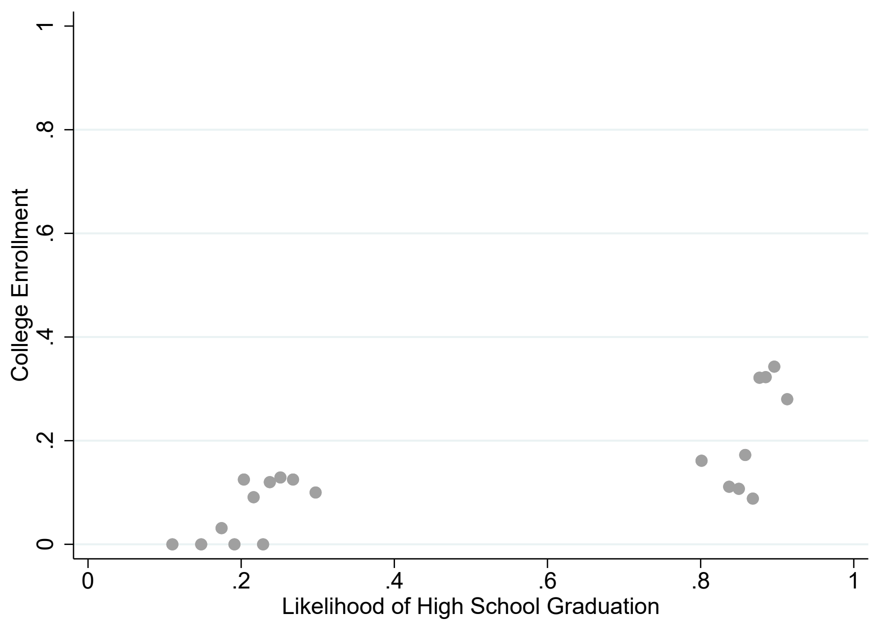
*Notes:* First, students were shown a concept of probability, and I provided 3 examples. Although this could be anchoring the beliefs of some students, during the piloting phase using more abstract concepts (or applied to other settings) was not helpful for students. At the end of the figure, I show the question used to ask about perceptions of their probabilities of graduation.

Figure A4: Overconfidence by Treatment Arm and Gender




*Notes:* Proportions of overconfident-underconfident students computed according to the classification shown in Figure 4.

Figure A5: Likelihood of High School Graduation and College Enrollment, Control Group – Binned Scatter




*Notes:* Sample limited to the control group. Data on college is actual college enrollment during the next academic year of my intervention and likelihood of high school graduation is the prediction estimated in Subsection 4.6. The graph shows the correlation between college enrollment and estimated likelihood of high school completion.

## Tables

Table A1: Selection into Participation

| | (1)<br>Participated | (2)<br>Participated<br>w/o 1 school |
|---|---|---|
| Production Function | -0.0985 | -0.0257 |
| | (0.0548) | (0.0353) |
| Returns to Education | 0.0183 | 0.0554 |
| | (0.0481) | (0.0419) |
| P-value: PF = RE | 0.008 | 0.034 |
| P-value: PF = RE = 0 | 0.028 | 0.103 |
| Mean (Control) | 0.65 | 0.65 |
| N | 2856 | 2688 |

*Notes*: Robust standard errors clustered at the school-shift level in parentheses. All regressions include strata fixed effects. Column (2) does not include the school with the lowest participation rate (see Figure A2).

Table A2: Impacts of Information on Graduation by Pending Subjects – No Additional Controls

| | (1) | (2) | (3) |
|---|---|---|---|
| | | Graduation | |
| | All | Zero Pending | At least One Pending |
| Production Function | 0.0465 | -0.0213 | 0.0509 |
| | (0.0306) | (0.0338) | (0.0331) |
| Returns to Education | 0.0827 | 0.0314 | 0.0935 |
| | (0.0319) | (0.0324) | (0.0354) |
| P-value: PF = RE | 0.214 | 0.141 | 0.286 |
| P-value: PF = RE = 0 | 0.041 | 0.325 | 0.029 |
| Mean (Control) | 0.50 | 0.87 | 0.21 |
| N | 1786 | 833 | 953 |

*Notes*: Robust standard errors clustered at the school-shift level in parentheses. All regressions include graduation from the cohort 2018 at the school-shift level, and strata fixed effects.

