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

Carolina Lopez*

January 7, 2025

In low- and middle-income countries, school completion rates remain low, making it critical to motivate higher levels of completion among students who reach this stage. This study uses a field experiment with 1,800 Argentine senior students to examine how providing targeted information during the last semester of their academic year—a critical period with limited opportunities to improve outcomes—can correct inaccurate beliefs about their chances of timely graduation and the economic value of education, ultimately increasing graduation rates. The first treatment provides information about the conditional probability of graduating, aiming to adjust misperceptions about students' chances of success and help them allocate effort more effectively during the academic year. The second treatment benchmarks this approach by offering information on the economic returns to education. The interventions increase high school graduation rates by 10% and 20%, respectively, compared to a control group graduation rate of 50%. Both treatments also improve college enrollment by 38%, compared to 13% in the control group. Notably, the improvement in graduation rates is most pronounced among students with the lowest baseline probabilities of graduating. Compared to other interventions aimed at improving educational outcomes in developing countries, this approach is highly cost-effective. Closing critical information gaps during the "last mile" of secondary schooling has transformative potential, particularly for the most disadvantaged students, offering a scalable strategy to improve lifetime outcomes.

JEL Codes: 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 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. I acknowledge the financial support I received 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). Completing high school is not only a critical signal to the labor market (Spence, 1973), but it also positively impacts wage prospects (Heckman and LaFontaine, 2010). In Argentina, for instance, while over 95 percent of school-aged teenagers are enrolled, only 50 percent of those who reach their senior year and complete their coursework ultimately receive a diploma. One potential explanation for this result is the presence of inaccurate beliefs about key outcomes, which may lead students to exert less effort than necessary to complete their degree. Addressing these inaccurate beliefs is essential to inducing greater effort and ultimately increasing school completion rates, particularly among students from low-income households (Dynarski et al., 2021).

In this paper, I examine how to improve graduation rates among students who reach the final year of secondary education but are at risk of not graduating in a context with minimal monetary constraints—the "last mile problem" (Mullainathan and Shafir, 2013) in an educational setting. Senior students may fail to take the necessary steps to graduate because they either perceive obtaining a diploma as unnecessary or lack accurate information about their chances of graduating and the effort required to achieve it within the limited time remaining. While there is evidence on providing information about the returns to education to correct inaccurate beliefs regarding its labor market value (see, e.g., Jensen, 2010; Bleemer and Zafar, 2018), less is known about addressing inaccurate beliefs concerning the likelihood of graduation. This paper examines these phenomena and evaluates the relative importance of both channels. Notably, for the second channel, the analysis relies exclusively on information from within the educational system, which is particularly relevant in many developing countries where data scarcity makes it challenging to provide commonly used information, such as the returns to education.

In 2019, I conducted a randomized controlled trial in 61 public high schools in Salta, Argentina, to test the impact of providing two separate pieces of information to senior students and their effects on high school completion. In this context, many students fail to graduate because they do not pass certain subjects in previous years and fail to take the required make-up exams within the designated time frame to receive their diploma on time.[1] Consequently, at the start of their senior year, students are classified as being in either good or

[1]Several Latin American countries use similar systems, such as make-up exams or extra homework for failed subjects after the academic year, in their secondary schools as a strategy to prevent grade repetition and reduce the risk of dropping out.

bad academic standing, depending on the number of pending subjects. The first intervention (*Production Function - PF*) provides information about the probability of graduation based on baseline academic standing. While students are generally aware of the requirements for graduation, this intervention offers easily digestible statistics highlighting the negative consequences of not being in good academic standing for on-time graduation. The second intervention (*Returns to Education - RE*) delivers information about the economic returns to education.

The interventions were delivered through a brief presentation using slides during a single visit to each school, reinforced with follow-up reminder messages. In the *PF* arm, the presentation included statistics on the previous cohort's graduation rates based on their academic standing at the start of their senior year, as well as information about the intermediate steps necessary to improve academic standing and ensure on-time graduation.[2] The goal of this session was to help students establish a link between their current academic standing (which they were aware of at the time of the intervention) and the observed graduation rates of students in similar academic conditions from 2018, enabling them to assess their own probability of graduating. In the *RE* arm, students were shown data on employment rates and wages by education level for young individuals living in the same city. To evaluate the effectiveness of the interventions, I used multiple individual-level data sources, including a baseline survey, hard copies of academic records (such as grades, pending subjects, and attendance at make-up exams), college enrollment records, and formal employment data. The study sample comprised nearly 1,800 senior students enrolled in public high schools.

I find that both interventions have a positive and significant impact on timely graduation. Notably, the magnitude of these effects is substantially larger than those observed in previous interventions with similar objectives in both developed and developing countries. The *RE* intervention increases the probability of graduation by 10 percentage points (nearly 20 percent relative to the control group), while the *PF* intervention increases graduation by 5 percentage points (10 percent). Additionally, I find an increase in observable effort, measured by attendance at retake exams and success in passing those exams. Both interventions also increase the probability of college enrollment by 5 percentage points compared to the control group (from a mean of 13 percent).

Analyzing heterogeneous treatment effects, I find that students who were less likely to graduate at baseline—based on predictions from past academic performance and observable demographic and socioeconomic characteristics—experienced the largest increase in the likelihood of graduation after receiving either of the information interventions. Furthermore,

[2] In the control group of this study, 55 percent of students had at least one pending subject at the beginning of their senior year. This meant that, in addition to passing their mandatory senior subjects, those students had to make an extra effort during their senior year to obtain a high school diploma.

students who were more likely to graduate at baseline and received the PF intervention were more likely to enroll in college in the following academic year.

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

To further investigate whether misperceptions about one's probability of graduation can be modified, I included a question in the baseline survey asking students to estimate their likelihood of graduating and re-elicited these beliefs after providing the information. I found that students who received the PF intervention became more accurate in their estimates when their beliefs were re-elicited. Additionally, to estimate heterogeneous effects on graduation based on confidence levels, I created confidence indicators by comparing the subjective measure (students' estimated likelihood of graduation) with the objective probability of graduation at baseline. Among overconfident students, the RE intervention has a larger effect on timely graduation than the PF intervention, whereas the effects for underconfident students are similar in magnitude across both treatment arms.

To disentangle the mechanisms driving these results, and following Bleemer and Zafar (2018), I found that the salience of the information, rather than its informativeness, explains the observed impacts on graduation. These effects appear to be driven primarily by students in poor academic standing at the start of their senior year. Regarding belief updating about their chances of graduation, students who received the PF intervention were more likely to adjust their perceptions, consistent with salience-based updating.

This study makes several contributions to the literature on improving educational outcomes. To the best of my knowledge, it is the first field experiment to examine how providing students with information about the educational production function impacts school achievement. Using data available from within the educational system, this research identifies misperceptions students face during the certification process, such as pending subjects or exit exams, and assesses the relative importance of this channel compared to providing information on the returns to education. Previous studies have primarily focused on the effects of information about the economic returns to education, with findings ranging from positive to null, and even negative impacts on educational achievement (see, e.g., Jensen, 2010; Loyalka et al., 2013; Avitabile and de Hoyos, 2018; Bonilla-Mejía et al., 2019). Additional evidence suggests that addressing information frictions related to students' academic standing can influence educational decisions (Andrabi et al., 2017; Dizon-Ross, 2019). In the context of the United States, Dynarski et al. (2021) demonstrates that reframing information about financial aid can significantly alter college decisions among low-income students. By emphasizing the value of small yet critical pieces of information, this study highlights their potential to significantly enhance student achievement in contexts where economic constraints to accessing education are minimal.

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 contribute to the literature exploring how behavioral patterns, such as present bias and overreliance on routine or defaults, influence students' decision-making (Dynarski et al., 2021). Empirical evidence suggests that individuals often overestimate the likelihood of important outcomes (Feld et al., 2017; Heger and Papageorge, 2018; Machado et al., 2018), leading to suboptimal decisions, particularly among unskilled individuals (Choi et al., 2014). In an educational context, overconfidence may specifically cause students to reduce their study efforts (Nowell and Alston, 2007). I extend this literature by demonstrating that, in the short term, after receiving information about actual probabilities of graduation, overconfident students adjust their perceived probabilities in the right direction and they are more likely to graduate as they adjust their effort accordingly.

My findings inform policy strategies to increase high school completion among disadvantaged teenagers at risk of failing to graduate on time. Even in contexts with unrestricted access to education, inaccurate beliefs about important outcomes, such as probabilities of graduation or returns to education, can lead students to underinvest in effort, limiting their medium-term economic opportunities by reducing their chances of attending college or competing in job markets where high school diplomas are critical signals.[3] At a cost of approx-

[3] I conducted qualitative interviews with the main employment agencies that medium and large firms in Salta hire to recruit employees. Recruiters stated that, even for jobs requiring minimal skills, such as cashiers and shelf stockers, employers mandate the completion of secondary school. Additionally, employers are increasingly favoring young candidates pursuing any level of education beyond high school as a way to

imately US$1 per treated student, this study provides evidence of the most cost-effective information-based intervention to improve academic outcomes in low- and middle-income countries, with an internal rate of return of 146 percent under the most conservative scenario (see Evans and Acosta, 2024, for a meta-analysis).

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 students drop out because they are not able to deal with school institutional guidelines."[5] But another important explanation 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.

compensate for their lack of experience and to serve as a "signal of responsibility and commitment."

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

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

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] Students are categorized by academic standing at the beginning of their senior year as either "in good standing" (zero pending subjects) or "in bad standing" (one or two pending subjects). According to school administrators, the main driver of low graduation rates is the prevalence of pending subjects, as students often fail the necessary exams to pass or do not attend them. Administrators also emphasized the importance of timely graduation, noting that once students leave the formal system, they are less likely to return. 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 and Senior Students Performance 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.

[6] There are no national or provincial exams to determine minimum levels of proficiency or to enroll to post public secondary education.

[7] Grade reports provided at the end of the academic year highlight failed subjects and list pending ones. These reports are sent to parents/guardians quarterly for signature, and interviews confirmed that parents are aware of their children's academic status but feel unable to enforce rules.

Poor performance and low graduation rates among students who reach their senior year are prevalent. According to self-reported data from an anonymous national survey conducted at the end of the 2017 academic year (Aprender, 2017), almost 40 percent of senior students were in bad standing (with at least one pending subject). Figure 1, Panel A, shows that in the control group (cohort 2019), over 55 percent of students began their senior year with at least one pending subject, indicating low chances of timely graduation and highlighting the prevalence of this issue. In Panel B, I show that pending subjects significantly hinder timely graduation: students with two pending subjects have a graduation rate of 12 percent, compared to 87 percent for those with none. Additionally, Table 1, Panel A, shows that the overall graduation rate in the control group is 50 percent. Despite this poor performance, students reported overly optimistic expectations about their likelihood of *timely* graduation in the baseline survey.

