Misperceptions and High School Graduation: Experimental Evidence on Information Interventions from Argentina

JMP: Click here for the latest version

Carolina Lopez*

This version: October 27, 2021

Abstract

Low high school graduation rates are a central challenge in the development of human capital in low-income countries. I conducted a field experiment in Salta, Argentina, to test if lack of knowledge about the production function of graduation among high school senior students explains low graduation rates (50 percent). To measure the relative importance of this treatment, I conducted a returns-to-education information intervention in a separate treatment arm. Baseline perceptions about own probability of graduation play an essential role in students' academic performance. Providing information about the probability of graduation conditional on failed subjects and discussing intermediate steps to transform inputs into graduation allows me to detect an increase in timely high school graduation by 5 percentage points, a 10 percent increase relative to the control group. The poor-performing students at baseline respond most to the treatment. The returns to education arm has a graduation rate that is increased by 10 percentage points. The magnitude is higher than the production function arm, but both treatments provide similar or even higher impacts than previous literature by targeting different sources of misperception. I also find that both treatments increase the probability of university enrollment by 5 percentage points (more than 30 percent relative to the control group). These findings indicate that targeted information may provide an inexpensive way to increase educational attainment in low-income settings.

*Brown University, Department of Economics. Email: carolina_lopez@brown.edu. I am incredibly thankful to Bryce Steinberg, Andrew Foster, Pedro Dal Bó, Anja Sautmann, and Neil Thakral for their support and encouragement. I am grateful to Simone Schaner, Jesse Shapiro, Alejandro Ganimian, Natalie Bau, seminar participants at Brown Applied Micro Lunch, and the Development Tea group members at Brown University for helpful discussions and feedback. Also, I would like to thank to Angélica Astorga, Felipe Brugués, Amanda Loyola Heufemann, Juan Ignacio Lopez, Ana Pacheco, Santiago Hermo, Diego Ramos-Toro, and Natalia Sández Pernas for their invaluable support. The field experiment would not have been possible without the authorization and support of the Directorate of Secondary Education of the Ministry of Education of Salta and the authorities of each participant school. I acknowledge the financial support to digitized academic records from the Orlando Bravo Center for Economic Research at Brown University and CAF–Development Bank of Latin America. The experiment was approved by Brown University Institutional Review Board and it was registered at AEA RCT registry (RCT ID: AEARCTR-0001299). All errors are my own.

1 Introduction

Investment in education is increasingly necessary to enable individuals to secure a quality job and to allow upward mobility within an economy (Cotton et al., 2020). This issue is becoming more urgent in the Global South; even in settings where barriers to education are minimal, a large educational achievement gap exists with high-income countries (Glewwe and Muralidharan, 2016). In Salta, Argentina, the setting of this study, secondary school enrollment is high but graduation rates remain low. It is possible that lack of information or cognitive bias lead students to think that they are exerting the right amount of effort in education, when in fact their effort is not sufficient (Kruger and Dunning, 1999).

Considerable evidence (see Ganimian and Murnane (2016) for a meta-analysis) shows that incentivizing academic achievement (outcomes) has no effect on performance, but incentives can improve educational performance when specific tasks (inputs) are targeted. Fryer (2016) and Fryer (2011) suggest that a potential explanation for students' failure to transform effort into academic achievement could be a lack of adequate knowledge about the education production function. In addition, empirical evidence in economics and psychology shows that individuals tend to overestimate the probability of important outcomes (see for example Feld et al. (2017), Heger and Papageorge (2018), Machado et al. (2018)), leading to suboptimal decisions, especially for unskilled individuals (Choi et al., 2014).

I conducted a randomized controlled trial in 61 high schools in the city of Salta, Argentina, to study whether targeted information improves students' graduation probabilities, depending on their academic standing at the beginning of the senior year. I examine if the provision of information on *how* to get a high school diploma—that is, the intermediate steps needed to effectively transform effort into educational achievement—has an impact on the likelihood of graduation of students who are currently attending their senior year. In a separate treatment arm, I present information about the returns to education. Both treatments were introduced through a brief presentation in a single visit to each school and reinforced with reminder messages. The average graduation rate of the control group is 50 percent, and I find that both information treatment arms have positive and significant impacts on the probability of timely graduation and university enrollment. These impacts exceed those reported in previous studies.

I combine a baseline survey, hard copies of individual academic records collected from each school, and administrative data of each school to analyze the impacts of the intervention. The participants included almost 1800 students attending public high schools, and the intervention was delivered by the research team via a 40-minute presentation. The presentation primarily focused on the importance of having a high school diploma to attend college or to find a formal job. I created the presentation to provide information about the intermediate steps to improve academic standing and on-time graduation, and I included statistics on the previous cohort's graduation rates based on their academic standing at the beginning of their senior year. This last piece of information was meant to generate a mapping between each student's academic standing (known by the student at the time of the intervention) and their chances of graduation. This treatment is the *Production function* arm.

In addition, given mixed evidence on the provision of economic returns to education in the literature, I provided such information in a separate treatment arm, *Returns to education*. Several studies mention economic constraints (such as tuition, fees, and school transportation) as forming the main barrier to obtaining more education, even after this information is received (Bonilla-Mejía et al., 2019) but those barriers are minimal in this setting. Students were shown information containing employment levels and wages by levels of education, using the same format as in the other treatment arm. All students, including those in the control group, received free access to an online platform with math content to remedy their standing in that subject, if needed. In total, my experiment has three arms: *Control*, *Production function*, and *Returns to education*.

For requirements made from school authorities to grant access to the school building, a free online math practice platform was introduced to students during the visit at each school to help them improve their performance in that subject. This platform was designed for the intervention using content sent by current math public school teachers at all high school grade levels. Through this platform, I was able to survey students in all public high schools located in the city of Salta to collect information on academic status, knowledge about earnings and employment, expectations about the future, time preferences, and self-reported probabilities of graduation from high school by the end of the academic year.

Based on the survey data, I find an overall positive and significant impact on graduation rates in both information treatment arms. Specifically, the *Returns-to-education* treatment increases the probability of graduation by 10 percentage points (almost 20 percent with respect to the control group), and the *Production function* arm increases graduation by 5 percentage points (10 percent). The effect of *Returns to education* is double the effect found in Jensen (2010) for the subgroup of less poor students, and *Production function* shows results of a similar magnitude compared to Jensen (2010) for the same subgroup; however, my results apply to the entire sample of students. The students with the greatest response in both treatment arms are those with the worst academic standing at the beginning of their senior year. In addition, an increase in observable effort can be observed among those students.

In the baseline survey, I asked students for their perceptions of the likelihood that they will graduate. I compared that subjective measure with the estimated probability of graduation based on observable characteristics of the students (as an objective measure) to create an indicator of confidence. In the *Control group*, students with a high level of confidence tend to be among those with the worst performance. After the presentation of the interventions, I again asked the students about their chances of graduation. I find that students' self-reported estimations of graduation are more accurate after receiving information about graduation probability in the *Production function* treatment arm. Importantly, I observe that students classified at baseline as underconfident and overconfident (concerning their chances of graduation) exert more effort under both treatment arms, but higher effects are found for overconfident students when they receive the *Returns to education* treatment arm. These results indicate that a single but targeted intervention for different types of students could help in other settings to facilitate dismantling a detrimental cognitive bias.

This paper contributes to the existing literature on how information can affect educational choices. The literature includes explorations of the provision of information on economic returns to education in contexts with low attendance rates (mainly due to economic constraints), with results showing an increase in school achievement (Jensen (2010); Loyalka et al. (2013)). The literature also finds that providing information about the relatively higher wages for unskilled labor may dissuade students from going to high school (Loyalka et al., 2013) or may not have an impact on college enrollment (Bonilla-Mejía et al., 2019). In addition, the economics literature on low school achievement has focused mainly on economic constraints such as tuition and other fees, clothes, books, and so forth. Although interventions that reduce those costs do increase attendance, they do not necessarily increase achievement (Ganimian and Murnane, 2016). Furthermore, interventions with non-monetary incentives also fail to increase educational achievement (Fryer, 2016).

My first contribution is to provide a novel piece of information to improve students' decisions in a high-stakes setting. While previous papers mostly focus on providing information about economic returns to education or on monetary or non-monetary incentives to motivate students to invest more effort in education, I study a suggested (Fryer (2016)) new source of low educational achievement: students' lack of knowledge of the educational production function. To fill this gap in the literature, my experiment aimed to provide information to students about the production function and how to transform inputs to outputs (graduation in this paper). I use a unique setting to identify different production functions depending on students' academic standing. This arm of my experiment (*Production function*) indicates a student's probability of receiving a diploma on time based on their standing at the beginning of their senior year; it additionally provides information about how to remedy academic standing if necessary.

In a second contribution, I test whether students ignore or discount new information on finishing high school because of biased beliefs about the information they already have (DellaVigna, 2009). Psychologists have long known that people tend to overestimate their own abilities. In particular, overconfidence in an educational context may lead students to study less (Nowell and Alston, 2007). I show how this biased belief in their own performance is detrimental to students' chances of graduation, and I demonstrate that those negative consequences can be ameliorated by providing accurate information on how to obtain a high school diploma. In a third contribution, I formalize an effort-choice model that incorporates the role of beliefs in the probability of graduation. I show that in a high-stakes setting, where senior students only have few months to remedy their standing and obtain a diploma, beliefs can be updated upon receiving information that targets different sources of misperceptions: labor outcomes by levels of education and probability of graduation. I allow for several results that depend on students' type (under- or overconfident) and their academic standing when they receive each piece of information. This model shows that under particular combinations of students characteristics, the update in beliefs could lead to no change, an increase or a decrease in effort, and as a consequence, the same direction of change in the probability of graduation.

In this paper, I study a vulnerable population in a high-stakes setting, with a high probability of not getting the high school diploma are among those not in education, employment, or training (NEET). An increasing concern in all Latin American countries. Improving graduation rates in developing countries constitutes a relevant topic for the design of effective educational policies.

I also show how having a deep knowledge of the social and current institutional context in which an intervention is conducted helps to target the appropriate audience, identify the information that is needed, and design how to deliver it to improve an outcome quickly and inexpensively. In the first stage of the research process, I heavily relied on qualitative research methods such as focus groups and open and semi-structured questionnaires aimed at provincial authorities, high school principals and vice principals, their supervisors, high school teachers, and importantly, students from previous and current cohorts. This step helped me to design the *Production function* arm and detect the presence of a cognitive bias—overconfidence—and to determine the feasibility of measuring it in the time available for conducting the surveys (Rao, 2002).

This paper is relevant for informing policy strategies to increase the demand for high school diplomas among teenagers, especially those who are disadvantaged and at risk of failing to complete high school on time.[1] I study a vulnerable population in a high-stakes

[1]Discussions are currently occurring in almost all countries and international organizations such

setting, with a high probability of not getting the high school diploma an as a consequence with high chances of being classified as not in education, employment, or training (NEET), which represent an increasing concern in Latin America. Although access to the educational system is not restricted in many settings, youths' lack of information can cause them to invest less than the optimum level of effort in education, which in the medium run will limit their economic opportunities (no access to post-secondary education and lack of opportunities in a job market that uses possession of a high school diploma as a signal). This paper provides evidence of a factor that could explain the underinvestment in education—the lack of knowledge on how to efficiently transform inputs into outcomes—and how it can be offset by providing accurate information on how to improve chances of graduation based on students' current academic standing.

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, secondary education is accessible for most teenagers: there are free public schools in every district and transportation is sometimes free for students. As a result, most teenagers are enrolled in high school (91 percent, CEDLAS and World-Bank (2018)). 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 (and attend until the last day of classes) may not obtain a high school diploma because they fail to pass all required subjects.

A possible reason for this result is that the students are not interested in pursuing educa-

as UNICEF (Annual Report 2020 https://www.unicef.org/reports/unicef-annual-report-2020) on how to recover from the consequences of the COVID-19 pandemic and the related closure of schools and the impacts on student achievement. Low high school diploma achievement was already a concern before the pandemic in Argentina. UNICEF has reported low school achievement (UNICEF-ARGENTINA, 2017), a referent from the private sector highlighted difficulties in hiring young people with a high school diploma, and civil associations, along with the current National Director of High School level, have expressed concerns related to low completion rates https://www.lanacion.com.ar/sociedad/crisis-educativa-por-que-toyota-no-consigue-200-jovenes-con-el-secundario-completo-para-\trabajar-en-nid05082021/, https://www.lanacion.com.ar/sociedad/preocupacion-por-que-la-mitad-de-los-alum esperado-nid07082021/.

tion beyond high school. However, another consequence of not getting a high school diploma is drastically lowered chances of obtaining a quality job.

2.1 Educational System and Students' Academic Standing

In Argentina, education is compulsory up to the end of secondary school. As a result, the share of youths of secondary school age who are attending secondary school is 91.2 percent, with 74.7 percent attending public schools (CEDLAS and World-Bank, 2018). Students 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; not being able to deal with school institutional guidelines".[2]

But another important explanation, which has attracted less research attention and is not even mentioned by the Director of Secondary Education at the national level, is that students who attend until the last day of high school may still not obtain a high school diploma. This topic has remained unexplored basically because there are no digitized data at the individual level that allow making conclusions about the magnitude of this issue.

In Argentina, the academic year begins in March and classes finish by December, but the year officially ends in February if the students have to remedy their academic standing. To finish high school, students must be in good standing in all subjects (10-12 per year). There are no national or provincial exams to determine minimum levels of proficiency or to enroll to post public secondary education.[3] In addition, students can have *pending subjects* (subjects with final grade lower than 6 over 10) from one year to any of the following years of high school (up to two pending subjects are allowed; if a student has three or more, they must repeat the year). Each student is fully aware of the number of pending subjects they have.[4] I use this concept throughout this paper to define students' type at the beginning of senior year: in good standing (zero pending subjects) or in bad standing (at least one pending subject). All high schools have three examination dates on which to pass pending subjects each year

[2]`https://www.lanacion.com.ar/sociedad/preocupacion-por-que-la-mitad-de-los-alumnos-no-\termina-el-secundario-en-el-tiempo-esperado-nid07082021/`

[3]According to a national law `https://www.argentina.gob.ar/normativa/nacional/ley-24521-25394/actualizacion` "All persons who pass secondary education can freely and unrestrictedly enter at the higher education level" to guarantee equal opportunities and conditions in access.

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

(July, December, and February). During phone interviews, school administrators said that the main issue related to low completion rates is the pending subjects; the administrators report that students either do not pass them or simply do not attend the examinations.