Table A3: Impacts of Information on Graduation by Pending Subjects – Excluding Observations from the School with Lowest Participation Rate

| | (1) | (2) | (3) |
|---|---|---|---|
| | | Graduation | |
| | All | Zero Pending | At least One Pending |
| Production Function | 0.0607 | -0.00411 | 0.0770 |
| | (0.0250) | (0.0252) | (0.0279) |
| Returns to Education | 0.108 | 0.0500 | 0.127 |
| | (0.0259) | (0.0215) | (0.0321) |
| P-value: PF = RE | 0.049 | 0.012 | 0.138 |
| P-value: PF = RE = 0 | 0.000 | 0.012 | 0.000 |
| Mean (Control) | 0.50 | 0.87 | 0.21 |
| N | 1768 | 823 | 945 |

*Notes*: Robust standard errors clustered at the school-shift level in parentheses. All regressions include graduation from the cohort 2018 at the school-shift level, and strata fixed effects. See notes in Table 3 for a list of potential controls.

Table A4: Impacts of Information on Self-estimated Probability of Graduation (after-before intervention)

| | (1) | (2) | (3) |
|---|---|---|---|
| | Difference: Confidence update | | |
| | All | Under-confident Students | Over-confident Students |
| Production Function | -2.049 | -0.276 | -2.409 |
| | (0.883) | (3.197) | (0.950) |
| Returns to Education | 0.546 | 2.431 | -0.521 |
| | (0.922) | (3.199) | (0.892) |
| P-value: PF = RE | 0.004 | 0.265 | 0.075 |
| P-value: PF = RE = 0 | 0.008 | 0.503 | 0.038 |
| Mean (Control) | 5.77 | 16.8 | 3.57 |
| N | 1765 | 336 | 1429 |

*Notes*: Robust standard errors clustered at the school-shift level in parentheses. All regressions include graduation from the cohort 2018 at the school-shift level, and strata fixed effects. See notes in Table 3 for a list of potential controls.

Table A5: Difference by Missing Employment Data

| | (1)<br>Dummy Missing<br>Employment |
|---|---|
| Production Function | 0.0453 |
| | (0.110) |
| Returns to Education | 0.0685 |
| | (0.0890) |
| P-value: PF = RE | 0.827 |
| P-value: PF = RE = 0 | 0.741 |
| Mean (Control) | 0.19 |
| N | 1786 |

*Notes*: Robust standard errors clustered at the school-shift level in parentheses. All regressions include graduation from the cohort 2018 at the school-shift level, and strata fixed effects. See notes in Table 3 for a list of potential controls.

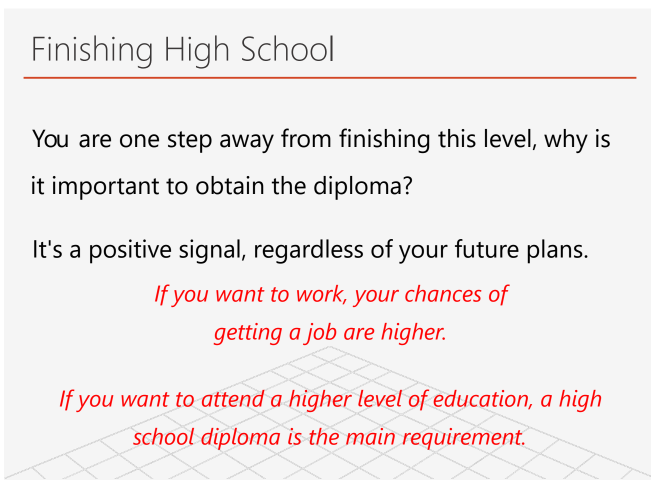
# B Appendix: Information Treatment Arms

## B.1 Information Interventions

I show the specific content introduced to the senior students that participated in each treatment arm. For both treatment arms, I discussed why it is important to finish high school, highlighting the fact that they already spent almost 5 years attending this level and that only a small fraction of the students that enter their senior year drop out at some point during the year (Anuarios Estadísticos, Ministerio de Educación de la Nación). See Figure B1.