At the onset of this study, qualitative fieldwork was conducted to understand why students who had already invested at least five years in high school were failing to obtain a diploma in their final year. Principals, other school authorities, and teachers consistently reported that students often fail to exert sufficient effort to pass pending subjects and frequently do not attend examination periods to address their standing. They also noted that these issues tend to worsen during students' senior year, as the final year of secondary education is marked by several institutional and non-institutional activities.[8]

Students in bad standing often stated that they skipped examination dates due to other "important" matters but believed they would use the next available date, pass the exam, and graduate on time. However, as shown in Table 1, graduation rates among these students are low. A possible explanation for this behavior is procrastination: despite having multiple opportunities to remedy their standing, they fail to act. I rule out this channel, as students are aware that they have limited chances remaining during their senior year. Additionally, evidence suggests that deadlines have no effect on educational outcomes in such settings because students face other constraints (Gershoni and Stryjan, 2023). Another explanation for poor graduation rates is overconfidence in their likelihood of graduating. Using a definition of confidence detailed in Subsection 4.7, I classified students as over- or underconfident. At baseline, over 80 percent of students were overconfident. Table 1, Panels B and C, shows that overconfident students in the control group performed worse than their underconfident peers. The overconfidence of students, particularly those in bad standing, suggests cognitive dissonance between their beliefs and the effort required to obtain a diploma. In the next

[8]These activities include: the *último primer día* (the last first day of classes, celebrated with a party), *presentacióní de la promo* (students choose colors and design t-shirts and hoodies), the commencement ceremony (in which all senior students, regardless of graduation status, participate and receive non-official diplomas), and *prom night* (a dinner organized and hosted by students).

section, I use this insight to develop a theoretical framework linking 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 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 β_h; type (2) students with low return to effort in senior year β_l. In addition, even if students do not exert effort there exists a positive probability to obtain the diploma given by α which captures students' ability and past effort, and also there are two types α_h and α_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 β_i and α_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 β_l and α_h.
- State 2 occurs with probability $(1-p)$ this state is represented by β_h and α_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 D 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 D 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,[9] Argentina, from August 2019 to November 2019. The population details and the experiment's design are discussed below.

4.1 Ethical considerations

Because the sample included minors (defined under Argentine law as individuals under 18 years of age), parental consent and student assent were obtained in accordance with the instructions of the IRB office at Brown University and the guidelines of school principals and the Ministry of Education of the Province of Salta. Additionally, the materials prepared for students—including content for the online platform, survey instruments, and presentations—were approved by the Ministry of Education.

4.2 Sample

The eligible population for this study consists of students attending their senior year at public high schools in Salta. While some schools operate more than one shift—a common feature in developing settings where the same school building serves different student populations across multiple shifts due to a shortage of buildings to accommodate all students simultaneously—I included only the morning and afternoon shifts due to logistical and budget constraints. Power calculations were conducted using data from the 2018 academic year (see Appendix C, Section C.1). In 2018, 2,933 students were enrolled in the senior class across all 63 school-shifts in Salta. The unit of randomization was the school-shift level, as school-shifts have different authorities and teachers and can be considered distinct units.[10]

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

[10]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."

4.3 Experimental treatments

My experimental design included three arms, randomly assigned at the school level and stratified by the number of students and the geographic area of Salta. Due to limited statistical power, it was not possible to combine both information interventions into a separate treatment arm. The arms are described below.

***Production Function - PF*:** Using data from a subset of students of the previous cohort (2018), I computed the rate of on-time graduation 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.[11]

Following recommendations from the IRB office, suggestions on *how* to improve academic standing were provided to all students, as academic standing was private information at the time of the visit. These suggestions, proposed by the Directorate of Secondary Education, outlined *intermediate steps* to effectively transform inputs into outputs. The recommendations included the following: request mock exams (*modelos de examen*) from teachers,[12] ask for study materials from classmates or students in younger cohorts (given that teachers and required materials can change over time), consult with teachers in advance to request study recommendations, or inquire about which teachers would be part of the committee for each subject.[13]

***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, INDEC, 2018), restricting the sample to employed individuals aged 18-30 who reside in Salta and are not currently

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

[12] These exams have been required to be available for every subject and year in all public high schools since 2018, as mandated by the Directorate of Secondary Education.

[13] Committees for each subject/year are typically composed of three to five teachers, depending on the number of students enrolled for a given exam period. Most exams are written to provide documented evidence of student performance in case of disputes with parents.

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.

***Control group*:** 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.[14]

4.4 Timeline

At the beginning of this project, I obtained approval from the Directorate of Secondary Education at the Ministry of Education of Salta, as well as from each school's principal and vice-principal, who provided detailed information about shift-specific activities such as school events, exams, and trips. In 2019, I piloted the survey instruments with 11th graders to refine the questions and ensure they could be completed within the one-hour time limit allocated by school principals to minimize disruptions. School visits, coordinated with vice-principals, were conducted between August and November 2019 (see Figure 2), before the final exam period. During these visits, I collected baseline survey data and implemented the interventions with the assistance of research assistants from the Department of Economics at Universidad Nacional de Salta (UNSa). Academic records containing the main outcomes of this project were scheduled to be collected by February 2020, after the academic year ended. However, the COVID-19 pandemic and the subsequent nationwide lockdown in March 2020 delayed data collection until March 2021.

4.5 Implementation

Two days before the intervention, the research team delivered consent forms for parents of senior students to school administrators. On the agreed date, the team met with participating senior students in a single room. Figure 3 outlines the activities conducted during these visits. During each school visit, school administrators introduced the implementation team. Tablets were distributed to students for completing the baseline survey, accompanied by a brief presentation on their use and a short explanation of the questionnaire.[15] Following this, all students were shown how to access a free online math platform developed for this study in collaboration with UNSa. This platform served as a "placebo" for control group schools. If applicable, the information provision was conducted using a short presentation with slides.

[14] 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.

[15] In schools with high attendance (over 80 students), questionnaires were distributed in paper format.

To reinforce the information treatments, SMS and/or email reminders were sent two weeks before the examination periods in December 2019 and February 2020, excluding control group students.[16] These reminders were designed to enhance the intervention's effectiveness, as supported by evidence on nudging interventions in education (see Damgaard and Nielsen, 2018, for a review).

4.6 Data

Baseline Survey. The questionnaire included sections on demographic characteristics, past academic performance, household characteristics, perceptions of labor market outcomes (employment and earnings) by level of education, and expectations about each student's future. Additionally, the survey included a question on self-perceived likelihood of timely graduation, serving as a *subjective* measure of confidence in the probability of graduation. After providing information to the students, I re-elicited their perceptions of their own probability of graduation to test for any changes following the presentation. This was the only experimental outcome included in the survey.

School Academic Records. I collected information about academic performance after the end of the 2019 academic year. These individual records included data on performance throughout the entire school year, graduation status, pending subjects (if any), and attendance at examination dates for senior students' pending and failed subjects. An example of an individual record is provided in Figure A1, Appendix A.

Administrative Records. To measure impacts beyond secondary completion, I also collected information on college enrollment and formal employment for my sample of students. 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 and the national tax authority.

4.7 Measuring Students' Confidence in Graduation

To measure students' self-confidence about graduation, I use two sources of data: the baseline questionnaire and administrative records that provide information on each student's graduation status. I use a question from the questionnaire that asks students to estimate their probability of graduation as a *subjective measure* (see Figure A3) and a set of observable characteristics of the students and their households to predict graduation probabilities

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

as an *objective measure*. For this step, I first consider only observations from 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.[17] 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.

5 Results

In this section, I first present balance checks across treatment arms, followed by a discussion of my main empirical strategy, results on high school graduation, and the mechanisms that could explain these impacts. Additionally, I examine heterogeneous impacts by socioeconomic status and gender, as well as the effects of the treatment arms on college enrollment and formal employment. All main analyses closely follow the specifications outlined in the pre-analysis plan for this intervention (registered in the AEA RCT registry as AEARCTR-0004511, Lopez, 2019).

5.1 Participation and Balance Checks

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. To test for the reason of participation differences, Figure A2

[17] 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.

in Appendix A shows that the difference is driven by a 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).

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

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

where y_{is} is the outcome of interest for student i who attends school-shift s, the dummy variables *Production Function*$_s$ and *Returns to Education*$_s$ indicate which information treatment school s received, δ_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 summarizes student characteristics and verifies balance across treatment arms. Panel A shows that, on average, 31 students participated per school visit, with no significant differences between treatment groups. Panel B indicates that participants are, on average, 18 years old; 60 percent are female, 6 percent have children or are pregnant (if female), 73 percent have an email address, and 86 percent have access to a cellphone. Most students live with their mother (87 percent), while 58 percent live with their father.

Panel C describes household characteristics: 76 percent of students report having a computer, 85 percent have internet access (via home, cellphone, school, or public places), and households average 1.74 persons per room. Thirty-three percent have at least one parent or guardian with some college education. Additionally, 45 percent of students work (in a family business or independently), and 20 percent care for a family member. No significant differences are observed between treatment groups.

Panel D reports academic performance: 38 percent of students have repeated at least one year, and 55 percent had at least one pending subject at the time of the visit. Panel E outlines expectations: 95 percent of students plan to attend college the next year, 87 percent intend to seek employment after graduation, and students estimate their chances of on-time graduation at 78 percent. None of these variables differ significantly between 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.

These results are both significant and larger in magnitude than those reported in previous studies aimed at improving educational outcomes. One potential explanation for these higher impacts is the focus on a target population comprised primarily of students who were close to finishing their senior year of high school. Furthermore, the setting of this study presented fewer economic barriers for students: enrollment and transportation were free, removing key obstacles that often hinder academic achievement. In contrast, Jensen (2010) observed only a 5-percentage-point increase in the likelihood of graduation among the least poor students in a study employing an intervention similar to *RE*.

Proposition 1 in the model presented in Section 3 states that the impacts of *PF* on effort (and subsequently on graduation) are undetermined. However, the results indicate an increase in the likelihood of graduation, suggesting either higher effort among students or better allocation of it.

In Table 3, columns 2 and 3 present the treatment effects by academic standing at the beginning of the senior year: students in good standing (zero pending subjects) and those in bad standing (at least one pending subject). For students in the *PF* arm who were in good standing, I do not observe a significant effect, with the magnitude of the effect close to zero. A plausible explanation is that these students already understand the level of effort required to succeed, as evidenced by their good standing. In contrast, for students in bad standing, although they received unfavorable news via the *PF* (that being in bad standing is correlated with a low probability of graduation), I observe an increase of 7 percentage points—more than 30 percent compared to the control group.

Proposition 2 states that the expected effect of the *RE* arm increases with perceived returns. Consistent with this, I found higher positive impacts of this treatment arm for the entire sample. Furthermore, as shown in columns 2 and 3 of Table 3, when I analyze the results separately by pending subject condition, I observe positive impacts for students in both 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

Under the theoretical framework presented above, behavior should only change if students update their beliefs about the returns to education or the level of effort needed to obtain their diploma. This is only possible if they receive information on the returns to education or the actual graduation probabilities, the required effort, and all the intermediate steps needed to successfully transform that effort into graduation.