A consequence of not getting a high school diploma is drastically lowered chances of obtaining a quality job.[5] In fact, most teenagers and young people who do not finish high school are among those not in education, employment, or training (NEET). The NEET issue represents an increasing concern in all Latin American countries (Tornarolli, 2016), and in Argentina, approximately one out of five individuals aged 15-24 are classified as NEET. Improving graduation rates in developing countries constitutes a relevant topic for the design of effective educational policies.

2.2 Educational Situation in Salta

The intervention was carried out in the city of Salta, the capital of the Argentinian province bearing the same name. In this setting, education and transportation are free for all students enrolled in all levels of formal schooling. In 2018, the province of Salta had the eighth-largest sub-national 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 had a "satisfactory" level in math.

According to self-reported data from an anonymous national survey of students collected at the end of the 2017 academic year (Aprender, 2017), almost 40 percent of senior students were in bad standing (had at least one pending subject) and had not remedied their status by the end of the academic year. This finding indicates that the chances of timely graduation for that cohort were low, and at the same time it reveals how common it is for students to have pending subjects at the beginning of the academic year.

At the onset of this study, qualitative field work was conducted to understand why students who had already invested at least 5 years of their lives attending high school were failing to obtain a diploma in the last year. Principals, other school authorities, and teachers were in accord in reporting that students do not make enough effort to pass failed subjects, they do not attend the examination periods to remedy their standing, and this issue gets worse during the senior year.[6] Students in bad standing stated that they did not use the

[5]At the onset of this project, I conducted qualitative interviews with the main agencies in Salta hired to recruit employees for medium and large firms located in the city. Recruiters stated that even for jobs that require minimum skills, such as cashiers and shelf stockers, employers require completion of secondary school. Employers are also starting to ask for young people attending any level of education beyond high school to compensate for their lack of experience and as a "signal of responsibility and commitment," see (Spence, 1973).

[6]The last year of secondary education is a expected year for the students owing to several institutional and non-institutional activities, with students beginning to make arrangements in 11th grade. Some of these

pass dates because they had other "important" matters but they would use the next one "for sure," pass the exam, and receive the diploma on time. Their confidence in being able to complete this process suggested to me some sort of cognitive dissonance regarding what they believe about their actions and effective effort to obtain the diploma. I use this insight in the next section to develop a theoretical framework that relates beliefs to effort.

3 Theoretical Framework

Previous literature in economics and psychology indicates that performance in education is inversely correlated with overconfidence, indicating that unskilled students are more confident than the skilled ones (Banks et al. (2019), Machado et al. (2018)). But what happens if they learn the true probabilities about the fact that they are confident about?

In terms of this paper, the question is how will students' beliefs and therefore their behavior change if they are informed about their true probabilities of graduation? Ex ante the answer is not obvious. Some 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 will become discouraged. Still others may become motivated to achieve the goal, and 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 with probability of graduation and beliefs. I show how the provision of information affects beliefs, but depending on students' types and under certain conditions, the change could lead to no response or to an increase or decrease of effort that consequently affects the probability of graduation.

3.1 A Model of Effort and Probability of Graduation

Students want to maximize their level of utility in the senior year, which depends on effort. I assume the utility function has two components: benefits and costs.

Definition 1 *The benefits are defined by the probability of graduation, which depends on a linear and scalar transformation of effort $g(\beta e + \alpha)$ times the return to the diploma V. I assume that $g(.)$ is a concave function.*

activities are the *Ultimo Primer Dia* (last first day of classes in the secondary level), *presentación de la promo* (every year students belonging to one class pick colors and a name that represent them in a way, they design and buy a t-shirt and hoodie personalized for each student, they introduce them to all the school by using music and a performance, and they invite their relatives to the school), commencement ceremony (regardless of whether they obtain a diploma, all senior students participate in a ceremony organized by the school where non-official diplomas are delivered to each student to celebrate their presence in the school after at least 5 years), *prom night* (a dinner organized and hosted by students, with the participation of school authorities, teachers, and students' relatives), and other private events hosted by students after December.

Definition 2 *The costs are a convex function that depends on effort. For simplicity I assume a linear function δe.*

Given the above the definitions, the maximization problem of a student can be written as:

$$\max_{e} \quad g(\beta e + \alpha)V - \delta e,$$

and the necessary and sufficient first-order condition that characterizes the optimal effort (assuming an interior solution) is given by:

$$g'(\beta e^{*} + \alpha)V = \delta$$

Note that there is certainty about the returns to effort and education, so there is no room to change students' beliefs. In the next subsection, I expand this framework to allow for the possibility of misperceptions.

3.2 A Model of Effort and Probability of Graduation with Uncertainty about Returns to Effort or to Education

The low graduation rate at the end of the academic year reflects a lack of knowledge of the diploma production function. The misinformation could involve the translation of study time into effective effort on exams in senior year courses or the amount of effective effort needed to pass pending subjects. In this case, students are not fully aware of the returns to effort or the returns to education, so beliefs will play a crucial role in graduation.

A student's subjective utility is defined following Definition 1 and Definition 2, but depends on students i (u -under- or o -overconfident-); and $\hat{V}$ is the perceived return to the high school diploma.[7]

Definition 3 *The probability of graduation is the probability that the minimum across scores of senior and pending subjects is above a threshold T.*

This implies that effort will be allocated to equate scores across given the state of the world. I assume there are two types of effort: (1) to pass each senior subject e_1 and (2) to pass each pending subject e_2. Therefore, the total effort needed to pass the senior year and obtain a diploma is the sum of the minimum effort the student has to exert to pass all subjects as shown below. Individuals have perfect information about their own standing

[7] From hereon, I omit the type of student (i) from the equations to reduce notation.

at the time of the interventions (how many pending subjects they have $n-1$) and I will consider the number of senior subjects as a numeraire and $n-1$ as the number of pending exams (with $n=1$ the student has zero pending subjects):

$$e = e_1 + e_2(n-1)$$

To achieve the threshold T, students must exert different types of effort, and in the case of pending subjects, effort includes ν_1 and ν_2 to indicate a slope and scalar transformation, respectively. Under both types of effort students have to reach the threshold T, so by using Definition 1 and considering the argument of the $g(.)$ function I have the following:

$$\beta e_1 + \alpha = \beta(1+\nu_1)e_2 + \nu_2 + \alpha$$

I use this equation along with the definition of total effort e to find the value of e_1, and I incorporate the pending subjects in the utility function. The results are the following: [8]

$$e_1 = \frac{\beta e \nu_1 + \beta e + n\nu_2 - \nu_2}{\beta(n+\nu_1)}; \quad e_2 = \frac{\beta e - \nu_2}{\beta(n+\nu_1)}$$

As stated above, students could have misinformation about the translation of study time into effective effort, and I assume there are two potential states of the world: (1) a good one, where students are certain about the returns to effort (ν_1 and ν_2 are equal to zero) and (2) a bad one, where the return to effort is affected by noise (there is a slope and scalar transformation on effort, with $\nu_1 > 0$ and $\nu_2 < 0$). Now, by using Definition 1 and considering these 2 states of the world, I have the following:

$$g_1 = g\left(\frac{\beta e}{n} + \alpha\right)$$

and

$$g_2 = g\left(\frac{(1+\nu_1)\beta e + \nu_2}{n+\nu_1} + \alpha\right)$$

[8] To obtain these results:

$$\beta e_1(n-1) = [\beta(1+\nu_1)e_2 + \nu_2](n-1)$$
$$\beta e_1 n = \beta e_1 + \beta(1+\nu_1)e_2(n-1) + \nu_2(n-1)$$
$$\beta e_1\nu_1 + \beta e_1 n = \beta(1+\nu_1)e + \nu_2(n-1)$$

and,

$$\beta e_1 - \nu_2 = \beta e_2(1+\nu_1)$$
$$\beta e_1 + \beta e_2(n-1) - \nu_2 = \beta e_2(n+\nu_1)$$

The true probability of state of the world 1 is given by p and that of state 2 is $(1-p)$ of state 2. Let $\hat{p}$ be the probability assessed by the student. The student maximizes their expected utility:

$$E(\tilde{g}) = \left[\hat{p}g\left(\frac{\beta e}{n} + \alpha\right) + (1-\hat{p})\,g\left(\frac{(1+\nu_1)\beta e + (n-1)\nu_2}{n+\nu_1} + \alpha\right)\right]$$

The maximization problem is now given by:

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

Given the assumptions about the functional forms, this problem has a unique solution given by $e^* = e(\hat{p}, \hat{V})$. To obtain the main predictions of this model, I apply the implicit theorem function to the first-order condition, but I first formalize the concept of under/overconfidence in this framework:

Definition 4 *The student is overconfident if and only if their subjective probability of graduation is higher than their objective probability. This is,*

$$\hat{p}g_1\left(e^*\left(\hat{p},\hat{V}\right)\right) + (1-\hat{p})\,g_2\left(e^*\left(\hat{p},\hat{V}\right)\right) \quad > \quad pg_1\left(e^*\left(\hat{p},\hat{V}\right)\right) + (1-p)\,g_2\left(e^*\left(\hat{p},\hat{V}\right)\right)$$

Remark 1.1 With

$$g_1\left(e^*\left(\hat{p},\hat{V}\right)\right) > g_2\left(e^*\left(\hat{p},\hat{V}\right)\right)$$

and $\hat{p} > p$, the student is overconfident.

Role of the Treatment Arms

I consider the effect of two separate treatments. The first arm consists of a shock to the beliefs about what state of the world the students are in. The second one consists of a change in the perceived returns to graduation, similar to Jensen (2010). I organize the results in two propositions.

Proposition 1 *(Production Function) The partial effect of changes in the assessed probability of the world is ambiguous, and depends on whether the student is over- or underconfident. Formally,*

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

Proof. See Appendix B.C for a full derivation. ■

The result of this derivative is *undetermined*, and it depends on the curvature of the $g(.)$ function and 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 and will respond with more effort, others will confirm that they are in the state of the world they believe in, so no response in effort is expected, while others could learn that they are in the bad state of the world and might become discouraged while others in the same position may become motivated and increase effort.

Proposition 2 *(Returns to Education) Optimal effort is increasing in the perceived returns:*

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

Proof. See Appendix B.C 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 underconfident or overconfident. An increase in perceived returns to education should lead to an increase in effort.

3.3 Predictions of the Model

To illustrate the potential predictions of this model, I use a functional form that fulfills the assumptions made so far:

$$g(x) = x - \omega x^2$$

and the hand-picked values are specified in the footnote of Figure 1.

I show how the provision of information could have an impact on beliefs and therefore the probability of graduation. The *Production function* arm can change perceptions of the state of the world ($\hat{p}$) of students by making them map their academic standing with the average graduation for that standing and the effort that it implies to achieve the diploma. Given that effort depends on the number of pending subjects the students have, the result will vary based on their baseline situation. The *Returns to education* is not expected to change beliefs about the perceptions of own ability, but to incentivize effort subject to how students react to the true values of earnings and employment.

Notice that the number of pending subjects at the beginning of the senior year is endogenous, meaning that students have some knowledge about the cost of effort they have to expend to pass subjects. 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 *Production function* only modifies the perception of $\hat{p}$, and *Returns to education* only modifies the perception of V.

3.3.1 Effects of the Production Function Treatment

Prediction 1 *Students with zero pending subjects are not likely to update their beliefs. The difference between states will be null, and as a consequence, their effort under the Production function treatment will remain the same as before.*

For this subset of students (in good academic standing at the beginning of the senior year), the main content of the intervention is not useful given that remedying the academic standing of students with pending subjects is given higher weight. Figure 1, Panel A, shows that there are no differences between the states of the world .

Prediction 2 *Students with pending subjects are more likely to update their beliefs under the Production function treatment.*

The treatment is more likely to affect the effort of these students given that their beliefs about the probability of graduation could change with the information provided. The question then becomes how they will adjust the weight in each potential state of the world Under different cost structures, different results can be observed, and they also depend on the parameters of $g(.)$. An explanation is presented in Figure 1 Panel B, which shows that the difference in marginal benefits will be greater according to the parameters in a good or bad state of the world. First, notice that the dashed gray line (expected self-perception of graduation) will move according to the perceived values of $\hat{p}$ and the updates after the intervention. Second, to explain the predictions I consider the extreme cases of $\hat{p}$—an overconfident student ($\hat{p} = 1$) and an underconfident student ($\hat{p} = 0$)—before my intervention.

- **Prediction 3** *Underconfident students will increase their effort after receiving the Production function treatment.*

  According to my definition of confidence, the probability of graduation for an underconfident student is higher than their perception. This student thinks they are in a bad state, when in fact they are in a good one. In the graph, the marginal benefit of graduation for this type of student is on the black line, after the intersection of curves with a lower marginal cost. At baseline their equilibrium is at e_{lc}^{u*}, where lc indicates that they perceive lower costs, and u^* indicates the baseline equilibrium. With the intervention, they learn that they have a better standing (their $\hat{p}$ will increase), meaning that their effort will increase too, as can be observed in the graph. Notice that the level of increase depends on the magnitude they update $\hat{p}$.

- **Prediction 4** *Overconfident students will increase their effort after receiving the Production Function treatment.*

According to my formal definition of confidence, overconfident students think that $\hat{p} = 1$ (they perceive they are in the good state of the world). Following the previous explanation, and for this particular case, this type of student thinks they are on the left hand side of the gray line (before the intersection of states). Notice these students have a high cost of effort, so the initial equilibrium is e_{hc}^{o*}, where hc indicates higher costs and o^* indicates the baseline equilibrium for an overconfident student. After the provision of information, students learn that $\hat{p} < 1$. If they update correctly, the equilibrium will move to the right, meaning that their effort will increase under this scenario *too.*

3.3.2 Effects of the Returns to Education Treatment

Prediction 5 *An increase in perceived returns to education should increase effort and, as a consequence, graduation.*

Under this treatment, previous literature Jensen (2010) and Nguyen (2008) indicates that students with low perceived returns to education should update educational achievement (in this case, graduation) in the correct direction (up).

Prediction 6 *A decrease in perceived returns to education should decrease effort and, as a consequence, graduation.*

Following the previous reasoning, those students who perceived high returns to education at baseline should decrease their effort.

3.3.3 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:

- *Production function:* Students update their beliefs about the right state of the world they are in, and they correct the level of effort to obtain a high school diploma.
- *Returns to education:* 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 Salta, Argentina, from August 2019 to November 2019. The details of the population and the design of the experiment are discussed below.