Each information intervention was delivered after the free online platform was introduced to the students (Appendix C, section C.2). In total, the presentation lasted 40 minutes.

Figure B1: Why to Obtain the Diploma



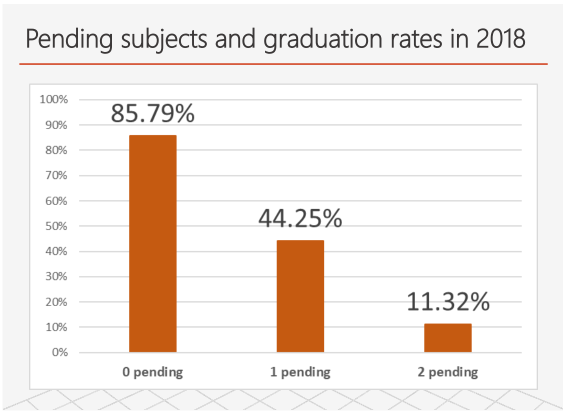

*Notes:* Common slide showed to all the students who received any of the intervention treatments.

### Production Function

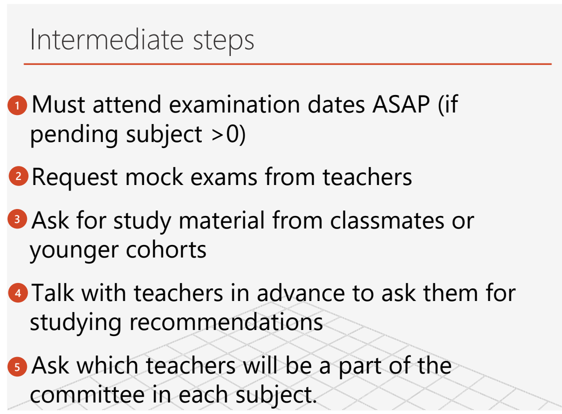
I showed information about graduation rates from the previous cohort (students who were seniors during the 2018 academic year). It was intended to emphasize how important it was for students to pass their pending subjects during their senior year. It underlined the pervasive effects of having pending subjects on the probability of obtaining a diploma. To construct these statistics, I asked the Directorate of Secondary Education for access to the academic records of "representative" schools. They asked school principals for permission before sending me a list of the schools with contacts who could give me access to their academic records. As mentioned previously, there was no previous information available about the correlation between pending subjects and graduation. Based on the sample I collected, I elaborated the statistics that were shown to the students (see Figure B2a).

Figure B2: Slides from the Production Function Arm

(a) Statistics Shown to the Students




(b) Tips to Remedy Academic Standing




*Notes:* Own estimations based on a sample of representative schools in the capital city of Salta including students who were seniors during the 2018 academic year.

Each student was aware of their own situation, but during the presentation, I could not observe their academic standing (number of pending subjects). The idea of showing these numbers was to help them create a mapping of their situation at the beginning of the senior year and how similar students performed in terms of graduation. Given that this could have been shocking news for students regardless of standing, I talked about the intermediate steps needed to transform inputs into outputs and I discussed how to remedy their situation: first, I opened a discussion of the options together, and then I showed a summary of the most relevant tips to effectively obtain a diploma on time (Figure B2b).

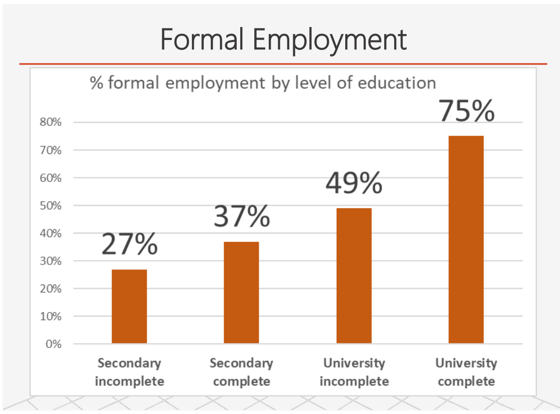
The key messages were (1) to devote more time and effort to studying students' senior year subjects and (2) for those with pending subjects, to attend the examination periods. Students' senior year includes several social activities (prom night, private parties, graduation trip, etc.). In interviews with the school principals and in some focus groups with students from the previous cohorts, these activities were mentioned as major distractions from academics.