5.3.1 Beliefs Updating of Perceptions on Graduation

To understand the drivers of these results, I examine the role of students' self-perception of graduation on their actual graduation outcomes (Table 4) by analyzing their responses regarding their own chances of graduation before and after the interventions. A key component of the *PF* treatment was to make students aware of the correct shape of the production function for obtaining a high school diploma based on their academic standing at the beginning of their senior year.

In Table 4, column 1, I analyze graduation rates by academic standing and their relationship to my definition of confidence. I examine the interaction between the treatment received and dummy variables indicating the level of confidence (under- or overconfident, as shown in Figure 4), demonstrating the treatment's impact on graduation across 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 rates is statistically significant at the 5 percent level.

In Table 4, columns 2-4 analyze beliefs updating on own probabilities of graduation. As expected, individuals who received the *PF* treatment became more accurate in assessing their own chances of graduation, with a statistically significant reduction observed for overconfident students.[18] Columns 3 and 4 show the results of belief updating by academic standing to show evidence of who is driving these results. Notice that this analysis might be underpowered. Students with pending subjects who received the *PF* arm are more likely to lower their own graduation estimates. While a greater impact was observed among underconfident students, it was not significant. However, overconfident students significantly revised their beliefs downward by 4 points statistically significant at the 1 percent level.

The improvement in the overconfident students' perception of their graduation probability, as a result of the *PF* intervention, did not translate into a statistically significant increase in graduation rates for this subgroup by the end of the academic year. This may suggest that the short-term impacts on their beliefs about their own graduation probabilities faded after the information was delivered.

5.3.2 Salience and Informativeness of the Treatment Arms

This subsection examines mechanisms such as salience (revisions in perceptions due to exposure to information) and informativeness (revisions systematically related to the informativeness of the information), as discussed in Bleemer and Zafar (2018). This study includes data on belief revisions about one's probability of graduation and actual behavior, such as graduation.[19] A key strength of this analysis is the direct elicitation of perceptions from participants who can influence outcomes through their actions.

First, I show the relationship between errors in self-reported probabilities of graduation and subsequent revisions following the provision of information. Panel I in Figure 6 presents binned scatter plots depicting the average belief revisions among participants in the *PF* treatment arm. These plots use the same variable defined in Table 4, columns 2 to 4, categorized by error bins (the difference between self-reported probability of graduation at baseline

[18] It is noteworthy that students in the control group's accuracy decreased as they became more optimistic about their graduation chances. A plausible explanation for this result is that a visit by a researcher from an American university and students from UNSa, might have elicited an optimistic response among students, especially given the almost nonexistent formal connection between secondary and post-secondary education levels in this setting. As discussed in Bleemer and Zafar (2018), being asked again about probabilities could have encouraged subjects to think more about their chances, "(t)he purpose of including a control group in the study design is precisely to purge these confounding effects from the treatment groups' revisions."

[19] I was unable to re-elicit students' perceptions of the returns to education because this section was time-consuming, and I had limited time to conduct the school visits.

and the statistics shown during the school visit). Panel A provides evidence of students systematically revising their perceptions of graduation, showing a clear negative relationship between the revisions and the error. Specifically, respondents who overestimated (or underestimated) their chances of graduation tended to decrease (or increase) their estimates upon reassessment after receiving the information. To account for potential variations in the impact of the information based on students' academic standing, I conducted a separate analysis by the number of pending subjects. Panels B and C confirm that the negative relationship persists in both cases.

Next, I present the impacts on graduation in Panel II of Figure 6. The figure shows that students with smaller errors in their perceptions of graduation (i.e., those with greater underestimation or less overestimation of their probabilities of graduation) experienced larger increases in their likelihood of graduating compared to those with larger errors. However, when separating the impacts by pending subjects, a different pattern emerges for both groups. Panel B, which focuses on students with zero pending subjects, shows that those with a positive error were more likely to graduate. In contrast, Panel C, which examines students with at least one pending subject, reveals that the fitted graduation line is approximately horizontal. Following Bleemer and Zafar (2018), I present regression results using the following equation:

$$
\begin{aligned}
y_{is} = \beta_0 + \beta_1 \ PF_s + \beta_2 \ RE_s + \beta_3 \ \mathit{Perceptions\ Error}_i + \beta_4 \ \mathit{Returns\ Error}_i + \\
\beta_5 \ PF_{\ s} \times \ \mathit{Perceptions\ Error}_i + \beta_6 \ RE_s \times \ \mathit{Returns\ Error}_i + \epsilon_{is}
\end{aligned}
\tag{3}
$$

where y_{is} is the outcome of interest for student i who attends school-shift s. In this case I will consider actual graduation status at the student level and the revision in i's reported probability of own graduation; the dummy variables PF_s and RE_s indicate which information treatment school s received (if β_1/β_2 are statistically different from zero would be indicative of salience-based updating); $\mathit{Perceptions\ Error}_i$ denotes the error on i's perception of own probability of graduation (self-reported probability of graduation at baseline minus statistics shown during the school visit) and $\mathit{Returns\ Error}_i$ shows the error on returns to complete secondary school, reported by the students, minus the average return shown during the school visit. If β_5/β_6 are statistically different from zero would imply information-based updating. Results for graduation and belief updating are shown in Table 5.

First, I present the results on graduation for the entire sample in column 1. The interaction terms are small in magnitude, indicating that impacts on graduation are not due to information-based updating. However, β_2 is positive and statistically significant, suggesting that the effects on graduation attributable to the RE are due to the salience of the information provided. By analyzing the mechanisms and considering the academic standing at baseline (columns 2 and 3), I observe that those with the poorest academic standing are driving the results on salience. Although not statistically significant, it appears that salience was also relevant for students in bad academic standing within the *PF* treatment arm.

Next, I focus on revisions in students' beliefs about their own chances of graduation. I do not observe any meaningful or significant changes in belief updating by under- or overestimation at baseline, but the salience of the information is statistically significant: those who received the *PF* treatment were more likely to revise down their expectations of graduation, becoming more accurate. In columns 5 and 6, I split the analysis by academic standing at baseline and observe that the results are driven by those with the worst academic standing. The fact that these changes in beliefs in the short run do not translate into impacts on graduation might indicate that the effects faded away, which is consistent with the results in the previous subsection.

5.4 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.

Table 6, 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.

I analyze the effect of the information treatments on two 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. Enrollment reflects effort, as only students who explicitly register are allowed to take the exam. Attendance provides an additional indicator of effort and includes all students, regardless of enrollment status. Table 6, Panel A (columns 2 and 3), shows positive impacts of the treatments on these outcomes, though only attendance for the *RE* treatment is statistically significant at the 1 percent level (column 3). Panel B examines effects by baseline confidence level, revealing that underconfident students respond more to the *PF* treatment, increasing attendance by over 40 percentage points compared to overconfident students (significant at the 1 percent level). Similarly, the *RE* treatment shows a 27 percentage point higher attendance rate for underconfident students compared to overconfident students.

To analyze the performance of students with pending subjects, Table 6, Panel A, column 4, reports the impact of the treatments on a dummy variable indicating whether a student had at least one pending subject remaining by the end of the academic year. Both treatments reduce the probability of having pending subjects: students in the *PF* arm are 7 percentage points less likely to have pending subjects (an 8 percent decrease, significant at the 5 percent level), while those in the *RE* arm are 12 percentage points less likely (a 15 percent decrease, significant at the 1 percent level).

5.5 Heterogeneous Effects

Time Preferences

The *RE* treatment encourages forward-looking behavior, as students must wait a considerable amount of time to benefit from improved labor market outcomes. To examine the role of time preferences in timely graduation, I used a set of questions in the baseline questionnaire following the standard Becker-DeGroot-Marschak procedure (Bursztyn and Coffman, 2012) to compute each student's discount factor. Students were then categorized as above or below the median discount factor. Results, shown in Table A4, indicate that the *RE* treatment effect is greater and statistically significant for students above the median. Although the difference compared to students below the median is not statistically significant, these findings underscore the importance of considering individual characteristics like time preferences when providing information to teenagers.

Additionally, the results show that the magnitudes for both groups of students (below and above the median discount factor) in the *PF* treatment arm are lower and nonsignificant. This aligns with the nature of the *PF* treatment, which does not aim to 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 (more than two people per room) 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."[20]

Table A5 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 A5 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.

5.6 Impacts on Other Outcomes

One objective of this paper was to analyze the effects of information treatments beyond secondary school. Due to certain data limitations (explained below), I only consider whether the student was enrolled in a university during the academic year following the interventions (2020) or entered formal employment between the last quarter of 2020 and the first quarter of 2021.

College enrollment

College enrollment reflects a student's intention to invest further in human capital, making it a key variable for assessing the medium-run effects of my information treatments. To construct this variable, I requested individual enrollment data for the 2020 academic year from UNSa (a public and free university) and UCASAL (a private university), the two most important universities in Salta. University enrollment in Argentina is open and unrestricted by law, with no entrance examinations, quotas, or requirements related to high school performance. The only prerequisite is a high school diploma; however, students with pending subjects can enroll provisionally. For my estimations, I only consider final enrollment. Unfortunately, it was not possible to obtain data from other tertiary educational institutions.[21]

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

[21] It is unlikely 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 elsewhere, they would face substantial costs for relocation and housing, which are higher than attending UCASAL. Additionally, there are no national-level data available to estimate the percentage of students who move to another province for post-secondary education. Therefore, my results represent a lower bound of the effects of the information treatments on tertiary education.

Table 7, column 1, shows that only 13 percent of students in the control group are enrolled in university, while 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 for the *PF* arm and at the 5 percent level for the *RE* arm. Bonilla-Mejía et al. (2019) present an experiment aimed at improving college enrollment in Colombia by providing information on returns to education for senior students, but no effects were found. A potential explanation for these differing results is the disparity in access to post-secondary education: in Argentina, there are minimal barriers to enrollment in higher education.

Formal Employment

Formal employment is an important outcome of interest following high school completion. To measure this, I used administrative records linked to students' national IDs. While this is not public information, participating students (and their parents or guardians, if the student was a minor) provided consent for me to access their employment status after the intervention. The dataset system only allows access to information from the six months preceding the inquiry. Therefore, I included data from the last quarter of 2020 (when some restrictions from the strict COVID-19 lockdown were lifted) to the first quarter of 2021. The variable *formal employment* is a binary indicator equal to 1 if the participant was registered as a formal employee for at least one month during this six-month period.

Column 2 of Table 7 presents the results for both treatment arms. As expected, the level of formal employment in the control group is low, with only 3 percent of students holding a formal job during the observed period. However, both treatment arms show a negative and statistically significant impact on formal employment: students in the *PF* treatment arm are 1 percentage point less likely to have formal employment compared to the control group, while those in the *RE* treatment arm are 2 percentage points less likely. These results are statistically significant at the 10 percent and 1 percent levels, respectively. A potential, though inconclusive, explanation is that the treatments increased students' reservation wage.