To characterize my sample, I show some statistics for the control group in Table 1. Only 50 percent of the students in their senior year finish high school in a timely manner. They overestimate their probabilities of graduation in the baseline survey and in the endline question. On average, students had 0.9 pending subjects at the beginning of the year and they do not tend to improve this situation; by the end of the academic year the average remains high (0.8 pending subjects). The percentage of students with at least one pending subject is 55.27 percent and the average number of pending subjects for those students is 1.6.

4.1 Ethical considerations

This research project required IRB approval. Given that some minors (according to the Argentinian law, individuals aged less than 18 years old) are included in the sample, consent from parents and students was sought following the instructions of the IRB office at Brown University, the school principals, and authorities from the Ministry of Education of Salta. In addition, material prepared for students was approved by the Ministry of Education (contents for the online platform, survey instrument, and presentations) without their being informed in advance which information treatment arm to which a school would be randomly assigned.

4.2 Sample

The eligible population for this study is students attending their senior year at public high schools in Salta.[9] While some schools can have more than one shift, I only considered the morning and afternoon shift due to logistic/budget constraints. Power calculations were conducted using information from the academic year 2018. There were 2933 enrolled students in the senior year across 63 school-shifts. The unit of randomization is at the school-shift level given that randomization at the individual or class level would be more likely to contaminate the control group.

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

4.3 Timeline

At the beginning of this project, by mid-October 2018, initial contacts were made with authorities of the Ministry of Education of Salta, and the office in charge of supervising my intervention was the Directorate of Secondary Education. They have overseen all the stages of the intervention. In addition to having their approval, I needed the direct approval of each school's principal and vice-principals, who were more aware of the specifics of each shift: school festivities, exams, and trips.[10]

This process finished in the first quarter of 2019 (see Figure 2). At the same time I requested from the directorate access to five "representative" schools to collect individual data about school performance and graduation. This administrative data were not available, so I followed their recommendation to collect data that was stored in secured rooms at each school building. The main intentions were to compute statistics at the individual level for use in the *Production function* treatment arm and to confirm that the graduation rate is in fact approximately 50 percent, in large part owing to the pending subjects issue (see more details in Appendix A).

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

A summary of the main milestones of the project follows below:

October 2018 — **Negotiations started with the Ministry of Education**.

[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".

November/December 2019	**Qualitative analysis started:** To understand the reasons behind low graduation rates, I conducted semi-structured interviews with senior students and schools' authorities. This step was extremely important, given that the issue with the pending subjects and connections with the self-confidence of students was detected in this way.
January 2019	**Approval was obtained and IRB procedure started**
March/June 2019	**Meetings with schools principals**
August/November 2019	**Launch of the information interventions and collection of the baseline survey**
February 2020 to March 2021	**Data collection of main outcomes:** Academic information at the individual level is not digitized and contains relevant information such as graduation. The process was interrupted because of the COVID-19 pandemic, and the entire process to analyze results was delayed. In the meantime, I started negotiations with the main post-secondary institutions in Salta to obtain information about university enrollments.

4.4 Data

Baseline Survey

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

A description of the baseline data collection process follows. At least 2 days before the intervention date, the research team visited and delivered to school administrators envelopes containing consent forms for parents of students attending the senior year. At a date and time agreed with the school administrators, the team met with all students of the school in a room.[11] A description of the activities conducted during each visit day is shown in Figure 3.

At the meeting with students, the research team was introduced by school administrators. Then, tablets were given to students, a short presentation (containing slides with pictures) was shown to instruct students on their use, and the students were asked to fill out the questionnaire. At the same time, a brief explanation of the questionnaire was provided.[12] Afterward, the research team showed a presentation introducing the online platform. If applicable, the information treatments were then conducted. After the presentation, the research team asked students to answer an additional question about their perceptions of their own graduation. This question was intended to test for any changes in students' perceptions after hearing the information presented.

As mentioned in the previous section, visits to the school lasted at most one hour. During that time the baseline survey was collected, the platform was introduced, and the information interventions were presented (if applicable). Given that a single presentation, including statistics and unknown facts for the students, could not have been enough to change students effort, I sent SMS and/or email two weeks before each examination date (to pass pending subjects and failed subjects during the senior year) to briefly reinforce the treatment received.[13] As was shown in previous papers, reminders can help to boost information interventions (Damgaard and Nielsen, 2018).

School Academic Records

I collected information about academic performance after the end of the 2019 academic year. The end was in February 2020 (during the February examination period). As shown in Figure 2, this process was heavily delayed by almost one calendar year because of the closure of schools in response to the COVID-19 pandemic. Those individual records contained data on performance during the entire school year and graduation, as well as information about the pending subjects (if any) and attendance on examination dates for pending and failed subjects during the senior year. An example of an individual record is in Figure C1, Appendix C.

[11] No authority knew beforehand which treatment was randomly selected for each school.

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

[13] Cellphone numbers and email addresses were collected during the baseline survey. See the reminders in Appendix A.

Administrative Records

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

4.5 Experimental treatments

The treatment assignment was randomly determined at the school level stratifying by number of students and geographic area of Salta. All school visits included the presentation of a free online platform with math content (see Appendix B), and its use is not part of this analysis. Information interventions considered in this study are described below.

Control: No information treatment was provided.

Production Function: Using data from a subset of students of the previous cohort (2018), I computed the mean of a dummy variable that indicates the rate of on-time graduation (by December 2018, after the December examination period) by pending subjects and no pending subjects at the beginning of the senior year of that cohort. 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 pass pending subjects and subjects from the senior year. A full description of the treatment is in Appendix A.

Suggestions about *how* to improve academic standing were provided to all students (because at the time of the visit the status of each student was unknown). All of these suggestions were *intermediate steps* to effectively transform inputs into outputs. The information provided included the following: request mock exams (*modelos de examen*) from teachers[14], ask for study material from classmates or

[14]These exams should be available for every subject and all years, as was requested by the Directorate of

students from younger cohorts (given that the teachers employed by the schools and the required academic material can change over time), talk with teachers in advance to ask them for studying recommendations, or ask which teachers will be a part of the committee in each subject. [15]

Returns to Education: Students might not be aware of the disadvantages of not finishing high school and the impacts on their labor market prospects. The provision of this information should incentivize students to obtain a diploma on time (in order to attend college or find a job in the formal sector). This piece of information is akin to Jensen (2010). In my case, I use data from the National Household Survey (second semester of 2018), restricting the sample to Salta and individuals aged 18-30 who are not currently attending any level of education and are employed. 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.

A description of the randomization and participation results are provided in Figure 4. Only one school principal with two shifts (out of 64 schools) refused to participate, even though I had the authorization from the Directorate of Secondary Education. After several conversations, the reasons were not disclosed and authorities of the Ministry of Education preferred not to force the school principal to participate. Another school was excluded from the analysis due to serious administrative complications in the implementation (following **?**).

Students' participation differed between the intervention treatment arms. A higher percentage of students, their parents, or both, decided not to participate in the *Production function* 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 allowed me to discard selection in participation. No school authorities knew beforehand which treatment was assigned to their school. The research team itself only knew which treatment should be implemented 30 minutes before the arrival to each school.

Secondary Education for all public high schools since 2018. Given that compliance of all the teachers could not be verified before the intervention, this information was included in the presentation, highlighting the fact that it was mandatory for teachers to prepare that material.

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

4.6 Measuring Students' Confidence in Graduation

To measure students' self-confidence about graduation, I use two sources of data: the baseline questionnaire and administrative data that provide information about the graduation of each student. I use a question that asks about the self-estimated probability of graduation as a *subjective measure* (see Figure C3, which was used in the questionnaire) and a set of observable characteristics of the students and their households to predict the probabilities of graduation, *objective measure.* For this last step, I first only consider observations in the control group and then extrapolate the predictions to the entire sample.

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

Figure 5 shows in Panel A the distribution of the estimated probabilities for students with zero pending subjects, and in Panel B the distribution of the difference with respect to the self-estimation of students' graduation. Figure 6 shows the same distributions for students with at least one pending subject. According to my definition of confidence, students with a positive difference are classified as underconfident (objective measure is higher than the subjective one) and those with a negative difference as overconfident.

5 Results

5.1 Description of the Control Group and Balance Checks

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

$$y_{is} = \beta_0 + \beta_{PF} ProductionFunction_s + \beta_{RE} ReturnsEducation_s + \delta_s + \epsilon_{is} \qquad (1)$$

where y_{is} is the outcome of interest for student i who attends school-shift s, the dummy variables $ProductionFunction_s$ and $ReturnsEducation_s$ indicate which information treatment school s received, δ_s indicates the strata fixed effects (Bruhn and McKenzie, 2009), and ϵ_{is} are robust standard errors 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). Each row shows results from a separate regression. Columns 4 and 5 show p-values of the tests if $PF = RF$ and $PF = RF = 0$, given that the comparison of the two information treatments is of special interest.

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

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

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

Panel E shows the variables that indicate expectations. Ninety-five percent of the participants stated that they want to attend college the next academic year and also 84 percent are interested in looking for a job after the end of the school year. At the time of the school visit, students perceived that their chances of on-time graduation were 78 percent. None of these variables exhibit statistically significant differences between treatment arms.

5.2 Empirical Strategy and Main Results

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

$$y_{is} = \beta_0 + \beta_{PF} ProductionFunction_s + \beta_{RE} ReturnsEducation_s + \delta_s + x'_{is}\omega + \eta_{is} \qquad (2)$$

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 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 depicted 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 *Production function* treatment arm are 5 percentage points more likely to graduate (10 percent with respect to the control group) and (2) those in the *Returns to education* are 10 percentage points more likely to obtain a diploma (20 percent with respect to the control group). I find that the differences associated with these treatments are statistically significant.

The effect of *Returns to education* is twice that found in a subgroup of less poor students in Jensen (2010) (the author does not find an impact for poor students). A potential explanation for the higher impact in the current study could be related to the fact that the target population was students who were closer to receiving their high school diploma. Additionally, my setting had fewer economic barriers: enrollment and transportation to school are free. The *Production function* effects are the same in magnitude as in Jensen (2010) but they apply to the entire sample in my study. This outcome shows that the treatment—simply talking about the probabilities of graduation (conditional on academic standing) and intermediate steps to transform inputs into outputs—is effective in increasing educational achievement, by only using data available in schools.

According to my hypothesis, not all students will experience the same impact from the *Production function* treatment. In Table 3, columns 2 and 3 show the treatment effects by academic standing, with students separated according to whether they are in good standing (zero pending subjects) or in bad standing (at least one pending subject). As expected, I observe no significant effect on students in good standing and the magnitude is close to zero. This is *Prediction 1* in my model. A likely reason for this finding is that these students already know how much effort they should devote to study to succeed. This is not the case for those students in bad standing. The information provided should help them to realize where to put the effort needed to obtain a diploma. For this subset of students, I observe an increase of 7 percentage points (more than 30 percent with respect to the control group). This is *Prediction 2* in my theoretical framework. With respect to the *Returns to education* arm, both groups are found to have positive and statistically.

5.3 Mechanisms for Production Function and Returns to Education

Perceptions on Graduation and Updating

To understand the drivers of these results, I tested the role of self-perception of graduation on actual graduation (Table 4) by using the answers to the questions about chances of graduation before and after the interventions. An important part of the *Production function* treatment was to make students aware of the correct shape of the production function of the high school diploma based on their academic standing at the beginning of the senior year. As previously mentioned, at the time of the intervention, the standing of the students was their private information and the goal was to allow students to create a *mapping* of their situation with regard to graduation rates of similar students from the previous year.[16]

The question about the self-perception of timely graduation was included in the baseline survey, conducted before the intervention treatments, and it was then repeated at the end of the visit. Under the theoretical framework shown above, perceptions of graduation should only change if students updated their beliefs about the level of effort needed to obtain their diploma. This is only possible if they receive information about the actual probabilities, the effort that is required, and all the intermediate steps needed to successfully transform effort into graduation. Table C1 shows that individuals who received the *Production function* treatment became more accurate with respect to their own chances of graduation: the variable decreases by 2 percentage points from the baseline response.[17] I observe that in the experimental outcome they become more accurate, but this result could not be transmitted into effective effort to remedy their standing. As expected by the design of the treatments, the most striking and significant differences are observed in the *Production function* arm.

I analyze graduation by academic standing and its relationship with my definition of con-

[16] Students' academic records are not digitized, and it was not possible to target the information in a separate way. Instead, all the information was shown to all students.

[17] Notice that the students in the control became less accurate (more optimistic about their chances of graduation). A reasonable explanation for this result is that the visit to the school from a member of a university in the United States and members from Universidad Nacional de Salta (UNSa) could have per se generated a optimistic response among students, given that there is almost no formal connection between secondary and post-secondary levels.

fidence in Table 4. Although differences exist between the underconfident and overconfident students (in both treatment arms), the differences with the largest magnitude are observed in the *Production function* arm for students in bad standing. *Prediction 3* and *Prediction 4* of the theoretical framework, and under the particular case analyzed in Section 3, are verified. There are positive and statistically significant effects (at the 5 percent level) for both under- and overconfident students, with a difference of 20 percentage points (but nonsignificant) in favor of those underconfident students.

This outcome indicates that even though after the presentation of that treatment students classified as overconfident became statistically more accurate (Table C1), but that effect fades away until the end of the academic year.

Effort

Table 5 measures the role of effort devoted to the study time of pending subjects. I analyze the effect of the information treatment on three variables that indicate direct measures of effort to pass pending subjects: (1) enrollment to examination period, (2) attendance to examination period, and (3) passed pending subjects before the end of the academic year (February 2020). The first variable indicates the degree of effort because according to high school rules, only students who explicitly register for the examination dates with a secretary are allowed to take the exam.[18] The second variable indicates if the students did in fact attend the examination, and the third variable is a dummy that indicates if the student passed at least one pending subject. I did not restrict the last two variables to enrollment or attendance, respectively.

Table 5 Panel A shows positive impacts of the information treatments on these outcomes. In particular, for those who received the *Returns to education* treatment, the effect is statistically significant at the 1 percent level for columns 2 and 3.

Panel B shows the effect of the information treatments according to confidence. As discussed above, underconfident students are those with a greater response to the treatment shown by exerting more effort; the difference in the *Production function* between the two types of students is more than 40 percentage points (significantly different at the 1 percent level). The *Returns to education* treatment arm also has differences in favor of the underconfident students, but they are lower; only in column 2 is the difference with respect to the overconfident students significant (at the 10 percent level).