**Returns to Education**

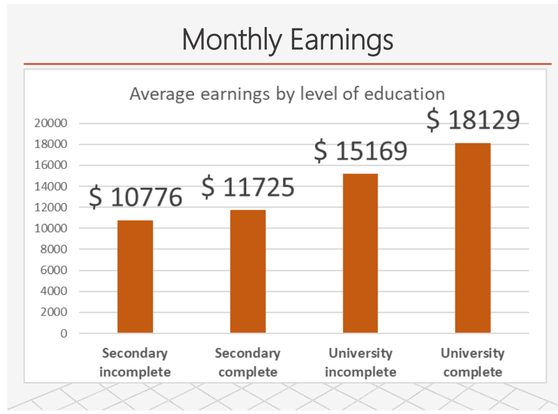
In this presentation, I used data from the National Household Survey 2018 (Encuesta Permanente de Hogares) to compute the averages of formal employment and earnings to be shown to the students. I only considered individuals from the province of Salta, between 18 and 30 years old. The statistics were computed according to the level of education and are shown in Figure B3.

Figure B3: Slides from the Returns to Education Arm

(a) Formal Employment by Level of Education




(b) Monthly Wages by Level of Education




*Notes:* Own estimations based on Encuesta Permanente de Hogares, 2018 (this survey only covers urban areas). Mincer equation was estimated considering age, gender, and marital status. After the presidential primaries of August 2019, the dollar became unstable but on average during October 2019, the exchange rate was \$1US ≈ \$64ARG.

## B.2 Reminders

Given that the intervention only included a single visit to each school, reminders via cellphone or e-mail were sent between 1 and 2 weeks before the December 2019 and February 2020 examination periods. This step was determined in the protocol approved by the Brown IRB and specified in the pre-analysis plan. The length of text messages was limited to 150 characters in Spanish (imposed by a private firm used to send the messages). To ensure a comparable reception of both reminders, the e-mail was also shortened. Both messages were sent if a student self-reported a valid cellphone number and/or e-mail address.

### Returns to Education Reminders

- SMS: Hi! Remember that a higher level of education increases the chances of finding a quality job and a higher salary!

  Team UNSa-Brown

- e-mail: Hi! In our visit to your school we showed you information about the labor market in Salta. Remember, a higher level of education increases the probability of finding a quality job and a higher salary!

Team UNSa-Brown

**Production Function Reminders**

- SMS: Hi! If you failed subjects this year or have pending subjects, remember, it is important to attend the available exam dates and pass them!

  Team UNSa-Brown

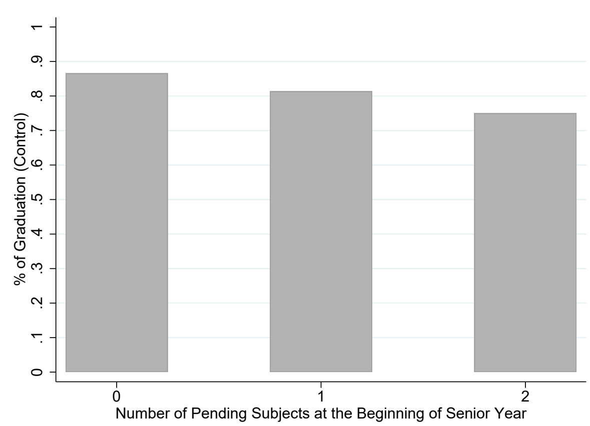
- e-mail: Hi! In our visit to your school we showed you that it is important to pass pending and subjects you failed this year as soon as possible. If you have failed subjects, remember to attend the available exam dates and study to pass them!

  Team UNSa-Brown

## B.3 Discussion about the Production Function

A potential concern on the design of the $PF$ treatment is that it could make students believe that moving from two to zero pending subjects will increase their probability of graduation by 74 percentage points (Figure B2a). In this context, deception will be present if passing the subjects is not enough to graduate, but passing those subjects is one requirement besides passing the senior subjects.