One key caveat is that the sample size in this analysis is smaller than the original sample because not all students could be matched to information in the administrative data. To test for potential attrition issues, I created a dummy variable equal to 1 if a student was not found and 0 otherwise. I then ran the main specification and found no significant differences across treatment arms (see Table A6 in Appendix A).

5.7 Discussion about Results on High School Graduation and College Enrollment

In summary, my study reveals two main findings. First, both treatment arms positively impact high school graduation, with the *RE* treatment proving more effective overall. The *RE* treatment addresses widespread inaccurate beliefs among all students, regardless of academic performance, while the *PF* treatment primarily benefits students with poor academic standing at baseline (55 percent of students in the control group had at least one pending subject at the start of their senior year). Second, both treatments increase college enrollment by the same magnitude (5 percentage points). To investigate this further, I analyze differential treatment effects based on students' baseline likelihood of graduation.

Table 8 presents the impacts of the interaction between baseline likelihood of graduation and the treatment arms on high school graduation (column 1). Among students in the *PF* arm, those less likely to graduate at baseline increased their chances by an additional 7 percentage points compared to those more likely to graduate, a result statistically significant at the 5 percent level. However, the difference between the coefficients for the two groups is not statistically significant. In the *RE* arm, students less likely to graduate at baseline increased their chances by an additional 4 percentage points, though this difference is also not statistically significant. These findings suggest that more disadvantaged students benefit the most from the treatments, as they needed additional information to overcome obstacles to obtaining their diploma.

Table 8, 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.

Taken together, these results provide suggestive evidence that the *RE* arm had a stronger impact on the most disadvantaged students, who gained a better understanding of the value of investing in education. In contrast, the *PF* arm had a greater impact on college enrollment among students more likely to graduate at baseline by emphasizing the importance of obtaining a diploma, while also supporting less likely graduates in achieving their high school diploma. Overall, both arms effectively increased high school graduation rates and college enrollment, though their impacts varied based on students' baseline characteristics.

6 Conclusions

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

By delivering accurate information tailored to each mistaken belief, this study demonstrates positive and significant effects of both interventions on timely graduation, with magnitudes exceeding those observed in prior research. Additionally, significant positive impacts on college enrollment were observed. The evidence suggests that receiving information prompted students to exert greater effort, improve their performance in senior-year subjects, and benefit disproportionately if they came from low socioeconomic backgrounds.

The findings of this study hold significant policy implications: graduation rates can be improved in low-income settings using an inexpensive intervention that corrects inaccurate beliefs, which are more prevalent in low-income households. At an approximate cost of US$1 per treated student, this study provides evidence of the most cost-effective information-based intervention to improve academic outcomes in low- and middle-income countries (see Evans and Acosta, 2024, for a meta-analysis), with an internal rate of return of 146 percent under the most conservative scenario (see Appendix E for details of the cost-benefit analysis).

In contexts where reliable data on returns to education is unavailable or potentially misleading for certain student groups, it is essential to explore alternative information sources that can incentivize educational investment. The evidence tested in this paper utilized information readily available within the educational system. The feature explored—"pending subjects"—is relevant in other countries, such as Uruguay. Similarly, other countries, like the Dominican Republic, have exit exams that form part of the certification requirements alongside strong performance in the final years of high school. Identifying potential sources of student misperceptions about certification can help design targeted information interventions that correct these inaccuracies. Future research could explore the optimal timing for providing this information and whether combining both types of interventions (PF and RE) could amplify the positive impacts observed in this study.

References

Andrabi, T., J. Das, and A. I. Khwaja (2017, June). 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.

Avitabile, C. and R. de Hoyos (2018). The heterogeneous effect of information on student performance: Evidence from a randomized control trial in mexico. *Journal of Development Economics 135*, 318–348.

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.

BCRA (2024). Banco Central de la Republica Argentina: Cotizaciones por fecha. https://www.bcra.gob.ar/PublicacionesEstadisticas/Cotizaciones_por_fecha.asp. Accessed: 2024-12.

Belloni, A., V. Chernozhukov, and C. Hansen (2014, May). 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.

Bleemer, Z. and B. Zafar (2018). Intended college attendance: Evidence from an experiment on college returns and costs. *Journal of Public Economics 157*, 184–211.

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.

Choi, S., S. Kariv, W. Mšuller, 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.

Dynarski, S., C. Libassi, K. Michelmore, and S. Owen (2021, june). Closing the gap: The effect of reducing complexity and uncertainty in college pricing on the choices of low-income students. *American Economic Review 111*(6), 155–189.

Evans, D. K. and A. M. Acosta (2024). Changing Perceptions of Educational Returns in Low- and Middle-Income Countries: A Meta-Analysis. CGD Working Paper 699.

Feld, J., J. Sauermann, and A. de Grip (2017, March). 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.

Gershoni, N. and M. Stryjan (2023, January). Do deadlines affect project completion? experimental evidence from israeli vocational colleges. *Journal of Economic Behavior amp; 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, May). 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.

INDEC (2018). Encuesta Permanente de Hogares. https://www.indec.gob.ar/indec/web/Institucional-Indec-BasesDeDatos.

INDEC (2024). Indice de salarios. https://www.indec.gob.ar/indec/web/Nivel4-Tema-4-31-61. Accessed: 2024-12.

Jensen, R. (2010). The (perceived) returns to education and the demand for schooling. *The Quarterly Journal of Economics 125*(2), 515–548.

Lopez, C. (2019, August). Understanding low completion rates of high school in a developing country. AEA randomized controlled trials.

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.

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.

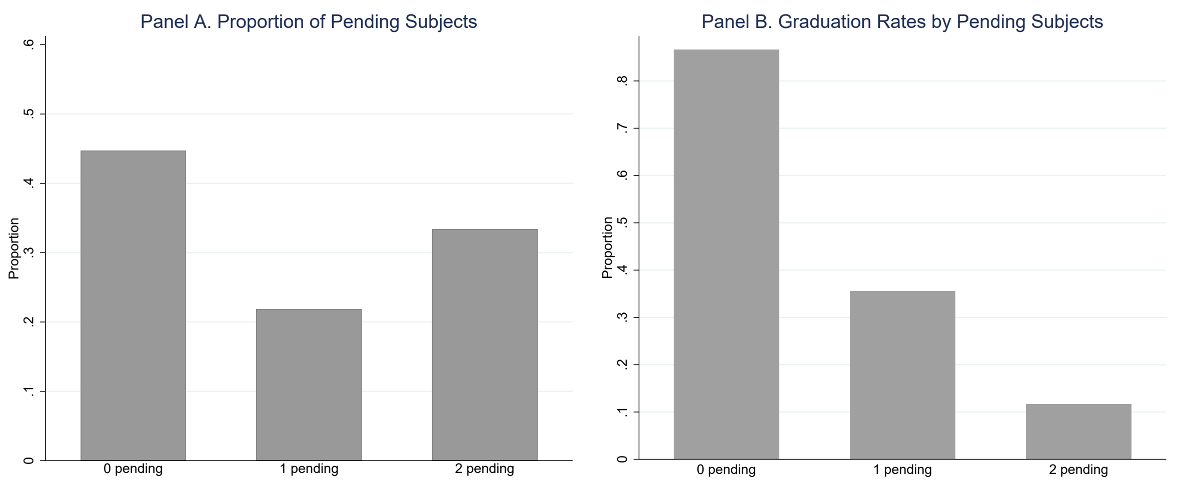
Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics 87*(3), 355.

UNICEF-ARGENTINA (2017). Para Cada Adolescente una Oportunidad: Posicionamientosobre Adolescencia. Technical report.