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

Perceptions of Labor Market Outcomes

In the baseline survey, I asked students to form a perception of expected earnings (employment and earnings, by level of education). They could have a positive misperception (meaning they overestimate the returns to education, relative to the true values) or a negative one (underestimation of returns to education). I was not able to collect the same information after the intervention (and check for updates in perceptions) because this section was very time consuming for the students and I had limited time to conduct the interventions.

According to previous findings (Jensen (2010), Nguyen (2008)), students who underestimate actual returns are those who are going to be positively affected by the returns to education treatment. I test this hypothesis by creating a variable of "expected returns" using the perceived earnings and probabilities of employment by level of education collected in the baseline survey. Then, considering the "actual" expected returns, I create two dummy variables: Misperception (+) when the student perceives that the expected return is higher than the actual return and Misperception (−) when the student perceives that the expected return is lower than the true value.

Table 6 shows the impact of these misperceptions at baseline on graduation, considering the returns to two levels of education: complete secondary and complete college. I focus here on the students who received the *Returns to education* treatment. Both those who misperceived expected earnings (for complete secondary and complete college) in a negative way and those whose misperceptions were positive at baseline have positive magnitudes, confirming *Prediction 5* but not *Prediction 6*. The magnitude of the effects is higher for students with a positive misperception, although the difference in coefficients is not statistically significant.

When I provide information about the true returns to education, students weight their prior beliefs with the new information, and they could subsequently decide which piece of information to assign a higher weight. Based on previous results in the literature, students with a negative misperception are expected to update their beliefs upward and graduation will increase. But the aggregated result depends on the percentage of students who assign a higher weight to their prior beliefs or to the new information, which may explain why I cannot verify *Prediction 6* in my results.

Time Preferences

The *Returns to education* treatment implies a forward-looking behavior on the students' side, given that they have to wait a considerable amount of time to see if the information provided actually improves their labor market outcomes.

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

It can also be observed that the magnitudes for both groups of students that received the *Production function* are lower, similar, and nonsignificant. This result is consistent with the information that was provided, that arm do not imply a forward looking behavior.

5.4 Heterogeneous Effects: Socioeconomic Status and Gender

In the baseline questionnaire, I did not include a question about family income due to the low response rate to that question in the pilot surveys. To generate a proxy for economic status, I use an index constructed by using variables indicating the ownership of air-conditioning, heating, a washing machine, a desktop or laptop, and whether the student's family lives in an overcrowded dwelling[19] and if 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 "least poor" otherwise.[20]

Table 8 shows that in the control group, students classified as poor have a lower graduation rate at 45 percent, which is 14 percentage points lower than the least poor students. In column 1, I demonstrate that contrary to previous findings (Jensen, 2010), less poor students are positively affected by both treatments: students in the *Production function* treatment arm are 8 percentage points more likely to graduate than the control group, and those in the *Returns to education* treatment arm are 14 percentage points more likely to graduate than the control group. Both results are statistically significant at the 5 percent level, and the difference of the magnitudes is also statistically significant at the 5 percent level.

Table 8 also shows the impacts by gender. Columns 3 and 4 show that female students are more likely to graduate than male students in the control group. However, both information treatments have a positive impact on both genders, with higher impacts observed for male students. I observe positive results of both treatments for both genders, and the differences are not statistically significant.

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

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

5.5 Other outcomes

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

University enrollment

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

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

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

Table 9 column 1 shows that only 13 percent of the students in the control group are enrolled in university, and both treatment arms increase the probability of enrollment by 5 percentage points (almost 40 percent). These effects are statistically significant at the 10 percent level. The difference between treatments is not statistically significant, but they both represent a substantial improvement in higher education access. Bonilla-Mejía et al. (2019) present an experiment aimed to improve college enrollment in Colombia by providing information on returns to education for senior students and no effects were found. A potential explanation for my results is that the settings are completely different regarding access to post-secondary education.

Formal Employment

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

The system only allows access to information from the 6 previous months at the time of the inquiry.[21] Given the strict quarantine imposed by the government in Argentina in response to the COVID-19 pandemic, I decided to include information from the last quarter of 2020 (when some restrictions were lifted) to the first quarter of 2021. The output formal employment is a dummy variable equal to 1 if the participant was registered as a formal employee for at least one month out of those 6 months.

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

One key caveat is that the sample size in this analysis is lower than the original sample because I did not find information for all students in the administrative data—there were errors in IDs in the data I received from the high schools. To test for potential issues of attrition, I create a dummy variable equal to 1 if a student was not found and 0 otherwise. Then I run the main specification (described in the next section) and I do not find difference across treatment arms, as shown in Table **??** in Appendix C.

6 Conclusions

This paper analyzes the effect of information interventions to improve high school graduation by correcting students' mistaken perceptions by using a novel intervention and a traditional one. The first intervention, and the main contribution of this paper, is aimed at making students aware of their chances of graduation based on their academic standing at the beginning of the senior year and it teaches them how to effectively transform inputs into outputs (*Production function*). The second intervention shows information about the returns to education

[21] See Subsection 4.4.

based on the achieved educational level (*Returns to education*). Targeting which information could be helpful to students is of great importance.

Perceptions about students' probabilities of graduation and returns to education could be modified by collecting the correct information that targets each mistaken belief. As reported in previous papers, overconfidence could be a detrimental personality trait in an educational setting. Overconfidence in graduation is widespread in my sample, but I provide evidence that a piece of information, returns to education, could help more than other types of information to ameliorate the consequences of this negative cognitive bias.

In contrast to previous studies, the experiment is conducted in a unique setting. Many of the main economic barriers to high school education are not present, but high economic instability is observed. I observed positive and significant effects in both treatment arms on timely graduation, and the magnitudes are more significant than those found in other studies. I also found positive and significant impacts on college enrollment, while previous studies aimed at driving demand for post-secondary education did not find this effect.

Findings of this study have substantive policy importance: graduation rates can be improved in low-income settings by using an inexpensive intervention, with data collected from a previous academic year at a low cost. In contexts in which there are no feasible ways to compute returns to education, the use of information that is already available offers a great opportunity to improve graduation rates, as shown in this paper. Students who are positively affected by this intervention now have a previously unavailable chance to achieve economic mobility.

References

Aprender, M. d. E. d. l. N. (2017). *Aprender 2017: Informe de resultados, Salta, 5to a˜o de secundaria.*

Banks, J., L. S. Carvalho, and F. Perez-Arce (2019). Education, decision making, and economic rationality. *101*(3), 428–441.

Belloni, A., V. Chernozhukov, and C. Hansen (2014). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives 28*(2), 29–50.

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 (2018). Socio-economic database for latin america and the caribbean. https://www.cedlas.econo.unlp.edu.ar/wp/en/estadisticas/sedlac/.

Choi, S., S. Kariv, W. MÃŒller, and D. Silverman (2014). Who is (more) rational? *104*(6), 1518–1550.

Cotton, C., B. Hickman, J. List, J. Price, and S. Roy (2020). Productivity versus motivation in adolescent human capital production: Evidence from a structurally-motivated field experiment. National Bureau of Economic Research.

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

DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic Literature 47*(2), 315–372.

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).

Fryer, R. G. (2011). Financial incentives and student achievement: Evidence from randomized trials. *The Quarterly Journal of Economics 126*(4), 1755–1798.

Fryer, R. G. (2016). Information, non-financial incentives, and student achievement: Evidence from a text messaging experiment. *Journal of Public Economics 144*, 109–121.

Ganimian, A. J. (2020). Growth-mindset interventions at scale: Experimental evidence from argentina. *Educational Evaluation and Policy Analysis 42*(3), 417–438.

Ganimian, A. J. and R. J. Murnane (2016). Improving education in developing countries. *Review of Educational Research 86*(3), 719–755.

Glewwe, P. and K. Muralidharan (2016). Improving education outcomes in developing countries. In *Handbook of the Economics of Education*, pp. 653–743. Elsevier.

Heger, S. A. and N. W. Papageorge (2018). We should totally open a restaurant: How optimism and overconfidence affect beliefs. *Journal of Economic Psychology 67*, 177–190.

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

Kruger, J. and D. Dunning (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology 77*(6), 1121–1134.

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.

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

Nowell, C. and R. M. Alston (2007). I thought i got an a! overconfidence across the economics curriculum. *The Journal of Economic Education 38*(2), 131–142.

Rao, V. (2002). Experiments in 'participatory econometrics': Improving the connection between economic analysis and the real world. *Economic and Political Weekly 37*(20), 1887–1891.

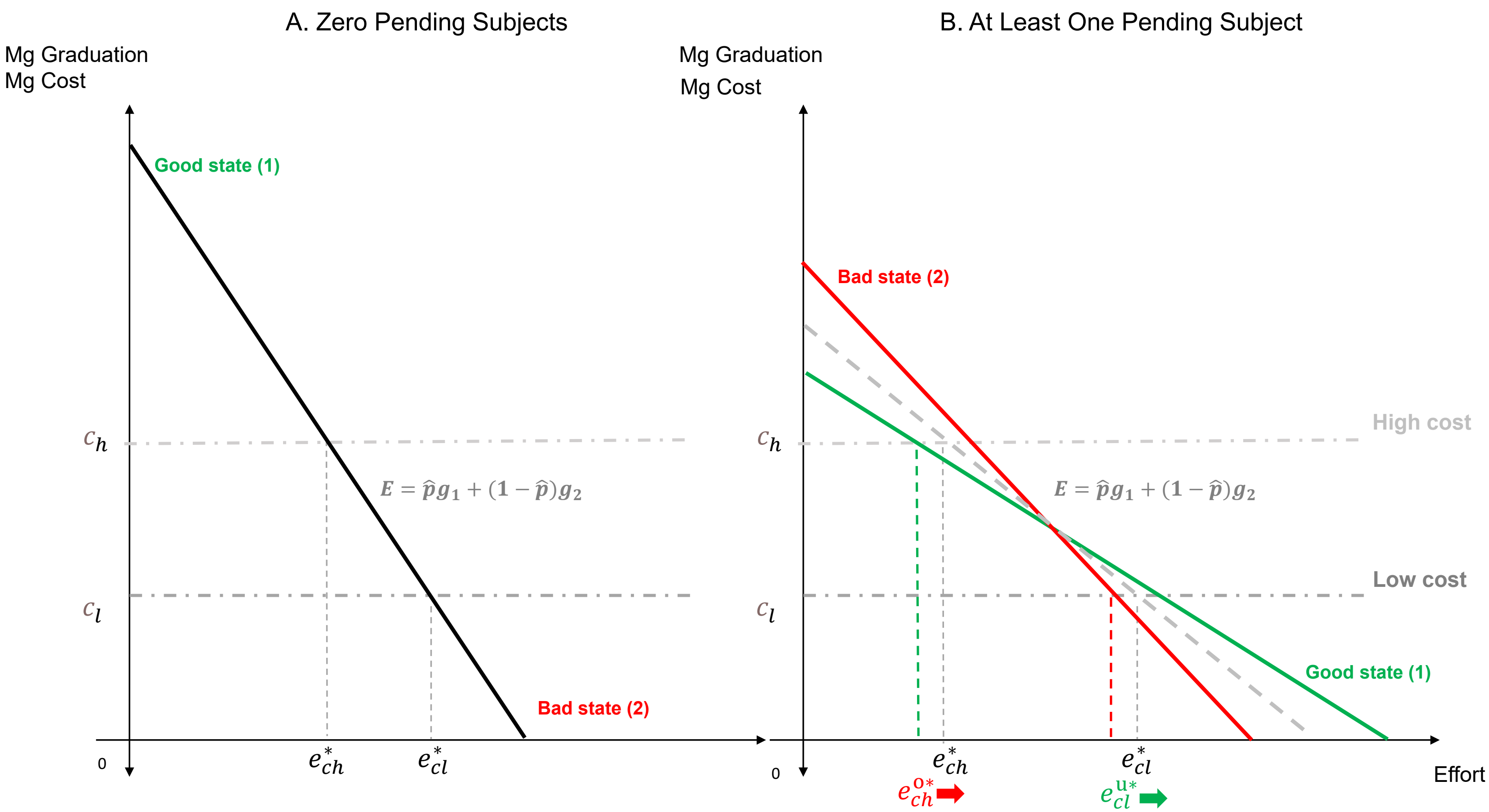
Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics 87*(3), 355.

Tornarolli, L. (2016). El fenomeno de los ninis en america latina. *CAF - Documento de trabajo 2016-18*.