Figure B4: Graduation of Students who Passed All their Pending Subjects by the End of Academic Year. Control Group




*Notes:* Even if students pass their pending subjects, they could fail senior subjects and do not graduate.

I use the control group to observe changes in the probability of graduation, considering the subset of students who had pending subjects but passed them by the end of the academic year. Figure B4 shows the graduation conditional on the number of pending subjects the students had at the beginning of the senior year. This subset of students passed their pending subjects and now moved to the "good standing bin."

After passing their pending subjects, I observe that the probability of graduation for those with 1 and 2 pending subjects is close to 80 percent, similar to the magnitude shown to the students in the *PF* arm. This evidence helps to rule out concerns about deceiving students in this treatment arm.

## B.4 Other planned interventions

The *PF* and *RE* treatments were cross-cut with two interventions. The first intervention, randomized at the school level, offered after-school math classes to help students prepare for the next examination period. Due to budget constraints, only one location was opened to provide this service. The tutors were UNSa math professors, but the office was located in the North of the city, not accessible for most of the students. While this free service was offered to the selected schools at the time of the school visits (35 percent of participants), some students complained about the distance from their schools. As expected, only some students attended these classes (5 percent of selected students). They were students from schools located nearby the location. Due to the lack of strategic locations and the low attendance of students, this treatment is not analyzed.

The other intervention consisted of a randomization at the individual level to inform or remind students via SMS and/or email about the availability of scholarships sponsored by the national and provincial government for college attendance. The randomization was conditional on having a cellphone or an email address, and this information was collected during the baseline survey. The messages were initially intended to be sent in November/December of 2019, but the deadlines and specific requirements for applying for these scholarships were not made public at that time. Additionally, there was a delay due to the COVID-19 outbreak. As a result, the messages were sent at the beginning of February 2020, with a reminder in March and a message informing about the deadline extension for the national scholarship in April. Although the message sent in February 2020 could have had an impact on high school graduation, the information on scholarships did not affect this outcome or college enrollment. There is also no significant evidence that this treatment impacted *PF* and *RE* treatment effects (see Tables B1 and B2).

Table B1: Impacts of Information about Scholarships on High School Graduation and College Enrollment

| | (1) Graduation | (2) College Enrollment |
|---|---|---|
| *Panel A. No Controls* | | |
| Scholarships Information | -0.022 | -0.003 |
| | (0.021) | (0.019) |
| Observations | 1618 | 1618 |
| *Panel B. With Controls* | | |
| Scholarships Information | -0.027 | -0.005 |
| | (0.018) | (0.017) |
| Mean (No Scholarships Info) | 0.550 | 0.151 |
| N | 1618 | 1618 |

*Notes*: Robust standard errors clustered at the school-shift level in parentheses. All regressions include graduation from the cohort 2018 at the school-shift level, and strata fixed effects. For Panel B, see notes in Table 3 for a list of potential controls.

Table B2: Impacts of Information about Scholarships on High School Graduation and College Enrollment

| | (1) Graduation | (2) College Enrollment |
|---|---|---|
| Scholarships Information | -0.017 | 0.025 |
| | (0.025) | (0.031) |
| Production Function | 0.062 | 0.069 |
| | (0.032) | (0.031) |
| Returns to Education | 0.11 | 0.084 |
| | (0.034) | (0.029) |
| PF $\times$ Scholarships Information | -0.017 | -0.039 |
| | (0.040) | (0.043) |
| RE $\times$ Scholarships Information | -0.014 | -0.054 |
| | (0.045) | (0.043) |
| Mean (No RE, No PF, No Scholarships) | 0.523 | 0.123 |
| N | 1618 | 1618 |

*Notes*: Robust standard errors clustered at the school-shift level in parentheses. All regressions include graduation from the cohort 2018 at the school-shift level, and strata fixed effects.. See notes in Table 3 for a list of potential controls.