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

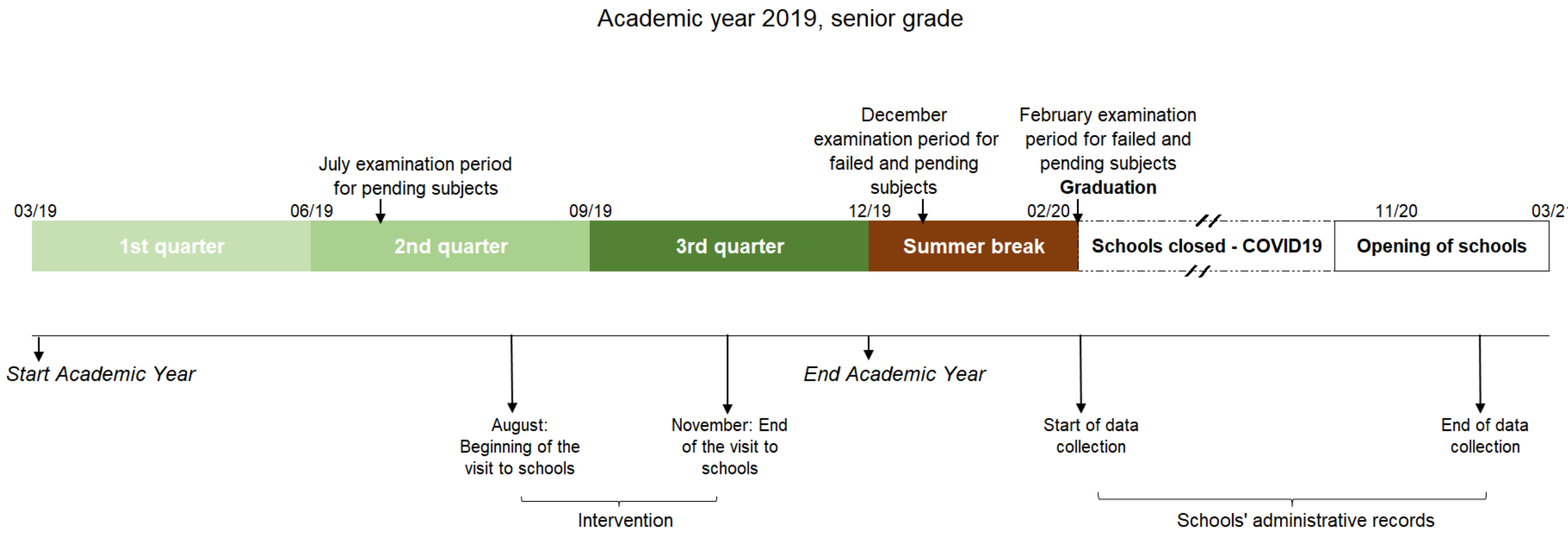
Figures

Figure 1: Senior Students and Pending Subjects, Control Group



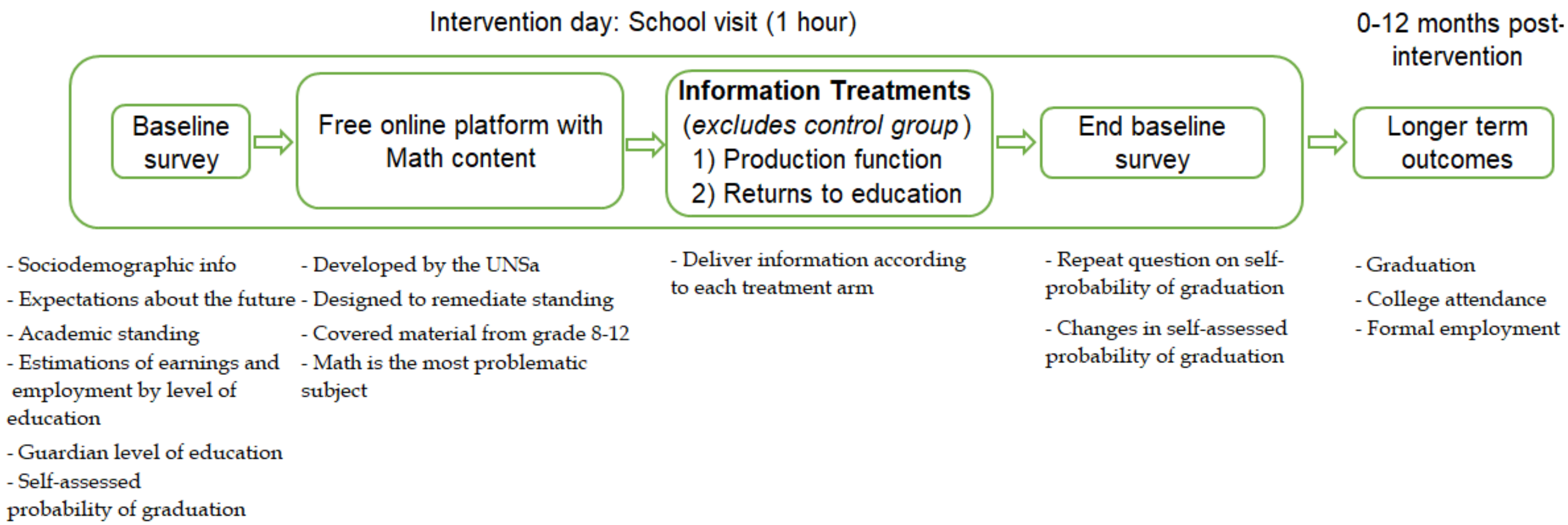
Notes: Sample limited to the control group. The horizontal axis displays the number of pending subjects at the beginning of the senior year. Panel A indicates the proportion of students with 0, 1, and 2 pending subjects at the start of their senior year. Panel B displays the average graduation rate for 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



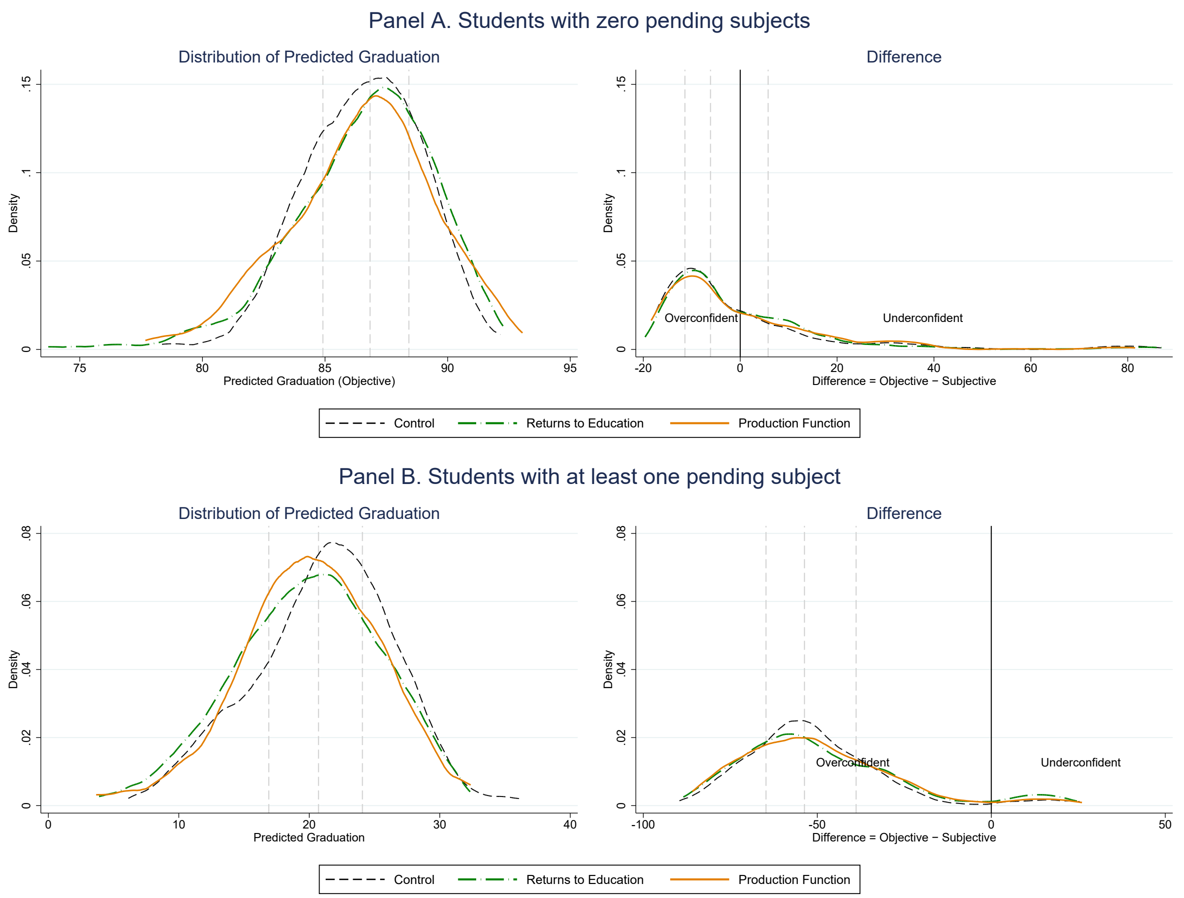
Notes: The intervention was designed for senior high school students in 2019. In 2018, discussions began with the Ministry of Education of Salta to define the scope of the intervention. The main survey instrument was tested during the first quarter of 2019. Subsequently, meetings were held with school 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



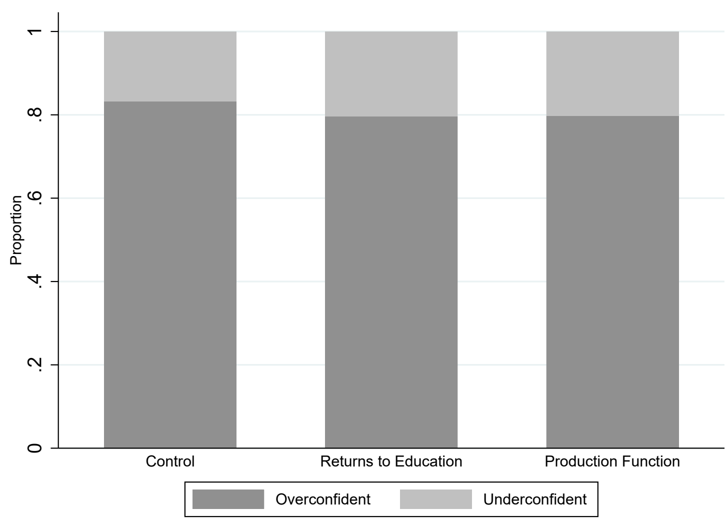
Notes: The intervention was conducted via one visit to each school and lasted no longer than one hour, as advised by school authorities. All senior students were gathered in one room. Activities included collecting a baseline survey from students at the beginning of the visit. Next, the research team 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



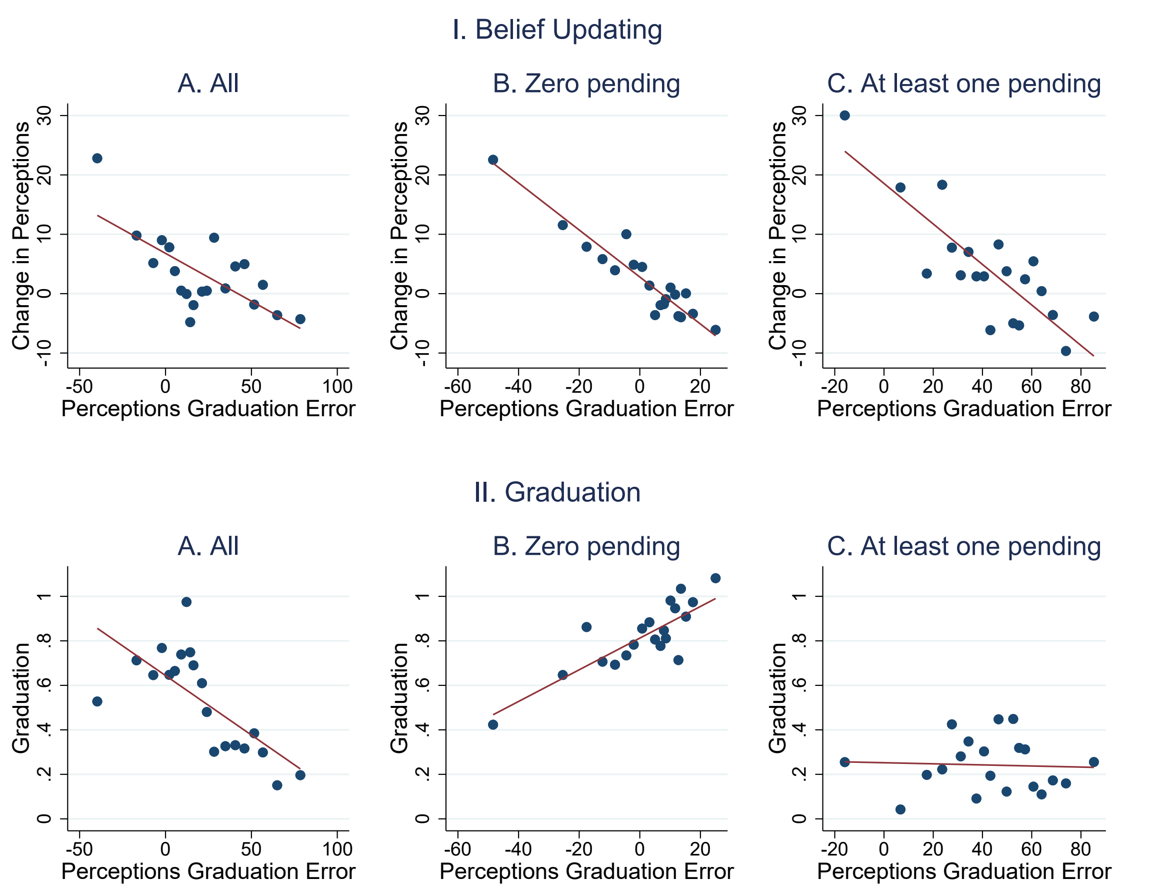
Notes: Graphs on the left show kernel density estimates of the distribution of predicted probability of graduation by treatment arm. Vertical dashed lines indicate 25th, 50th, and 75th percentiles of overall distribution, respectively. Graphs on the right show the difference between the predicted probabilities 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: Binned scatter plots of revisions in own perceptions of graduation and actual graduation by error in perceptions of graduation. Production Function Arm.