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

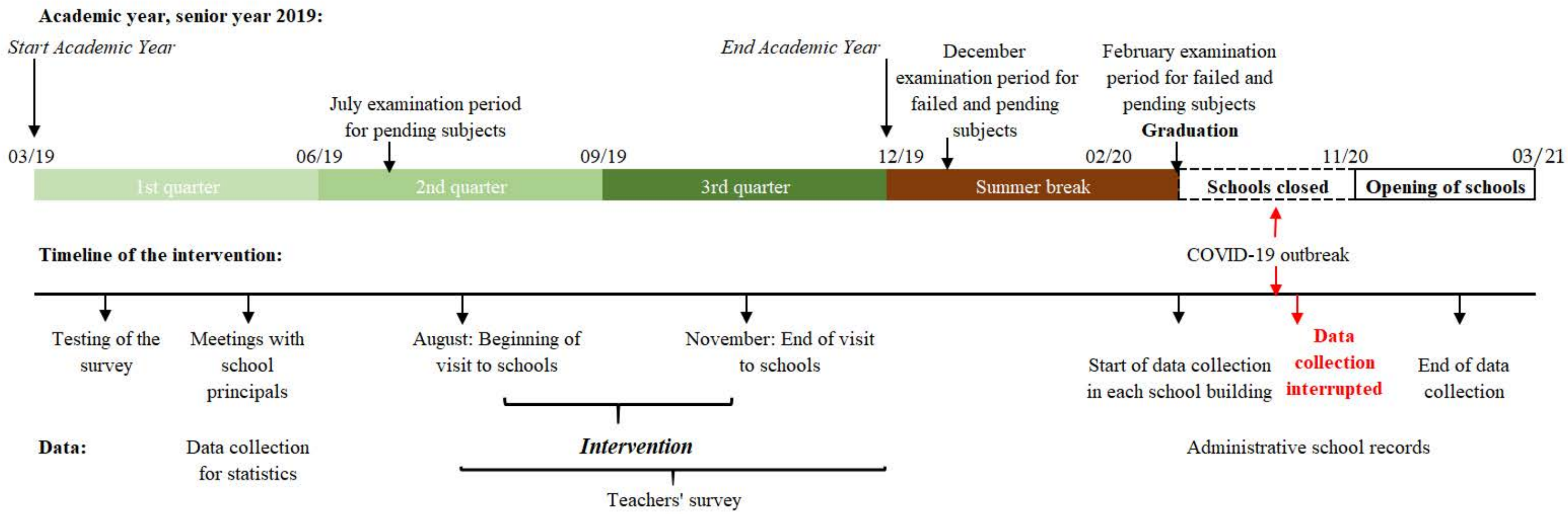
Figures

Figure 1: Marginal Benefits and Marginal Costs



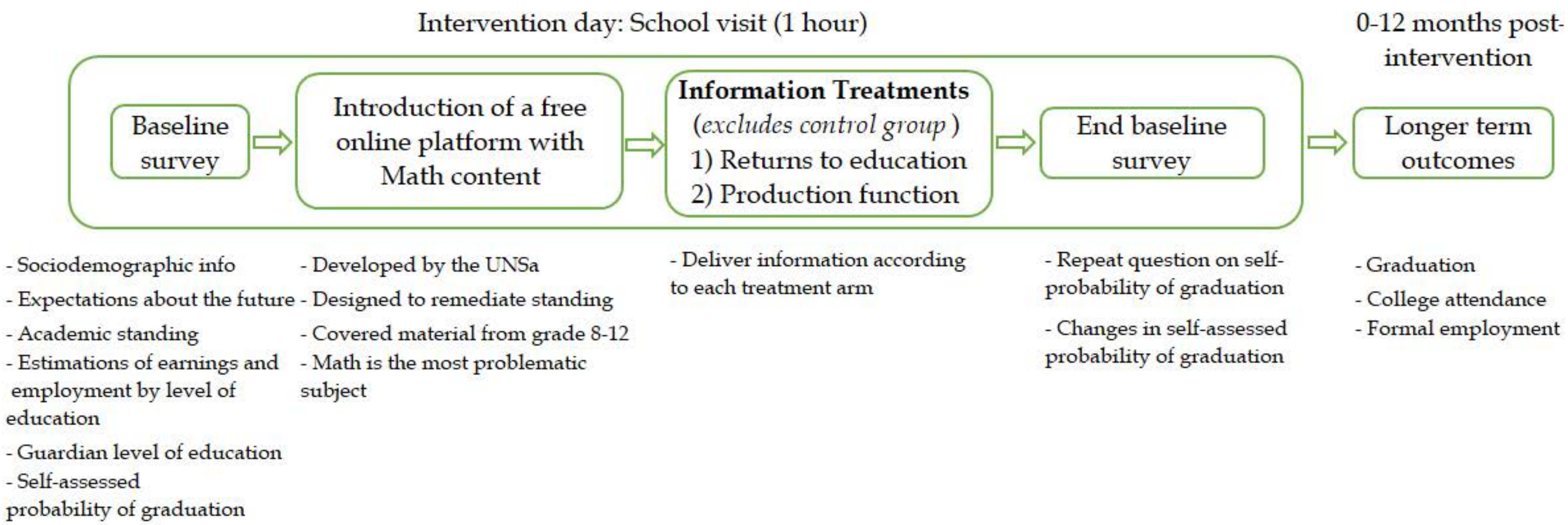
Notes: Values selected: $\beta = 1$, $\alpha = 1$, $\omega = 0.1$, $\nu_1 = 0.8$, and $\nu_2 = -0.4$. Panel A, students with zero pending subjects, shows that a change in $\hat{p}$ does not alter the equilibrium points for the two different cost structures. In Panel B, students with pending subjects, the marginal benefits of graduation are more salient in each state, with a change in $\hat{p}$, students' effort will increase no matter their type.

Figure 2: Timeline, Intervention and Data Collection



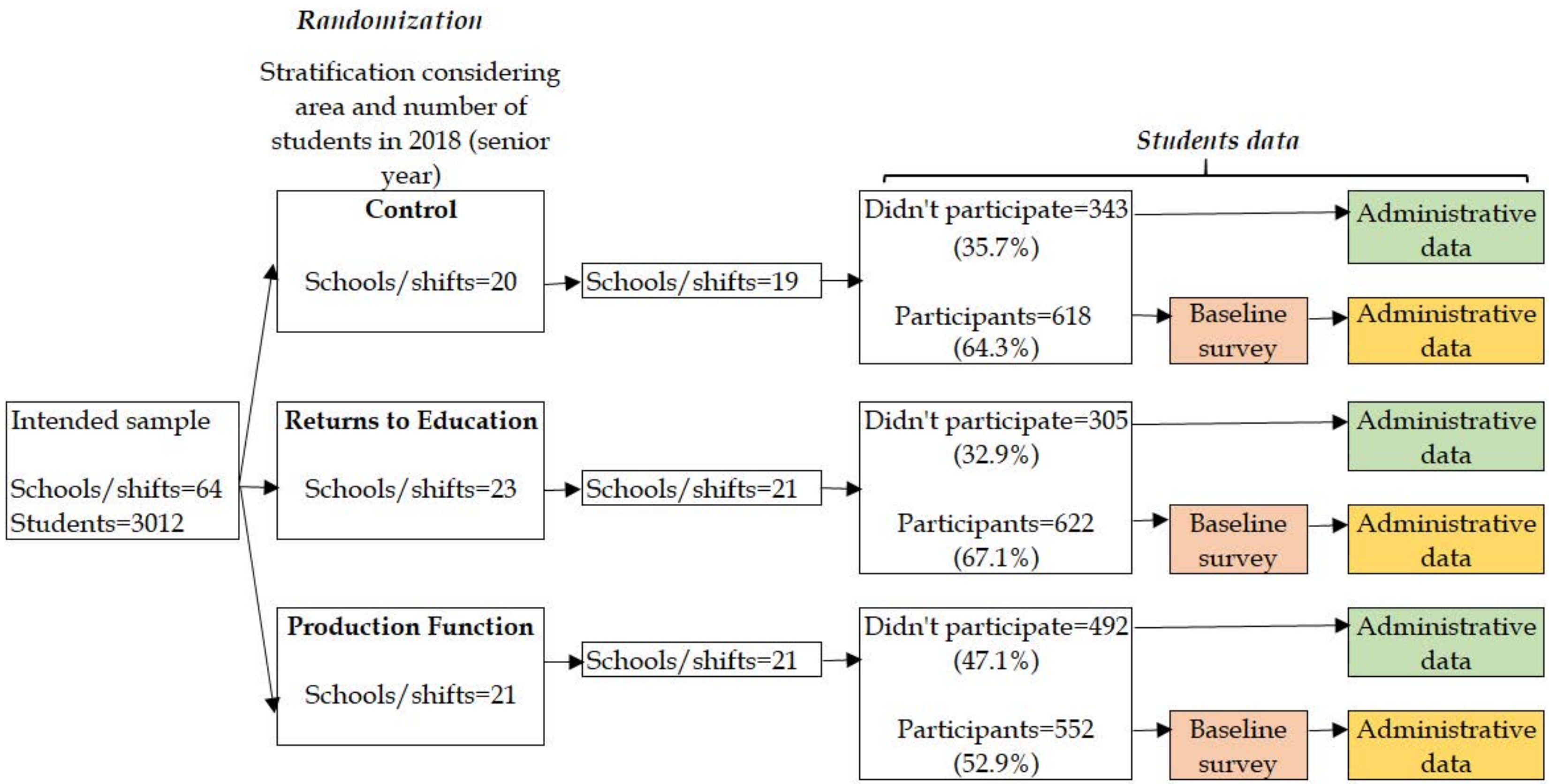
Notes: The design of the intervention was tied to the academic year of senior students in 2019.

Figure 3: The Intervention Day



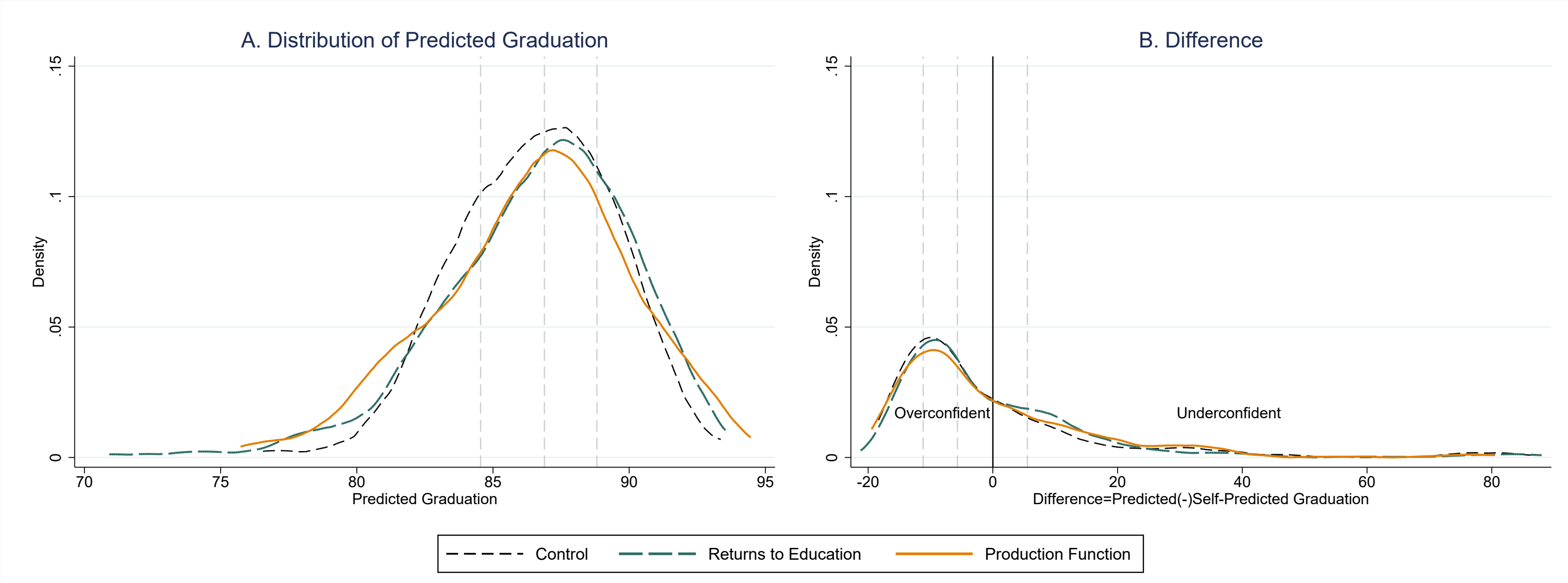
Notes: At the start of this intervention the questionnaire was tested in several rounds. Several corrections were made to improve students' understanding. The main change was related to the question used to ask probabilities of own graduation. A higher variability in responses was found using Figure C3 in Appendix C, so the question was asked in that way.

Figure 4: Randomization Design and Sample



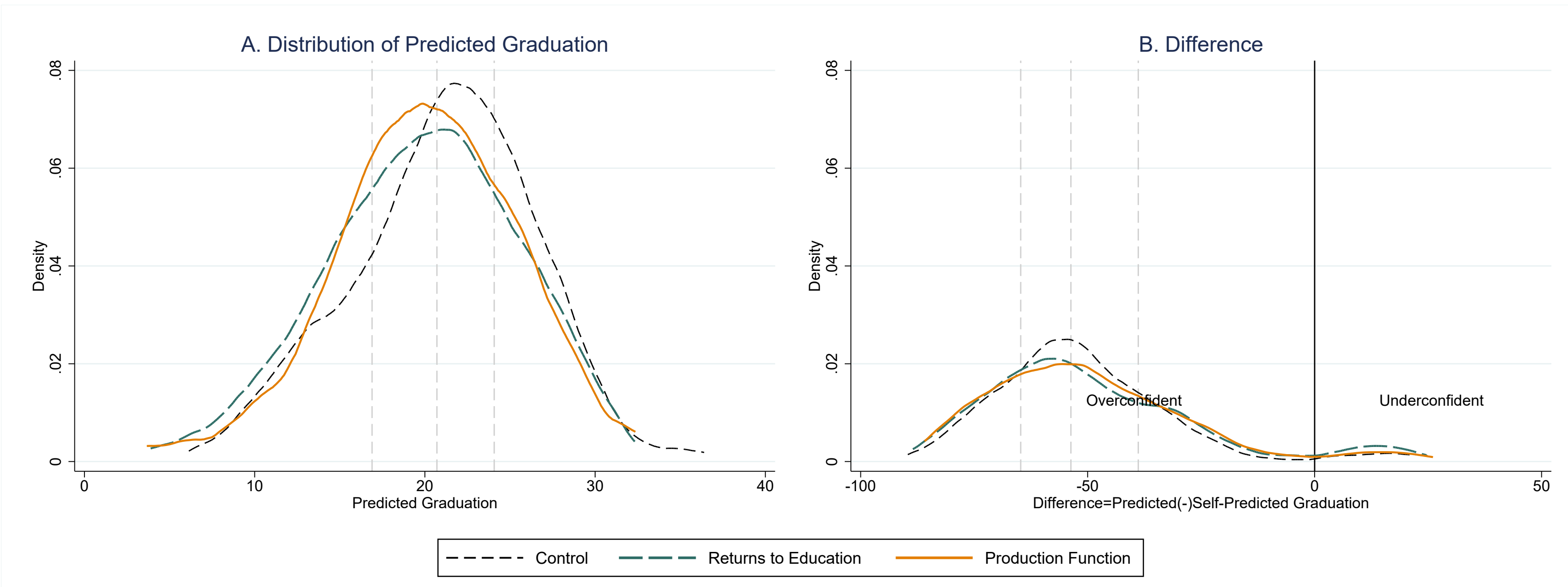
Notes: The first two columns of the diagram show the intended sample. The rest of the diagram shows the participation values.

Figure 5: Distribution of Predicted Graduation and Difference with Self-estimation by Treatment Group: Students with Zero Pending Subjects



Notes: Kernel density estimates. Vertical dashed lines indicate 25th, 50th, and 75th percentiles of overall distribution, respectively.

Figure 6: Distribution of Predicted Graduation and Difference with Self-estimation by Treatment Group: Students with at Least One Pending Subject



Notes: Kernel density estimates. Vertical dashed lines indicate 25th, 50th, and 75th percentiles of overall distribution, respectively.

Tables

Table 1: Descriptive Statistics from Control Group

	(1) Full Sample	(2) N	(3) Underconfident	(4) N	(5) Overconfident	(6) N
Graduation (by February 2020)	0.504	617	0.612	103	0.482	514
Students' Graduation estimation at baseline	0.784	615	0.569	101	0.826	514
Students' Graduation estimation at endline	0.842	601	0.740	101	0.863	500
Number of pending subjects	0.887	617	0.272	103	1.010	514
Number of pending subjects (if any)	1.604	341	1.867	15	1.592	326

Notes: Column 1 reports the number of non-missing observations of variables among all students in the Control group.