# C Appendix

## C.1 Statistical Power

To compute the statistical power, I used data from the previous cohort (2018, subsample of five schools), and I focused only on the information interventions. Given the small number of clusters, I was not able to include the interaction of the treatments. By considering three groups (control, returns to education, and production function), with a graduation rate in the control group of 50 percent, $\alpha = 0.05$, average cluster size of 47 students, ICC=0.05 (computed using data from that subsample), I am able to make comparisons between the two main treatments by estimating an effect of 3.5 percentage points in graduation rate with a statistical power of 76 percent.

## C.2 Free Online Platform: MOODLE

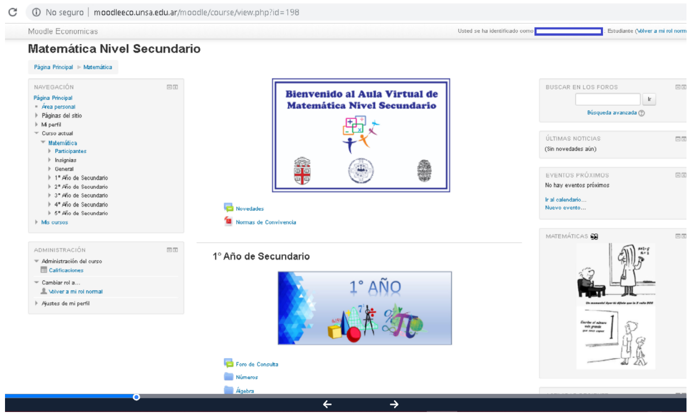
The Directorate of Secondary Education of Salta required that I provide some useful information to all students; otherwise, I would encounter resistance from school principals reluctant to give me access to their schools. So, to provide something in exchange for their participation, I designed a free online platform with math content for all the years of high school. This platform could help to improve the academic standing of students in that subject.

At the onset of the project I had two rounds of meetings with principals, vice principals, and senior-level math teachers to hear their opinions about my agreement with the directorate and to incorporate their feedback. The agreement was that the software would use material sent directly from math teachers. I partnered with the Department of Mathematics at the Faculty of Economics at Universidad Nacional de Salta to unify the content and create new material useful to all students from public schools. In addition to this material, professors of mathematics at UNSa, offered office hours to senior students from the participant schools (online).

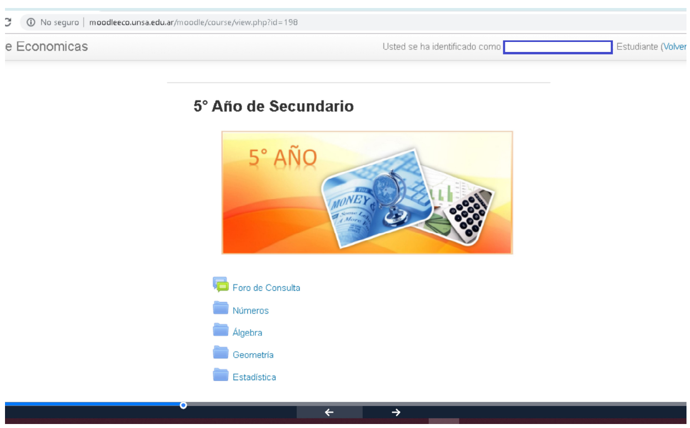
As mentioned above, the platform is not a part of the intervention, but rather enabled me to conduct the baseline surveys in all schools. After being introduced, we first explained the contents of the platform and then gave instructions on how to obtain free access (for security reasons, a unique code was determined for each school). Figure C1 shows the homepage of the platform (Panel A), with all the content year by year, Panel B shows a representative image of the content available by topics covered during students' senior year, and Panel C shows files with the available material.

Figure C1: MOODLE Platform

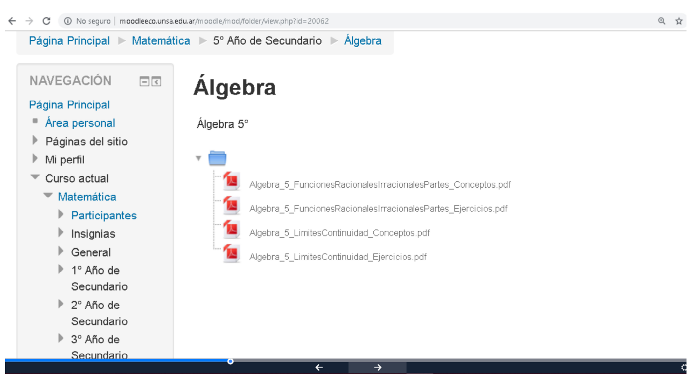
(a) Homepage




(b) Senior Year Overview




(c) Senior Year Specific Content




*Notes:* Screenshots of the platform designed by the Department of Mathematics at Faculty of Economics (UNSa).