Notes: Perception graduation error is defined as individual baseline belief minus true graduation values (shown during the presentation, by number of pending subjects). Panel I: Experimental revisions in own probabilities of graduation on baseline errors by pending subjects. Panel II: Graduation on baseline errors by pending subjects.

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 4.7.

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 and Beliefs Updating by Confidence on Graduation

	(1)	(2)	(3)	(4)
		Belief Updating: Own probability of graduation		
	Graduation	All	Zero Pending	At least One Pending
Production Function × Overconfidence	0.0300	-2.204**	0.929	-4.097***
	(0.0287)	(1.008)	(0.972)	(1.485)
Production Function × Underconfidence	0.0820*	-4.663	-3.143	-9.047
	(0.0450)	(3.534)	(2.898)	(13.01)
Returns to Education × Overconfidence	0.0920***	-0.372	-1.208	0.155
	(0.0298)	(1.238)	(1.644)	(1.401)
Returns to Education × Underconfidence	0.115**	1.481	-1.064	-0.626
	(0.0461)	(4.101)	(3.423)	(13.24)
Overconfidence	-0.109**	-13.02***	-12.19***	-44.54***
	(0.0478)	(3.249)	(2.404)	(12.14)
P-value: PF × Overconfident = PF × Underconfident	0.381	0.515	0.189	0.710
P-value: RE × Overconfident = RE × Underconfident	0.696	0.693	0.973	0.953
P-value: PF × Overconfident = RE × Overconfident	0.020**	0.116	0.196	0.008***
P-value: PF × Underconfident = RE × Underconfident	0.406	0.069*	0.493	0.278
Mean (Control, Underconfident)	0.61	16.8	12.0	49
N	1786	1765	826	939

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 Self-estimated Probability of Graduation (after–before intervention)

	(1)	(2)	(3)	(4)	(5)	(6)
	Graduation			Belief Updating: Own probability of graduation		
	All	Zero Pending	At least One Pending	All	Zero Pending	At least One Pending
Production Function	-0.013	-0.046	0.088	-3.815**	-1.382	-11.842***
	(0.05)	(0.03)	(0.07)	(1.74)	(1.34)	(4.05)
Production Function × Perceptions Error	0.001	0.002	-0.000	-0.007	0.121	0.114
	(0.00)	(0.00)	(0.00)	(0.04)	(0.13)	(0.07)
Returns to Education	0.109***	0.037	0.122**	-0.395	-1.330	1.612
	(0.03)	(0.03)	(0.05)	(1.39)	(1.71)	(1.67)
Returns to Education × Returns Error	-0.000*	-0.000	0.000	-0.000	-0.000	-0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Perceptions Graduation Error	-0.006***	0.005***	-0.000	-0.152***	-0.498***	-0.400***
	(0.00)	(0.00)	(0.00)	(0.03)	(0.09)	(0.04)
Returns to Complete Secondary Error	0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Mean (Control)	0.51	0.88	0.21	5.73	3.26	7.81
N	1564	750	814	1547	744	803

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 6: Impacts of Information on Performance with Senior Subjects and Observable Effort with Pending Subjects

	(1) Passed all senior subjects	(2) Enrollment for Examination Period	(3) Attendance to Examination Period	(4) At least one pending subject left
Panel A. No Interactions				
Production Function	0.013	0.030	0.055	-0.067**
	(0.024)	(0.065)	(0.036)	(0.028)
Returns to Education	0.049**	0.042	0.13***	-0.12***
	(0.022)	(0.074)	(0.039)	(0.032)
P-value: PF = RE	0.152	0.859	0.048**	0.095*
P-value: PF = RE = 0	0.074*	0.832	0.005***	0.000***
Mean (Control)	0.65	0.62	0.44	0.79
Panel B. Interactions with Students' Confidence				
Production Function × Overconfidence	-0.0055	0.027	0.034	-0.051*
	(0.030)	(0.066)	(0.038)	(0.029)
Production Function × Underconfidence	0.050	0.020	0.46***	-0.39***
	(0.050)	(0.12)	(0.13)	(0.14)
Returns to Education × Overconfidence	0.035	0.033	0.11***	-0.11***
	(0.024)	(0.072)	(0.041)	(0.036)
Returns to Education × Underconfidence	0.093**	0.11	0.38***	-0.32***
	(0.046)	(0.12)	(0.13)	(0.100)
Overconfidence	0.0028	-0.087	0.21*	-0.21***
	(0.038)	(0.066)	(0.11)	(0.055)
P-value: PF × Overconfident = PF × Underconfident	0.378	0.958	0.002***	0.017**
P-value: RE × Overconfident = RE × Underconfident	0.257	0.449	0.058*	0.078*
P-value: PF × Overconfident = RE × Overconfident	0.183	0.931	0.031**	0.090*
P-value: PF × Underconfident = RE × Underconfident	0.405	0.514	0.518	0.620
Mean (Control, Underconfident)	0.64	0.71	0.21	1
N	1786	853	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 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 8: Impacts of Information on Graduation and College Enrollment by Likelihood of High School Graduation Based on Observable Characteristics at Baseline

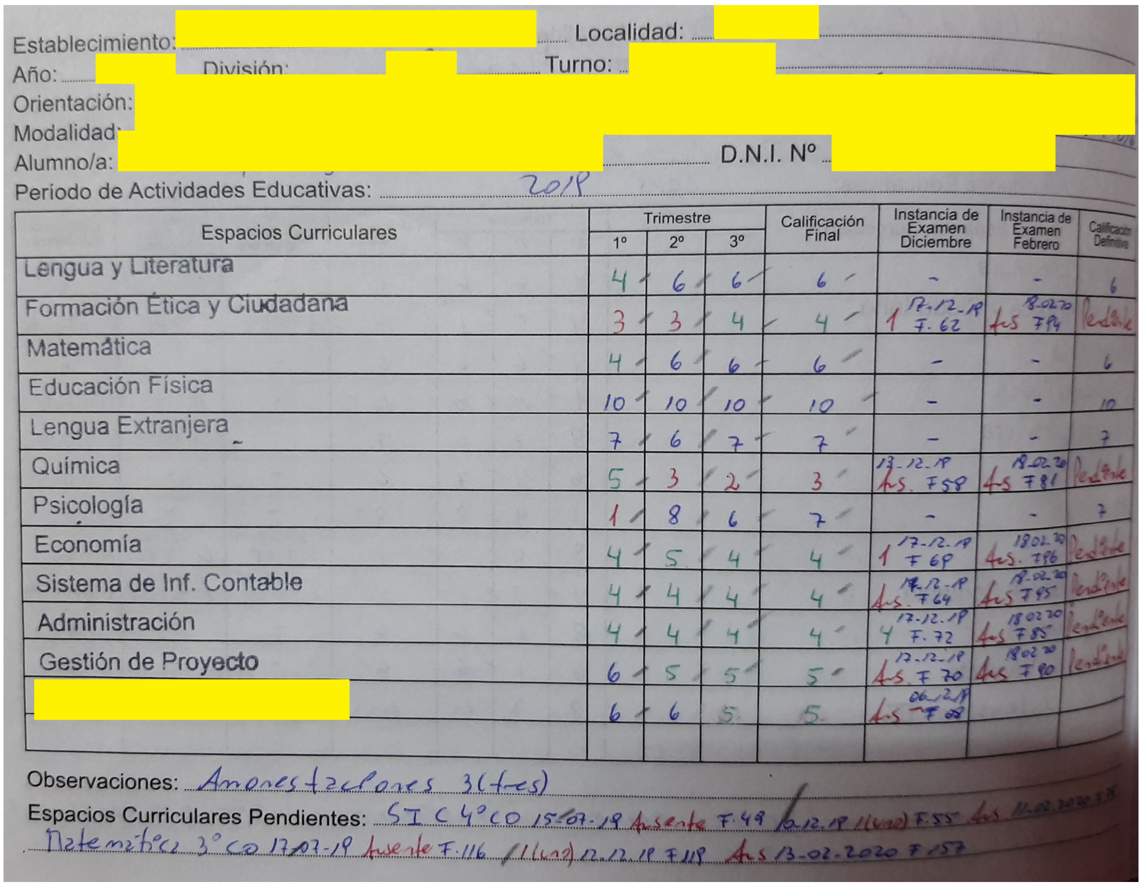
	(1) Graduation	(2) College Enrollment
Production Function × More Likely	-0.00420	0.0865*
	(0.0298)	(0.0505)
Production Function × Less Likely	0.0708**	0.0153
	(0.0316)	(0.0195)
Returns to Education × More Likely	0.0581**	0.0453
	(0.0292)	(0.0426)
Returns to Education × Less Likely	0.105***	0.0607***
	(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 4.7.

A Appendix: Additional Information

Figures

Figure A1: Example of Student Academic Report.



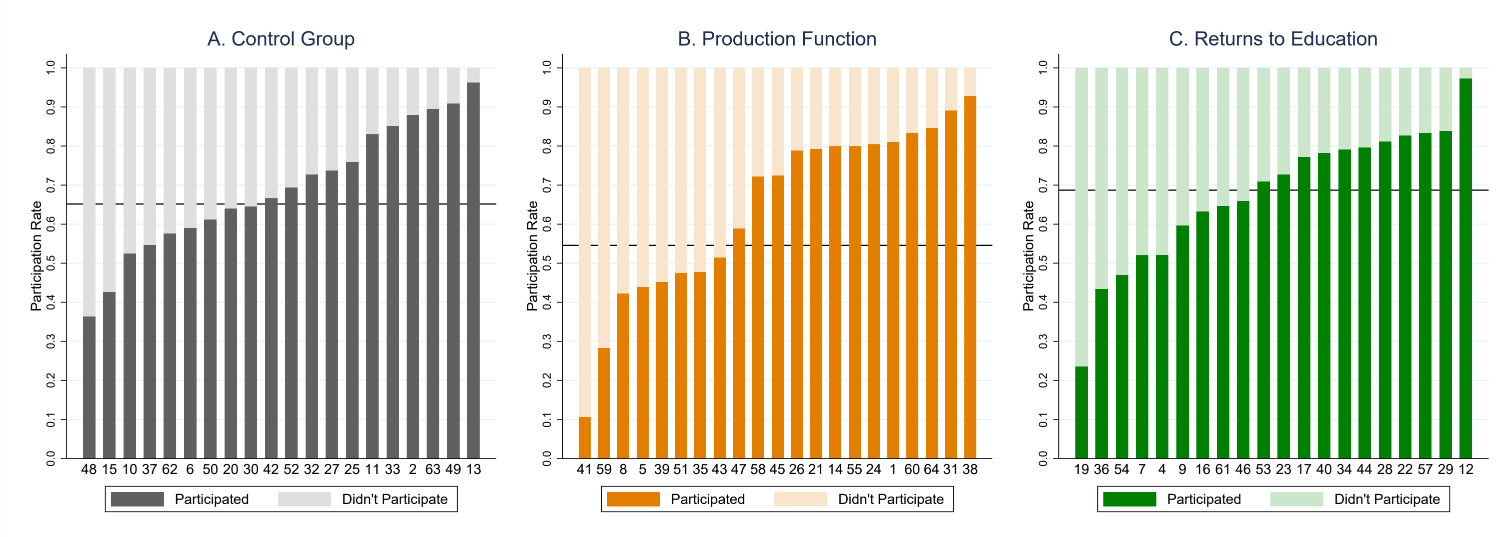
Establecimiento: [redacted] Localidad: [redacted]
Año: [redacted] División: [redacted] Turno: [redacted]
Orientación: [redacted]
Modalidad: [redacted]
Alumno/a: [redacted] D.N.I. Nº [redacted]
Período de Actividades Educativas: 2018