Table 2: Randomization Verification

	(1)	(2)	(3)	(4)	(5)	(6)
		Regression Coefficients		P-Value		
	Control Mean	Returns to Education	Production Function	Joint test R=PF	Joint test R=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. Student 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. Household 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 higher education	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. Student 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. *, **, and *** denote statistical significance at the 10, 5, and 1 percent levels, respectively.

Table 3: Impacts of Information on Graduation by Pending Subjects

	(1)	(2)	(3)
		Graduation	
	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 students. Missing values are recoded to the sample mean and separately dummied out. These missing dummies are also used to construct pairwise interactions. *, **, and *** denote statistical significance at the 10, 5, and 1 percent levels respectively.

Table 4: Impacts of Information on Graduation by Pending Subjects and Confidence on Graduation

	(1)	(2)	(3)
		Graduation	
	Graduation All	Zero Pending	At least One Pending
Production Function × Overconfidence	0.0300	-0.0372	0.0630**
	(0.0287)	(0.0234)	(0.0276)
Production Function × Underconfidence	0.0820*	0.0184	0.262**
	(0.0450)	(0.0591)	(0.131)
Returns to Education × Overconfidence	0.0920***	0.0184	0.123***
	(0.0298)	(0.0260)	(0.0346)
Returns to Education × Underconfidence	0.115**	0.0786	0.182**
	(0.0461)	(0.0544)	(0.0836)
Overconfidence	-0.109**	0.0975**	0.155***
	(0.0478)	(0.0410)	(0.0579)
P-value: PF × Overconfident = PF × Underconfident	0.381	0.376	0.139
P-value: RE × Overconfident = RE × Underconfident	0.696	0.358	0.549
P-value: PF × Overconfident = RE × Overconfident	0.020**	0.025**	0.089*
P-value: PF × Underconfident = RE × Underconfident	0.406	0.301	0.579
Mean (Control, Underconfident)	0.61	0.72	0
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.. See notes in Table 3 for a list of potential controls. *, **, and *** denote statistical significance at the 10, 5, and 1 percent levels respectively.

Table 5: Impacts of Information on Performance Conditional on Having Pending Subjects in December 2020

	(1) Enrollment for Examination Period	(2) Attendance to Examination Period	(3) At least 1 pending subject passed by the end of senior year
Panel A. No Interactions			
Production Function	0.030	0.055	0.062
	(0.065)	(0.036)	(0.041)
Returns to Education	0.042	0.13***	0.16***
	(0.074)	(0.039)	(0.039)
P-value: PF = RE	0.859	0.048**	0.041**
P-value: PF = RE = 0	0.832	0.005***	0.000***
Mean (Control)	0.62	0.44	0.28
Panel B. Interactions with Students' Confidence			
Production Function × Overconfidence	0.027	0.034	0.041
	(0.066)	(0.038)	(0.041)
Production Function × Underconfidence	0.020	0.46***	0.45***
	(0.12)	(0.13)	(0.13)
Returns to Education × Overconfidence	0.033	0.11***	0.15***
	(0.072)	(0.041)	(0.040)
Returns to Education × Underconfidence	0.11	0.38***	0.24**
	(0.12)	(0.13)	(0.11)
Overconfidence	-0.087	0.21*	0.11
	(0.066)	(0.11)	(0.082)
P-value: PF × Overconfident = PF × Underconfident	0.958	0.002***	0.001***
P-value: RE × Overconfident = RE × Underconfident	0.449	0.058*	0.431
P-value: PF × Overconfident = RE × Overconfident	0.931	0.031**	0.018**
P-value: PF × Underconfident = RE × Underconfident	0.514	0.518	0.099*
Mean (Control, Underconfident)	0.71	0.21	0.14
N	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. *, **, and *** denote statistical significance at the 10, 5, and 1 percent levels respectively.

Table 6: Impacts on Graduation by Perceptions on Expected Earnings by Level of Education

	(1)	(2)
	Graduation: Perceptions by Level of Education	
	Complete Secondary	Complete College
Production Function × Misperception (+)	0.0511	0.0772*
	(0.0312)	(0.0449)
Production Function × Misperception (-)	0.0717	0.0336
	(0.0438)	(0.0300)
Returns to Education × Misperception (+)	0.116***	0.126***
	(0.0346)	(0.0440)
Returns to Education × Misperception (-)	0.101**	0.101***
	(0.0425)	(0.0348)
Misperception (+) by Level of Education	0.00367	-0.0164
	(0.0336)	(0.0424)
P-value: PF × Misperception (+) = PF × Misperception (−)	0.711	0.433
P-value: RE × Misperception (+) = RE × Misperception (−)	0.777	0.646
P-value: PF × Misperception (+) = RE × Misperception (+)	0.024**	0.163
P-value: PF × Misperception (−) = RE × Misperception (−)	0.542	0.043**
Mean (Control, Misperception (-))	0.48	0.52
N	1609	1593

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 Misperception (-) by level of education (level showed at the top of each column), I consider that a student is accurate or is underestimating employment and earnings are being underestimated. See notes in Table 3 for a list of potential controls. *, **, and *** denote statistical significance at the 10, 5, and 1 percent levels respectively.

Table 7: Impacts on Graduation by Time Preferences

	(1) Graduation
Production Function × Above Median	0.0349
	(0.0364)
Production Function × Below Median	0.0394
	(0.0371)
Returns to Education × Above Median	0.117***
	(0.0347)
Returns to Education × Below Median	0.0438
	(0.0487)
Above Median Discount Factor	-0.0208
	(0.0402)
P-value: R × Very Patient = R × Not Very Patient	0.238
P-value: PF × Very Patient = PF × Not Very Patient	0.928
Mean (Control, Not Very Patient)	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 was 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. *, **, and *** denote statistical significance at the 10, 5, and 1 percent levels respectively.

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

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

Notes: Robust standard errors clustered at the school-shift level in parentheses. All regressions include graduation from the cohort 2018 at the school-shift level, and strata fixed effects. To classify students as Poor or Less Poor I created an index variable that includes ownership of household items and a dummmy 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. *, **, and *** denote statistical significance at the 10, 5, and 1 percent levels respectively.

Table 9: Impacts of Information on Other Main Outcomes

	(1) College Enroll-ment	(2) Formal Employ-ment
Panel A. No Interactions		
Production Function	0.052*	-0.014*
	(0.027)	(0.0087)
Returns to Education	0.054**	-0.022***
	(0.024)	(0.0076)
P-value: PF = RE	0.909	0.227
P-value: PF = RE = 0	0.059*	0.012**
Mean (Control)	0.13	0.032
Panel B. Interactions with Students' Confidence		
Production Function × Overconfidence	0.035	-0.0080
	(0.027)	(0.010)
Production Function × Underconfidence	0.092*	-0.040**
	(0.049)	(0.016)
Returns to Education × Overconfidence	0.047*	-0.026***
	(0.024)	(0.0088)
Returns to Education × Underconfidence	0.074	-0.0086
	(0.046)	(0.022)
Overconfidence	0.024	-0.00091
	(0.033)	(0.018)
P-value: PF × Overconfident = PF × Underconfident	0.160	0.098*
P-value: RE × Overconfident = RE × Underconfident	0.556	0.485
P-value: PF × Overconfident = RE × Overconfident	0.606	0.021**
P-value: PF × Underconfident = RE × Underconfident	0.637	0.064*
Mean (Control, Underconfident)	0.13	0.035
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 and 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. *, **, and *** denote statistical significance at the 10, 5, and 1 percent levels respectively.

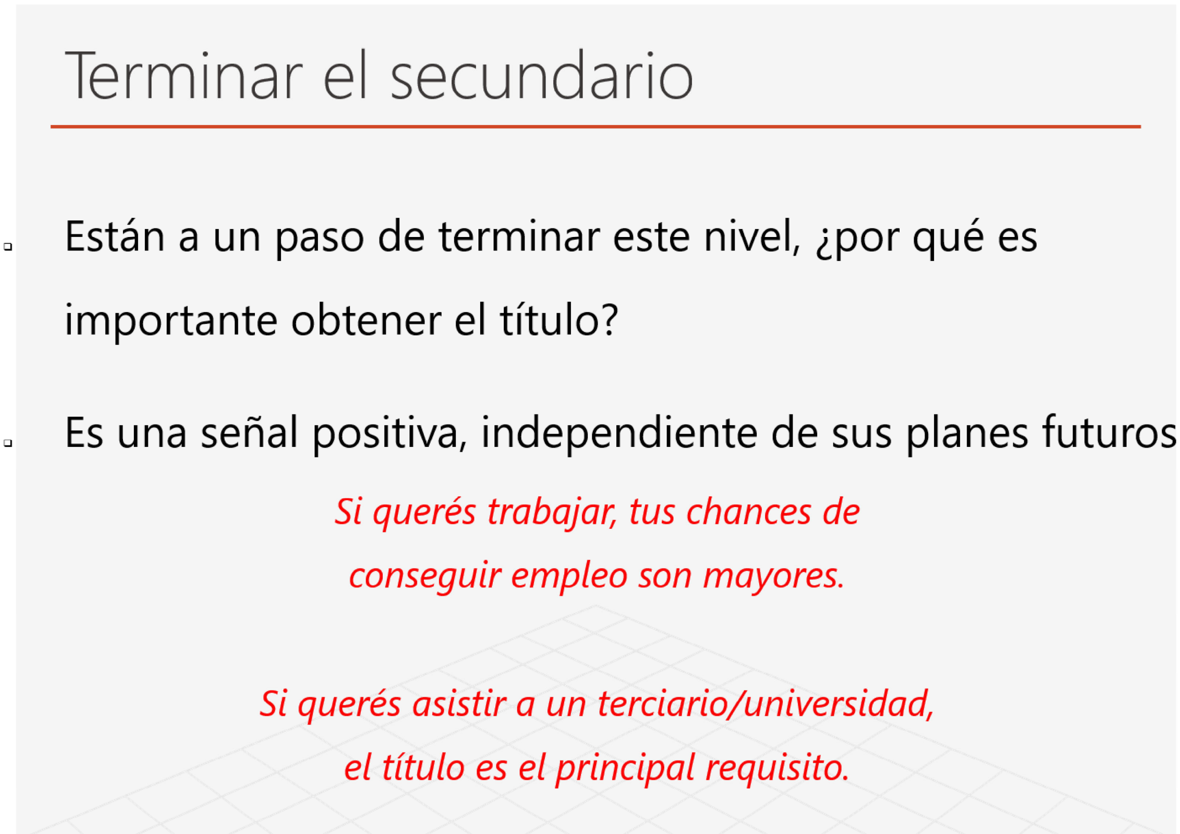
A Appendix: Information Treatment Arms

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 A1.

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

Figure A1: Why to Obtain the Diploma



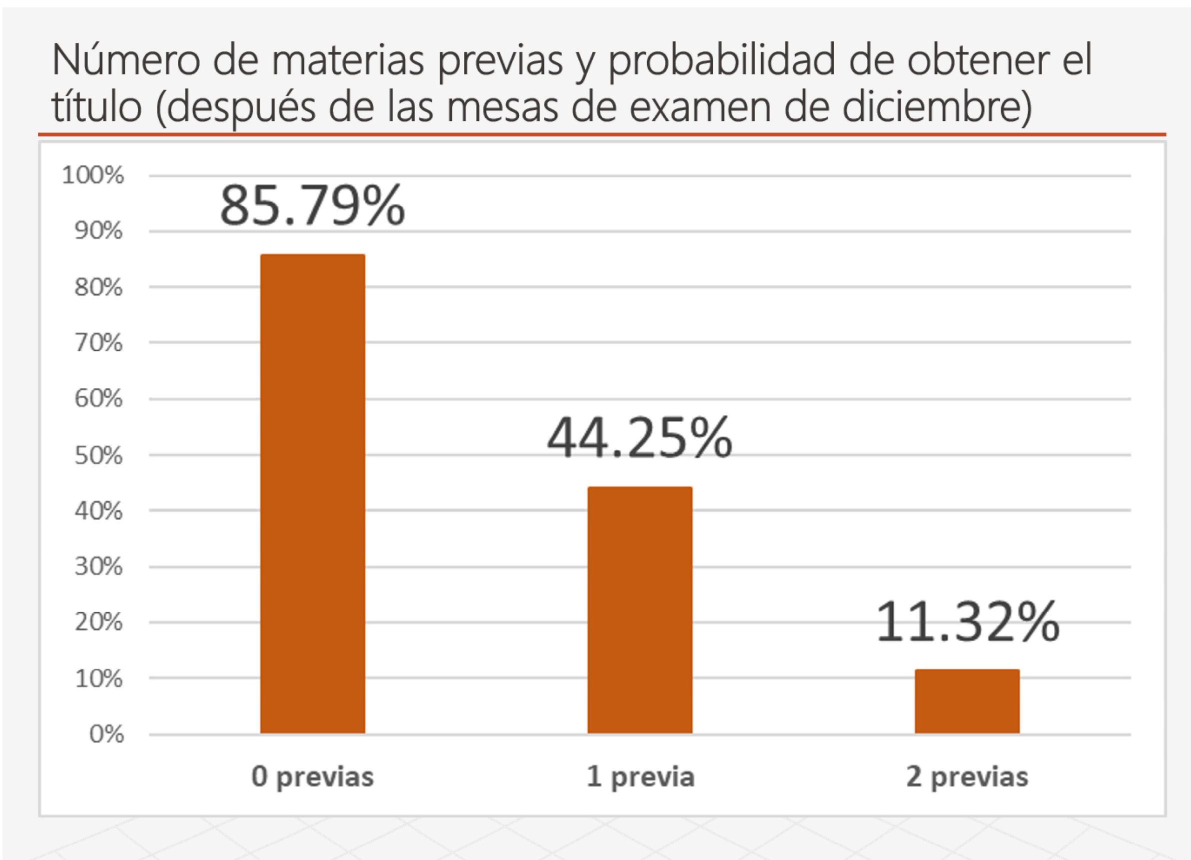
Notes: Common slide showed to all the students who received any of the intervention treatments. Translation: Finish high school, you are really close to finish this level of equation, but why it is important? It is a positive signal that does not depend on your future plans: If you want to work, your chances to get a job are higher or if you want to attend a higher level of education the high school diploma is the main requirement.