## C.3 Full Derivatives: Model with Uncertainty

The maximization problem the student faces is:

$$\left[\hat{p}g\left(\hat{\beta}_l e+\hat{\alpha_h}\right)+(1-\hat{p})\,g\left(\hat{\beta}_h e+\hat{\alpha}_l\right)\right]\hat{V}-\delta e$$

with FOC:

$$\left[\hat{p}g'\left(\hat{\beta}_l e+\hat{\alpha_h}\right)\hat{\beta}_l+(1-\hat{p})\,g'\left(\hat{\beta}_h e+\hat{\alpha}_l\right)\hat{\beta}_h\right]\hat{V}-\delta=0$$

**Proof.** Production Function

$$\begin{aligned}g'\left(\hat{\beta}_l e+\hat{\alpha_h}\right)\hat{\beta}_l+\hat{p}g''\left(\hat{\beta}_l e+\hat{\alpha_h}\right)\left(\hat{\beta}_l\right)^2\frac{de^*}{d\hat{p}}&\\ -g'\left(\hat{\beta}_h e+\hat{\alpha}_l\right)\hat{\beta}_h+(1-\hat{p})\,g''\left(\hat{\beta}_h e+\hat{\alpha}_l\right)\left(\hat{\beta}_h\right)^2\frac{de^*}{d\hat{p}}&=0\end{aligned}$$

$$\frac{de^*}{d\hat{p}}=\frac{-g'\left(\hat{\beta}_l e+\hat{\alpha_h}\right)\hat{\beta}_l+g'\left(\hat{\beta}_h e+\hat{\alpha}_l\right)\hat{\beta}_h}{\hat{p}g''\left(\hat{\beta}_l e+\hat{\alpha_h}\right)\left(\hat{\beta}_l\right)^2+(1-\hat{p})\,g''\left(\hat{\beta}_h e+\hat{\alpha}_l\right)\left(\hat{\beta}_h\right)^2}\lesseqgtr 0$$

the second derivative of $g(.)$ is negative, but the sign of the numerator cannot be determined without additional assumptions about $g(.)$ function and the parameters of relevance. ■

**Proof.** Returns to Education

$$\begin{aligned}\hat{p}g'\left(\hat{\beta}_l e+\hat{\alpha_h}\right)\hat{\beta}_l+(1-\hat{p})\,g'\left(\hat{\beta}_h e+\hat{\alpha}_l\right)\hat{\beta}_h+&\\ \hat{p}g''\left(\hat{\beta}_l e+\hat{\alpha_h}\right)\left(\hat{\beta}_l\right)^2\frac{de^*}{d\hat{V}}+(1-\hat{p})\,g''\left(\hat{\beta}_h e+\hat{\alpha}_l\right)\left(\hat{\beta}_h\right)^2\frac{de^*}{d\hat{V}}&=0\end{aligned}$$

$$\frac{de^*}{d\hat{V}}=-\frac{\hat{p}g'\left(\hat{\beta}_l e+\hat{\alpha_h}\right)\hat{\beta}_l+(1-\hat{p})\,g'\left(\hat{\beta}_h e+\hat{\alpha}_l\right)\hat{\beta}_h}{\hat{p}g''\left(\hat{\beta}_l e+\hat{\alpha_h}\right)\left(\hat{\beta}_l\right)^2+(1-\hat{p})\,g''\left(\hat{\beta}_h e+\hat{\alpha}_l\right)\left(\hat{\beta}_h\right)^2}$$

By assumption, the second derivative of the $g(.)$ function is negative, so the entire denominator is negative. The numerator is positive (also by assumption). This means that the entire expression is positive. ■