Espacios Curriculares	Trimestre 1º	2º	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. F86	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	17-12-18 4 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 F90	Pendiente
[redacted]	6	6	5	5	06-12-18 Aus F 08		

Observaciones: Amonestaciones 3 (tres)
Espacios Curriculares Pendientes: SIC 4º CO 15-07-19 Ausente F.49 12.12.18 1 (uno) F.55 Aus 11-02-2020 F.76
Matemática 3º CO 17-07-19 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

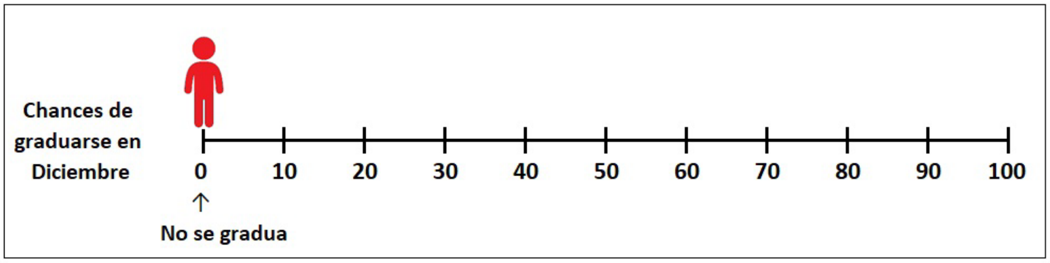


Notes: Horizontal axis shows random numbers assigned to each school. In each panel, the horizontal black lines indicates the participation rate for the entire treatment arm.

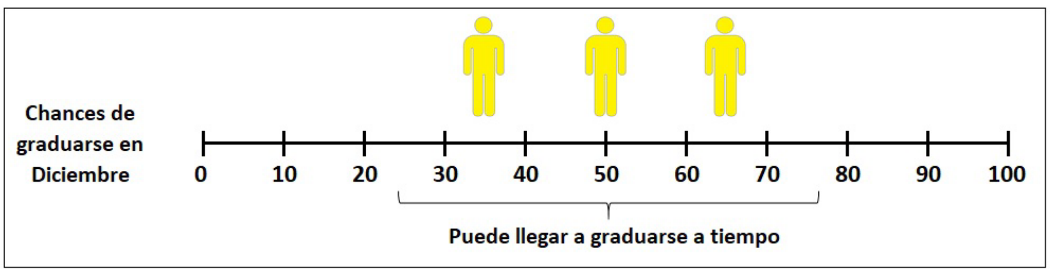
Figure A3: Prompts used to ask own probability of graduation

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.*

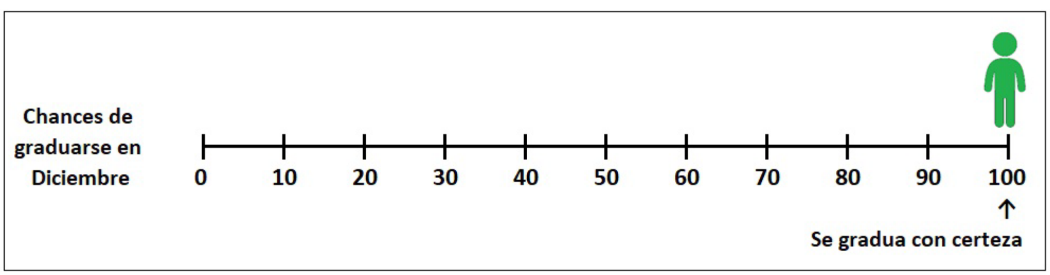
Example 1: A student who does not study, frequently skips classes. Has pending subjects and does not attend the exam periods, who does not pass all the subjects this year, has a 0% probability of receiving the diploma in December.



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



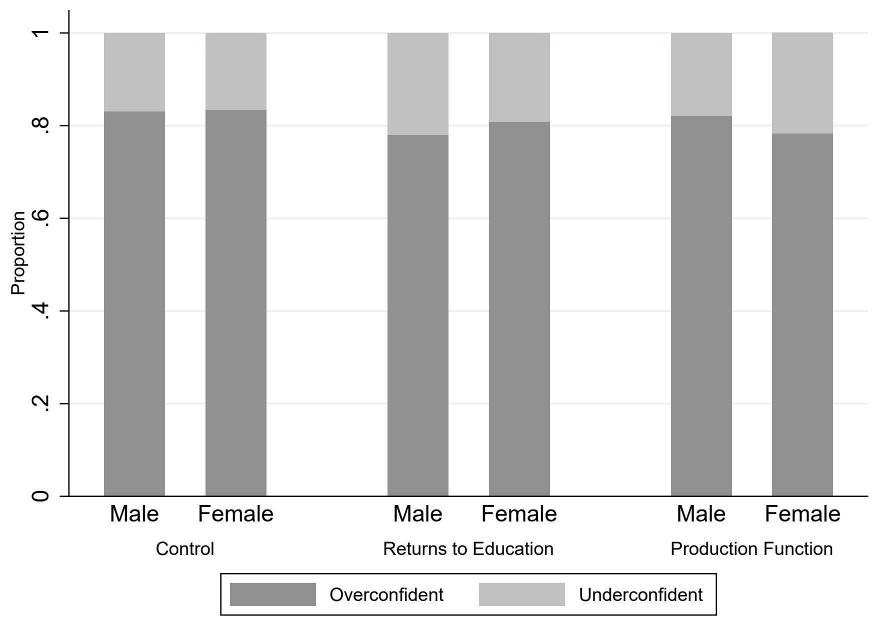
Example 3: A student who always studies, never skips classes, does not have pending subjets, with grade 10 in all subjects this year, has a 100% probability of receiving the diploma.



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

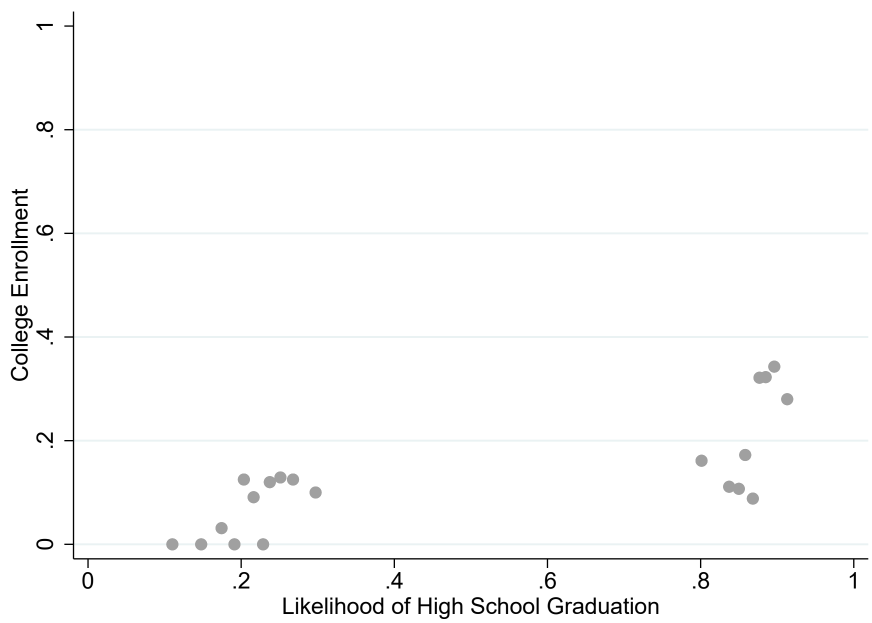
Notes: First, students were shown a concept of probability, and I provided 3 examples. Although this could be anchoring the beliefs of some students, during the piloting phase using more abstract concepts (or applied to other settings) was not helpful for students. At the end of the figure, I show the 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 4.7. The graph shows the correlation between college enrollment and estimated likelihood of high school completion.

Tables

Table A1: Selection into Participation

	(1) Participated	(2) Participated 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 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 A5: 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 A6: Difference by Missing Employment Data

	(1) Dummy Missing 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.

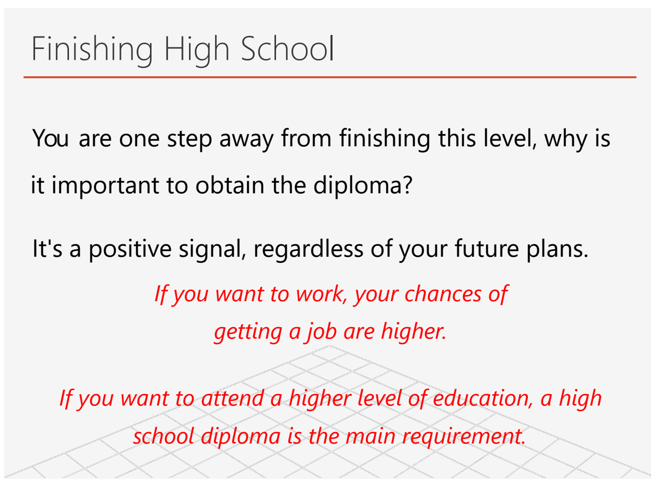
B Information Treatment Arms

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



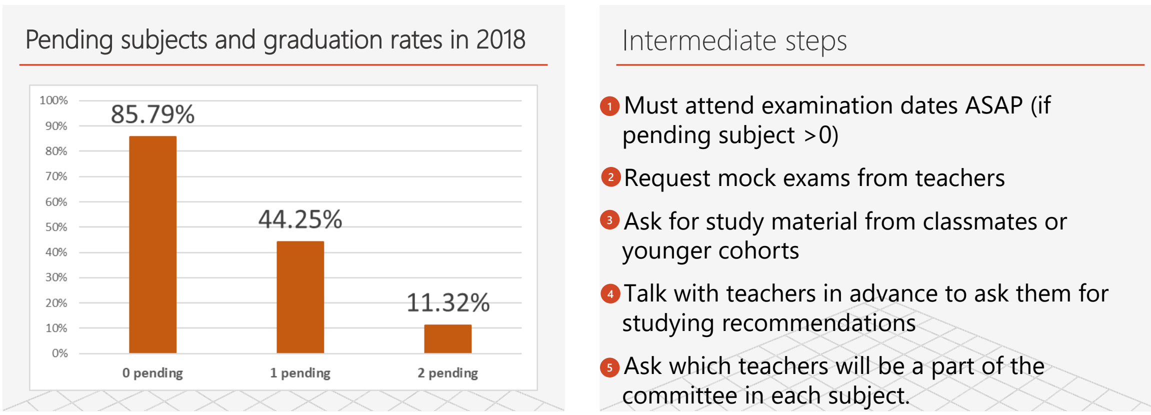
Notes: Common slide showed to all the students who received any of the intervention treatments.