Production Function

I showed information about graduation rates from the previous cohort (senior students in the 2018 academic year). It was intended to emphasize the pervasive effects of the pending subjects that the students do not pass during their senior year on the probability of obtaining a diploma and also how important it was to pass the subjects during the senior year. 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 the 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 A2). 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 the students in any 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 (Figure A3), and then I showed a summary of the most relevant tips to effectively obtain a diploma on time.

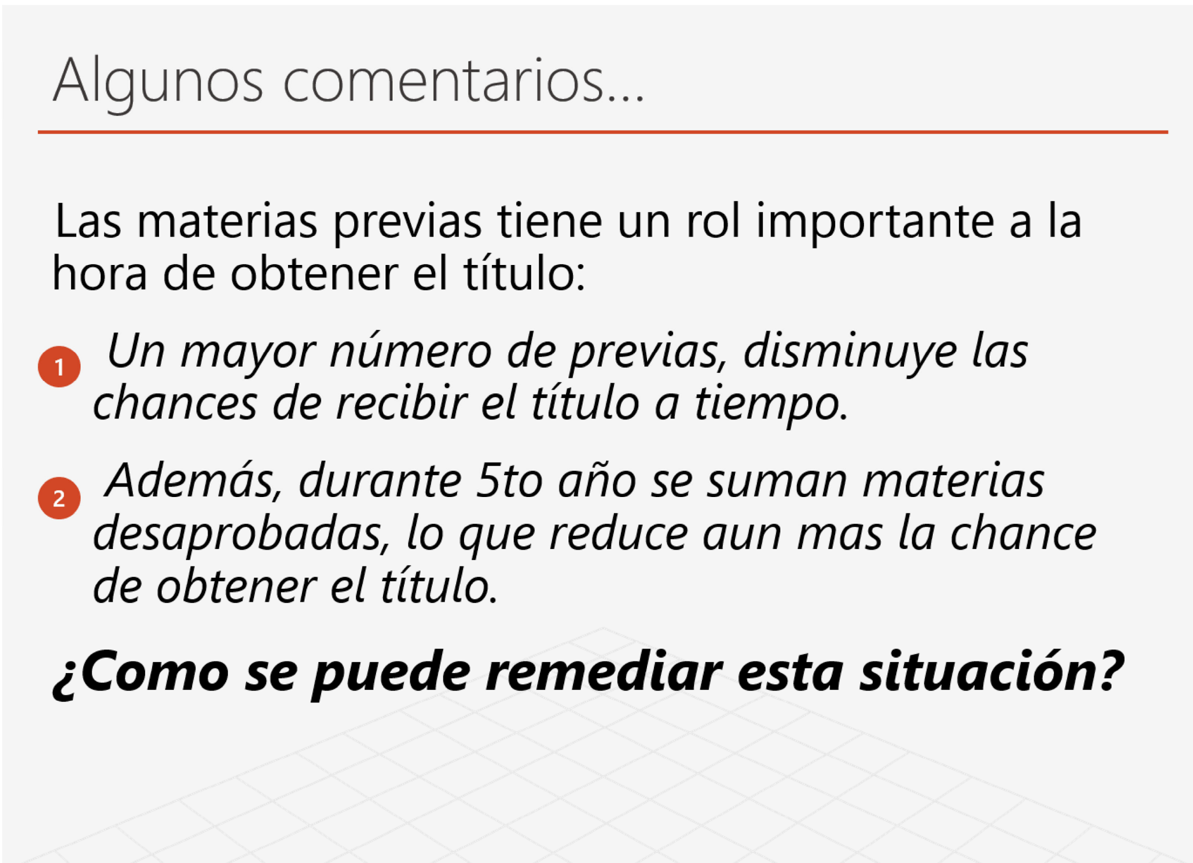
Figure A2: Statistics Shown to the Students



Notes: Own estimations based on a sample of representative schools in the capital city of Salta including students from the senior year during 2018.

The key messages were (1) to devote more time and effort to study the senior year subjects and (2) to attend the examination periods (for those with pending subjects). The 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.

Figure A3: The Role of Pending Subjects

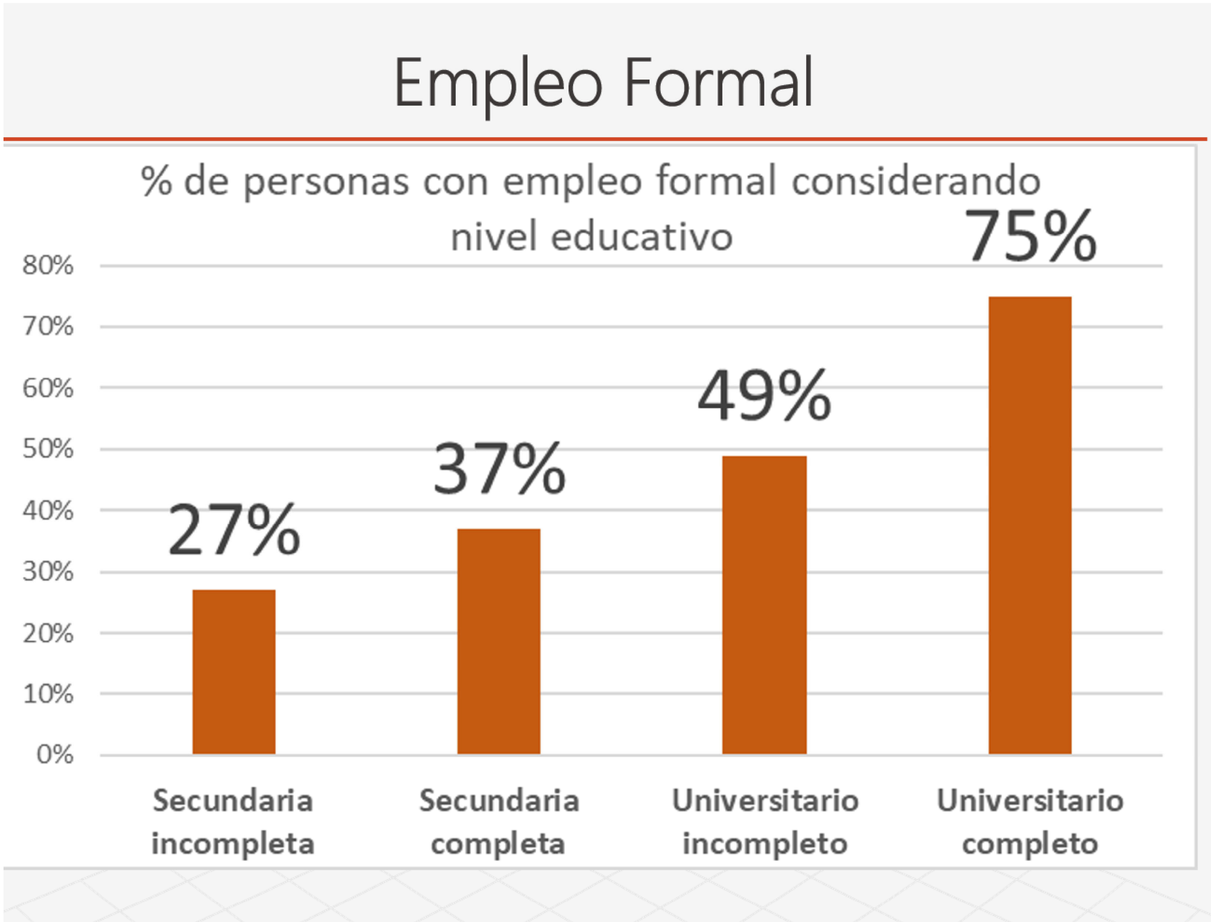


Notes: In this part of the presentation, I highlighted the role of the pending subjects and passing senior year subjects on timely graduation. Then I opened the discussion with a question, "How can this situation be remedied?"

Returns to Education

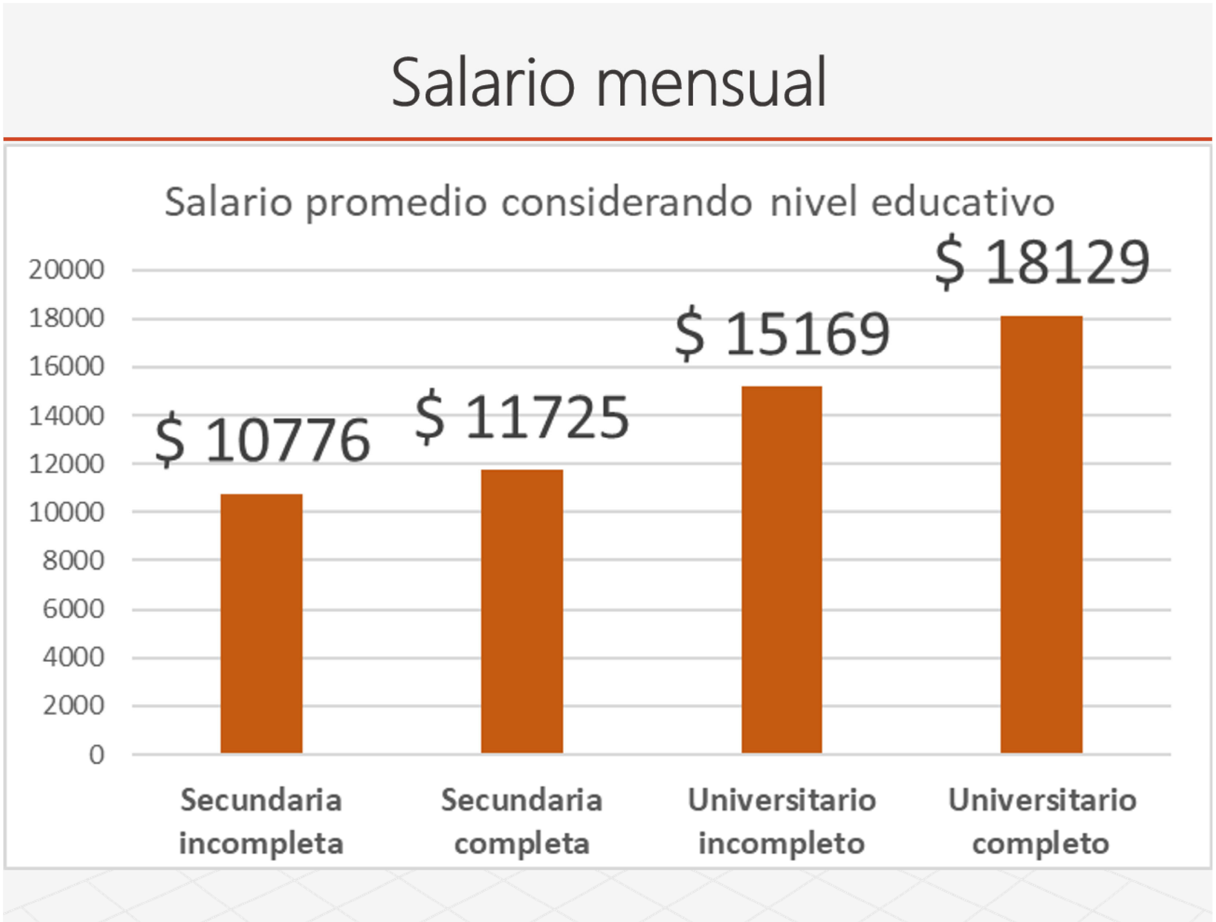
In this presentation I used data from the National Household Survey 2018 (Encuesta Permanente de Hogares) to compute the averages of formal employment and earnings to be shown to the students. I only considered individuals from the province of Salta, between 18 and 30 years old. The statistics were computed according to the level of education and are shown in Figures A4 and A5.

Figure A4: Formal Employment 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.

Figure A5: 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 $\approx$ \$64ARG.

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 examination period. 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 the reminder, the e-mail was also shorten. 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, 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 Appendix

B.A 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 arms (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.

B.B 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 at least 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 in 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).

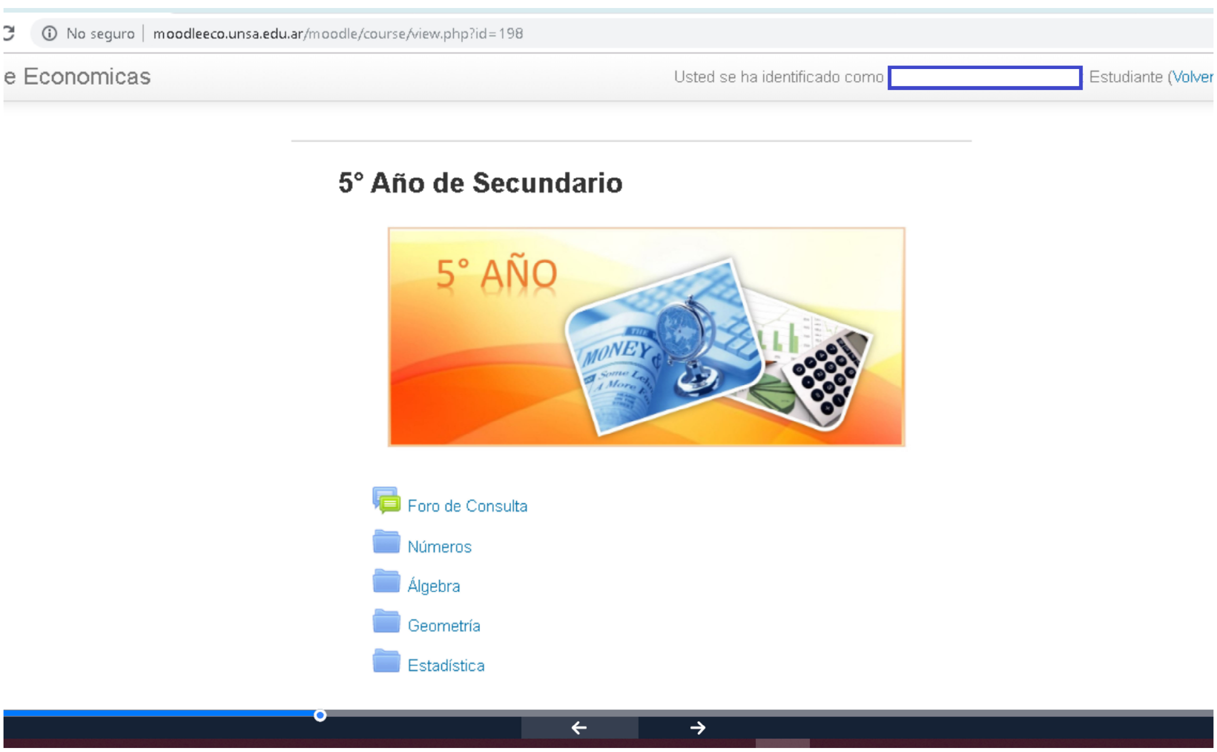
As mentioned above, the platform is not a part of the intervention, but rather enabled me to conduct the baseline surveys in all schools. After being introduced, we first explained the contents of the platform and then gave instructions on how to obtain free access (with a code that was determined for each school, for security reasons). Figure B1 shows the homepage of the platform, with all the content year by year. Figure B2 shows a representative image of the content available by topics for the senior year. Figure B3 shows pdf files with the available material.

We also showed how to post questions (public or private) with the commitment on our side to reply to each question within 48 hours. Students were allowed to upload pictures for assistance with exercises involving mathematical notation.

Figure B1: MOODLE Platform: Homepage

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

Figure B2: MOODLE Platform: Senior year overview



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

Figure B3: MOODLE Platform: Senior year specific content

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

B.C Full Derivatives: Model with Uncertainty

The maximization problem the student faces is:

$$\left[\hat{p}g\left(\frac{\beta e}{n}+\alpha\right)+(1-\hat{p})\,g\left(\frac{(1+\nu_1)\beta e+(n-1)\nu_2}{n+\nu_1}+\alpha\right)\right]\hat{V}-\delta e$$

with FOC:

$$\left[\hat{p}g'\left(\frac{\beta e}{n}+\alpha\right)\frac{\beta}{n}+(1-\hat{p})\,g'\left(\frac{(1+\nu_1)\beta e+(n-1)\nu_2}{n+\nu_1}+\alpha\right)\frac{1+\nu_1}{n+\nu_1}\beta\right]\hat{V}-\delta=0$$

Proof. Production Function

$$\hat{p}g'\left(\frac{\beta e}{n}+\alpha\right)+\hat{p}g''\left(\frac{\beta e}{n}+\alpha\right)\left(\frac{\beta}{n}\right)^2\frac{de^*}{d\hat{p}}-g'\left(\frac{(1+\nu_1)\beta e+(n-1)\nu_2}{n+\nu_1}+\alpha\right)\frac{1+\nu_1}{n+\nu_1}\beta+$$
$$+(1-\hat{p})\,g''\left(\frac{(1+\nu_1)\beta e+(n-1)\nu_2}{n+\nu_1}+\alpha\right)\left(\frac{1+\nu_1}{n+\nu_1}\beta\right)^2\frac{de^*}{d\hat{p}}=0$$

$$\frac{de^*}{d\hat{p}}=\frac{-g'\left(\frac{\beta e}{n}+\alpha\right)\frac{\beta}{n}+g'\left(\frac{(1+\nu_1)\beta e+(n-1)\nu_2}{n+\nu_1}+\alpha\right)\frac{1+\nu_1}{n+\nu_1}\beta}{\hat{p}g''\left(\frac{\beta e}{n}+\alpha\right)\left(\frac{\beta}{n}\right)^2+(1-\hat{p})\,g''\left(\frac{(1+\nu_1)\beta e+(n-1)\nu_2}{n+\nu_1}+\alpha\right)\left(\frac{1+\nu_1}{n+\nu_1}\beta\right)^2}\lesseqgtr 0$$

the second derivative of $g(.)$ is negative, but the sign of the numerator cannot be determined without additional assumptions on the $g(.)$ function and the parameters of relevance. Consider, for example, the case of Figure 1, with $g(x)=x-\omega x^2$:

- For underconfident students, it can be observed that $\frac{de^*}{d\hat{p}}>0$.
- In the case of overconfident students, we have $\frac{de^*}{d\hat{p}}<0$.

■

Proof. Returns to Education

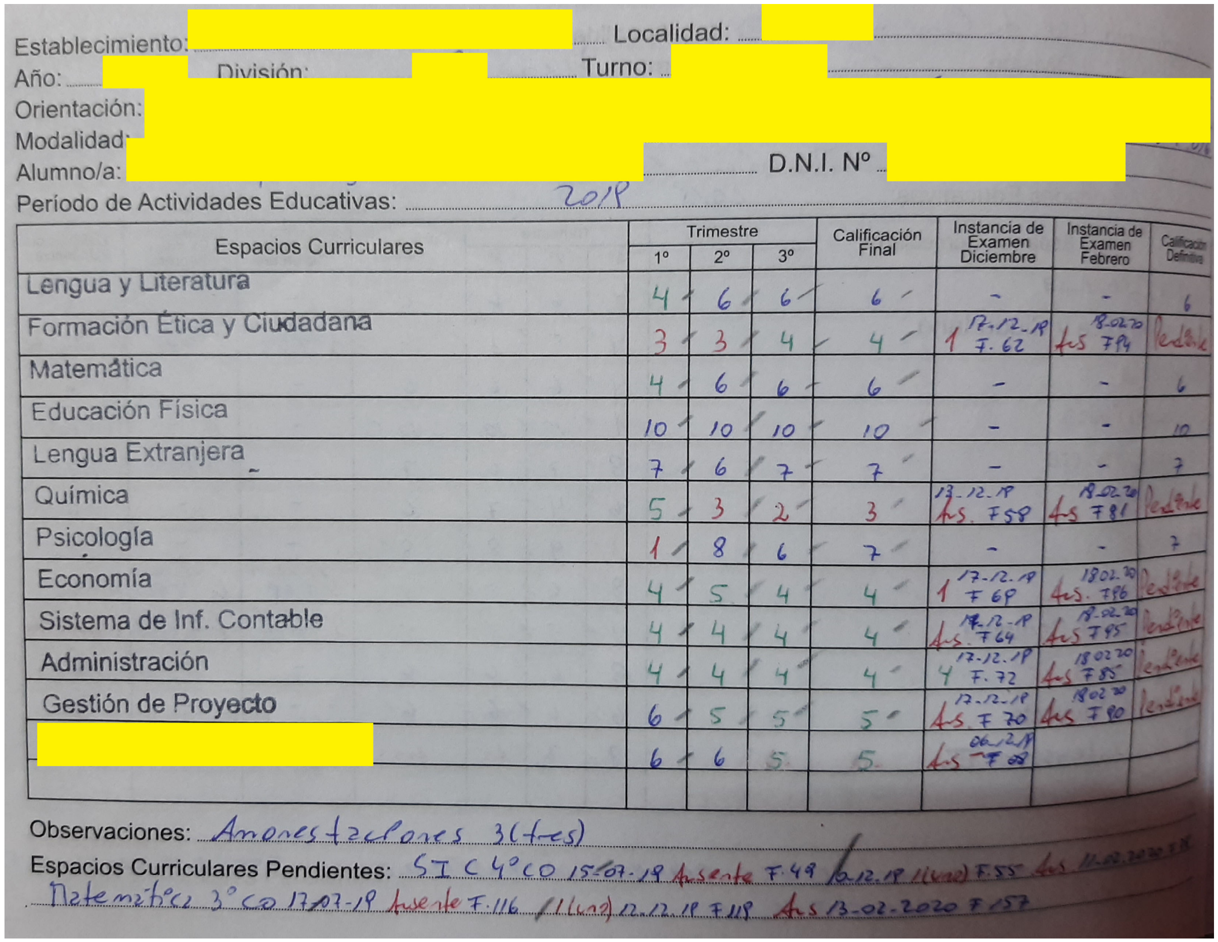
$$\hat{p}g'\left(\frac{\beta e}{n}+\alpha\right)\frac{\beta}{n}+(1-\hat{p})\,g'\left(\frac{(1+\nu_1)\beta e+(n-1)\nu_2}{n+\nu_1}+\alpha\right)\frac{1+\nu_1}{n+\nu_1}\beta+$$
$$\hat{p}g''\left(\frac{\beta e}{n}+\alpha\right)\left(\frac{\beta}{n}\right)^2\frac{de^*}{d\hat{V}}+(1-\hat{p})\,g''\left(\frac{(1+\nu_1)\beta e+(n-1)\nu_2}{n+\nu_1}+\alpha\right)\left(\frac{1+\nu_1}{n+\nu_1}\beta\right)^2\frac{de^*}{d\hat{V}}=0$$

$$\frac{de^*}{d\hat{V}} = -\frac{\hat{p}g'\left(\frac{\beta e}{n}+\alpha\right)\frac{\beta}{n} + (1-\hat{p})\,g'\left(\frac{(1+\nu_1)\beta e+(n-1)\nu_2}{n+\nu_1}+\alpha\right)\frac{1+\nu_1}{n+\nu_1}\beta}{\hat{p}g''\left(\frac{\beta e}{n}+\alpha\right)\left(\frac{\beta}{n}\right)^2 + (1-\hat{p})\,g''\left(\frac{(1+\nu_1)\beta e+(n-1)\nu_2}{n+\nu_1}+\alpha\right)\left(\frac{1+\nu_1}{n+\nu_1}\beta\right)^2}$$

By assumption, the second derivative of the $g(.)$ function is negative, so the entire denominator is negative, the numerator is positive (also by assumption), so the entire expression is positive. $\blacksquare$

C Appendix: Supplementary Figures and Tables

Figure C1: Student Academic Report. The format is similar in all secondary schools.

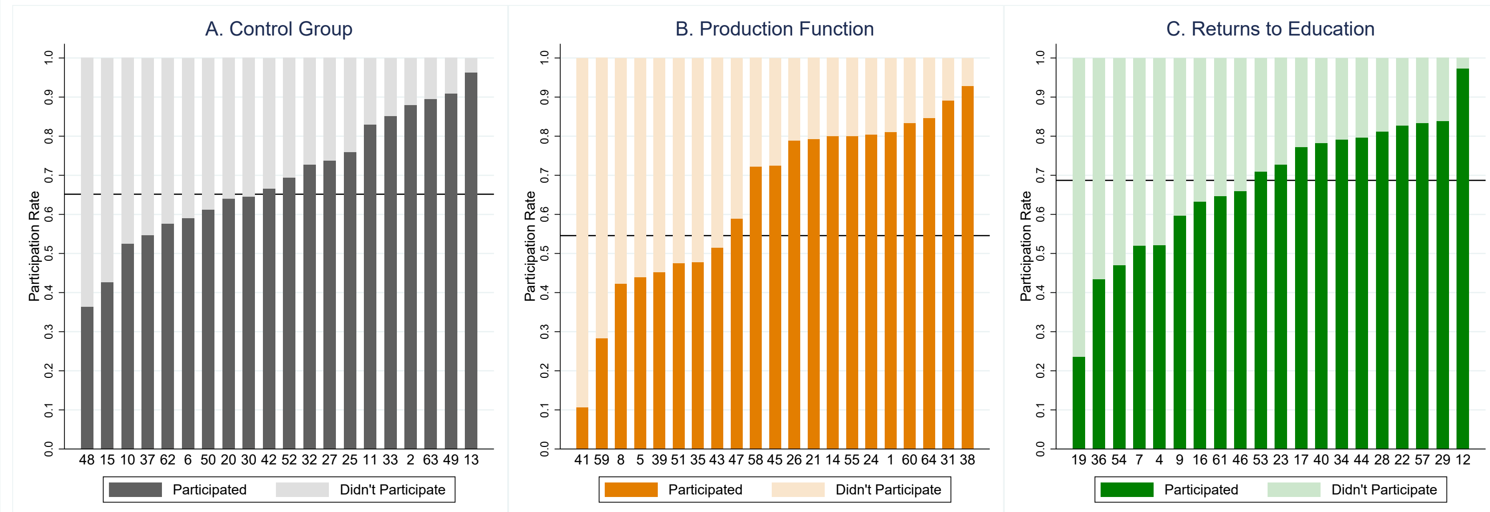


Establecimiento: [redacted] Localidad: [redacted]
Año: [redacted] División: [redacted] Turno: [redacted]
Orientación: [redacted]
Modalidad: [redacted]
Alumno/a: [redacted] D.N.I. Nº [redacted]
Período de Actividades Educativas: 2018

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. F96	Pendiente
Sistema de Inf. Contable	4	4	4	4	Aus. 17-12-18 F64	18-02-20 Aus F95	Pendiente
Administración	4	4	4	4	4 17-12-18 F. 72	18-02-20 Aus F85	Pendiente
Gestión de Proyecto	6	5	5	5	17-12-18 Aus. F 70	18-02-20 Aus F90	Pendiente
[redacted]	6	6	5	5	06-12-18 Aus F 68		

Observaciones: Amonestaciones 3 (tres)
Espacios Curriculares Pendientes: SIC 4º CO 15-07-19 Ausente F.49 10.12.19 1(uno) F.55 Aus 11.02.2020 F.8[illegible]
Matemática 3º CO 17-07-19 Ausente F.116 1(uno) 12.12.19 F.118 Aus 13-02-2020 F.157

Figure C2: Participation Rates at the School Level

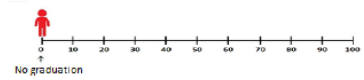


Notes: Horizontal axis shows random numbers assigned to each school. In each panel, the horizontal black lines indicates the participation rate for the entire treatment arm.

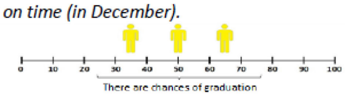
Figure C3: Question used to ask own probability of graduation

<u>Probability:</u> 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 degree? in December? after the exam dates of that month. 0 means no chance of receiving the title and 100 means that you will receive the title with certainty.

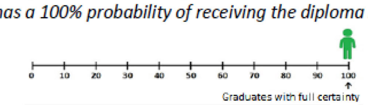
Example 1: A student who does not study, frequently skips classes, has pending subjects and does not appear at the exam periods, who disapproves of all the subjects this year, has a 0% probability of receiving the diploma in December.



Example 2: A student who studies sometimes, sometimes skips classes, with some pending subject, has a chance to receive the diploma on time (in December).



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



16- *What are your chances of receiving a high school diploma in December? (after exams period) From 0 to 100.* ________

Notes: First, a notion of probability was provided.

Table C1: Impacts of Information on Self-estimated Probability of Graduation (after-before intervention)

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

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 shift and strata fixed effects. See notes in Table 3 for a list of potential controls. *, **, and *** denote statistical significance at the 10, 5, and 1 percent levels respectively.

Table C2: Impacts of Information on Graduation by Pending Subjects

	(1)	(2)	(3)
		Graduation	
	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: R = PF	0.049**	0.012**	0.138
P-value: R = PF = 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 shift 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 students. Missing values are recoded to the sample mean and separately dummied out. These missing dummies are also used to construct pairwise interactions. *, **, and *** denote statistical significance at the 10, 5, and 1 percent levels respectively.