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).

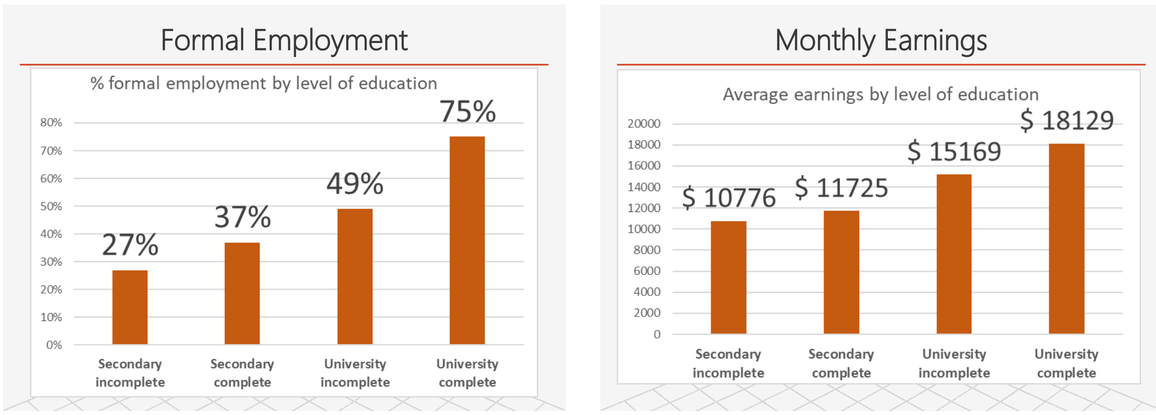
The key messages were (1) to devote more time and effort to studying students' senior year subjects and (2) for those with pending subjects, to attend the examination periods. Students' senior year includes several social activities (prom night, private parties, graduation trip, etc.). In interviews 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

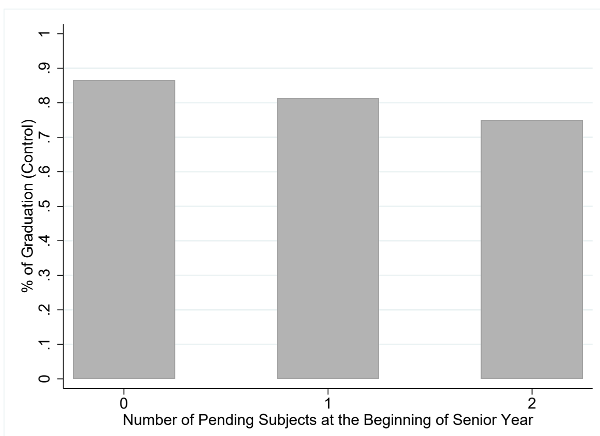
- e-mail: Hi! In our visit to your school we showed you that it is important to pass pending and subjects you failed this year as soon as possible. If you have failed subjects, remember to attend the available exam dates and study to pass them!

  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 × Scholarships Information	-0.017	-0.039
	(0.040)	(0.043)
RE × 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 Other intervention details

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

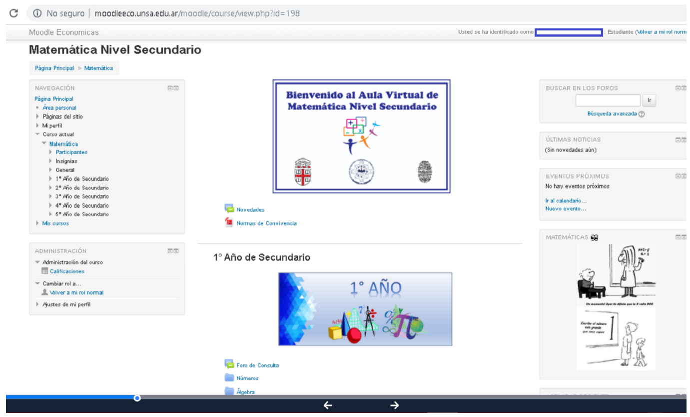
The Directorate of Secondary Education of Salta required that I provide some useful information to all students; otherwise, I would encounter resistance from school principals reluctant to give me access to their schools. So, to provide something in exchange for their participation, I designed a free 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).

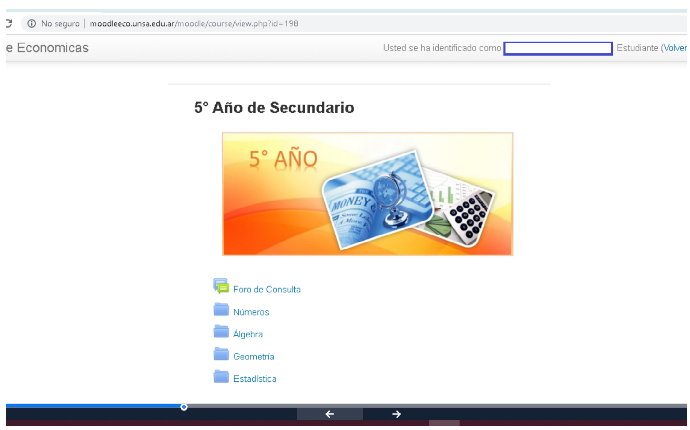
As mentioned above, the platform is not a part of the intervention, but rather enabled me to conduct the baseline surveys in all schools. After being introduced, we first explained the contents of the platform and then gave instructions on how to obtain free access (for security reasons, a unique code 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

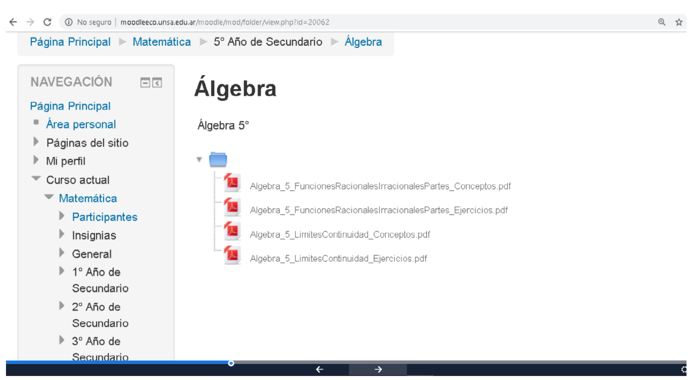
(a) Homepage



(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).

D 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

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

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

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

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

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

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. ■

E Cost-Effectiveness

To simplify the analysis, this section presents the costs and benefits of the combined interventions. The impacts are translated into additional years of education for direct beneficiaries, which are expected to lead to higher productivity and earnings in the labor market. All values are expressed in 2019 US dollars to facilitate comparison. The steps required to compute this analysis are discussed below.

Direct Beneficiaries. The intervention targeted a unique cohort: students attending their senior grade during the 2019 academic year. To simplify estimations, the analysis assumes 600 students per arm. In the control group, 50 percent graduated on time (Table 3), meaning 300 students graduated under the status quo. The interventions resulted in an additional 30 graduates in the *PF* arm and 60 graduates in the *RE* arm. Additionally, students in this cohort were more likely to enroll in college. Based on results from Table 7, 78 students in the control group enrolled in college the following academic year, while the interventions led to 60 additional enrollments (30 per treatment arm). Since it is unclear whether the students who graduated from high school due to the interventions are the same as those newly enrolled in college, two scenarios are considered:

- **Low estimate:** The 60 students enrolled in college are among those who graduated from high school as a result of the intervention, resulting in 90 beneficiaries.
- **High estimate:** The 60 students enrolled in college are distinct from those who graduated from high school due to the intervention, resulting in 150 beneficiaries.

In both scenarios, and given Argentina's low post-secondary graduation rates, it is assumed that only 30 percent of college enrollees will graduate within the expected timeframe (five years). Those who do not complete college are assumed to drop out after their first year. It is expected that beneficiaries entered the labor market in 2020.

Costs. The fixed cost of developing both information packages is US$180:

- *PF intervention:* US$130, which includes five visits to schools to collect data (US$6 per round trip) and two hours per school to digitize academic records.
- *RE intervention:* US$50, which includes five hours of coding.

Variable costs include in-person visits to schools and associated expenses. A total of 41 visits were conducted, with one person responsible for providing information and additional research assistants organizing the room. For each visit, it is assumed that two individuals were paid for one hour and received one hour of training. The variable cost amounts to US$1,086 (US$246 for transportation + US$840 for personnel). The total cost per treated student is US$0.95.

Monetizing Benefits. Senior students who directly benefit from this intervention now hold a diploma of significant value: it serves as the minimum requirement to attend college and acts as a positive signal in the labor market. While quantifying the improvement in terms of additional years of schooling for students who have already reached their senior year is challenging, I assume an increase of 0.3 years of education. The potential increase in future earnings from additional years of schooling is calculated as follows. On average, each additional year of schooling in Salta is associated with a 5 percent increase in labor income (own estimations, INDEC (2018)). The average annual per capita salary in Salta for 2019 is US$2,345 for individuals with complete secondary education, US$3,033.80 for those with incomplete college, and US$3,625.80 for those with complete college (calculated as a 23 percent projected increase from 2018 data INDEC (2024), using an exchange rate of 1 US$=60 Argentine Pesos, BCRA (2024)). Based on the estimated mean increase in years of education and the corresponding percentage increase in future earnings, the annual incremental increases in earnings are US$58.63, US$75.85, and US$90.65, respectively. Discounted benefits are calculated for the intervention's estimated impact through 2040, applying a discount rate of 5 percent.

Results The benefits of the project exceed its costs under all scenarios (Table E1). In the most optimistic scenario, the project's internal rate of return (IRR) is 289 percent, while in the most pessimistic scenario, the IRR is 146 percent—still well above the discount rate of 5 percent. Notably, this economic analysis likely underestimates the true impact of the intervention, as it does not account for additional social and private returns that are difficult to quantify in monetary terms.

Table E1: Cost and Benefits

Year	Cost (USD)	LOW Benefits (USD)	HIGH Benefits (USD)
2019	1,266	0	-
2020	0	1,005	3,015
2021	0	2,691	4,605
2022	0	2,563	4,386
	… *Years omitted* …		
2038	0	1,561	2,397
2039	0	1,487	2,282
2040	0	1,416	2,174
Total (USD)	**1,266**	**35,828**	**71,586**
NPV: Benefits - Costs		**34,562**	**70,320**
IRR: Benefits - Costs		**146%**	**289%**

Notes: Own estimations based on interventions results. Discount rate 5% (IMF and WB standard).