Perceptions, Information Interventions and High-School Graduation: Experimental Evidence from Argentina

Job Market Paper: Preliminary Version

Carolina Lopez*

This version: September 2021

Abstract

Low high-school graduation rates are a central challenge for the development of human capital in developing 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, in a separate arm, I conducted a returns-to-education information intervention. Baseline perceptions about own probability of graduation play an important role in students' academic performance. Providing information about the correct shape of the production function of the high school diploma 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 worst performing students at baseline, and those with a low perception of the probability of graduation (compared to an objective probability of graduation computed for each student), are those who respond most to the treatment. The returns to education arm increases graduation by 10 percentage points. The magnitude is higher to the production function arm, but both treatments provide similar or even higher impacts than previous literature by targeting different sources of misperceptions. I also find that both treatment arms increase the probability of university enrollment by 5 percentage points (more than 30 percent compared to the control group). In contexts where returns of education are not feasible to compute, the historical data available in each school becomes an important source to improve educational outcomes.

*Brown University, Department of Economics. Email: carolina_lopez@brown.edu. I am incredibly thankful to Pedro Dal Bó, Andrew Foster, Anja Sautmann, Bryce Steinberg, 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, 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. The experiment was approved by Brown University IRB 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 secure individuals a quality job and allow upward mobility within an economy (Cotton et al., 2020). This issue is becoming more urgent in the Global South where there are significant barriers to access, lower levels of education and significant inequality of opportunities. Even in poor countries where barriers to education are minimal, the educational achievement gap with rich countries is large (Glewwe and Muralidharan, 2016). In some contexts, secondary enrollment is high but graduation rates remain low. This is a barrier to sustained economic mobility given that a high school diploma is a minimum requirement to get a quality job or to obtain higher levels of education.

Considerable evidence shows that incentivizing academic achievement (outcomes) has no effect on performance, but incentives work better to improve educational performance when targeting specific tasks (inputs). Fryer (2016) and Fryer (2011), discuss that a potential explanation on why students fail to transform effort in academic achievement could be the lack of adequate knowledge of the education production function. In addition, empirical evidence in economics and psychology shows that individuals tend to overestimate the probability of important outcomes (Heger and Papageorge, 2018), leading to sub-optimal decisions. To test these two potential channels in an educational context, I designed an experiment that tests both, and their interaction, on high school graduation, by using a unique setting that allows me to identify the existence of different production functions and students' types.

I conducted a randomized controlled trial in high schools in the city of Salta, Argentina, to study if targeted and valuable information could improve 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, the intermediate steps to transform effort in educational achievement effectively, has an impact on the likelihood of graduation of students attending their senior year. In a separated treatment arm, I present information about the returns to education. In this setting, where the main economic constraints are not the main barrier to attend and finish high school, providing this information could have a different result concerning other contexts in which have been tested so far. Both treatments were introduced through a quick presentation in a single visit to each school. The average graduation rate of the control group is 50 percent, and I find that both information treatment arms have higher, positive, and significant impacts on the probability of timely graduation and university enrollment to previous studies in the literature.

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

In addition, given mixed evidence on the provision of economics returns to education in the literature, I provided that information in a separate treatment arm. Economic constraints (such as tuition, fees, and school transportation) are mentioned in several papers as the main barrier to obtain more education even after receiving this information (Bonilla-Mejía et al., 2019). For that reason, they fail to observe positive impacts. I use my unique setting to test whether those economic constraints are a feasible explanation or not. The information containing employment levels and wages by levels of education was shown to students using the same format used for the other treatment arm. In total, my experiment has 3 arms: *Control*, *Production function* and *Returns to education*.

For requirements made from school authorities to grant access to the school building, during all the school visits (one per school), a free online math practice platform was introduced to students to help them improve their performance in that subject. This platform was designed for this intervention using content sent by current math public school teachers from all the high school years. This allowed me to conduct students' surveys 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.

I find that both information treatment arms have an overall positive and significant impact on graduation rates. Specifically, the *Returns to education* treatment arm 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 than the effect found in Jensen (2010), the *Production function* shows results of a similar magnitude compared to Jensen (2010) for his subgroup of less poor students, but in my case, all my results apply to the entire sample of students. The students who respond most to both treatment arms are in 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 (subjective measure). I compared that measure with an 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 the worst-performing. I asked the question about chances of graduation after the presentation of the information treatments as an experimental outcome and I find that students give more accurate estimations of their self reported estimations of graduation after receiving information about graduation probability by type of student (*Production function* treatment arm). Importantly, I observe that those students classified at baseline as underconfident (concerning their chances of graduation) exert more effort under both treatment arms. These results indicate that a single but targeted intervention for different types of students could help in other settings to facilitate a detrimental cognitive bias.

This paper contributes to the existing literature that explores how information can affect educational choices. The literature has explored the provision of information on economics returns to education in contexts with low attendance rates (mainly due to economic constraints) and observed 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 not have an impact on college enrollment (Bonilla-Mejía et al., 2019). In addition, the economics literature that looks at low school achievement has focused mainly on economic constraints (tuition, other fees, clothes, books, etc.); interventions that reduce those costs increase attendance but do not necessarily increase achievement (Ganimian and Murnane, 2016). In addition, interventions with non-monetary incentives also fail to increase educational achievement (Fryer, 2016).

While most of the previous papers focused on providing information about economics returns to education or monetary or non-monetary incentives to motivate students to invest more in education, my first contribution is to provide a novel piece of information to improve students' decisions in a high stake setting. I study a suggested (Fryer (2016)) but a new source of low educational achievement: lack of knowledge of the educational production function among students. To fill this gap in the literature, my experiment aimed to provide information 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*) gives one's probability of receiving a diploma on time by standing at the beginning of senior year and information about how to remedy academic standing if it is necessary.

In a second contribution, I test if students ignore or discount new information because of biased beliefs about the information they already have to finish high school (DellaVigna, 2009). Psychologists have long known that people tend to overestimate their abilities. In particular, overconfidence in an educational context may lead students to study less if they are overconfident (Nowell and Alston, 2007). I show how this biased belief in own performance is detrimental to students' chances of graduation.

In a third contribution, I show how having a deep knowledge of the historical, social, and current institutional context in which the intervention is conducted helps to target and design how and which information is needed to deliver to improve an outcome with a quick and cheap intervention. In the first stage of the research process, I heavily relied on qualitative research methods, such as focus groups, open and semi-structured questionnaires aimed at provincial authorities, high schools 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 —confidence—, and how feasible it was to measure it in the available time to conduct the surveys.

This paper is relevant to inform policy strategies to increase the demand for high school diplomas among teenagers, especially the disadvantaged ones at risk of failing to complete high school on time.[1] Although access to the educational system is not restricted in many settings, youths' lack of information can make them decide to invest less than the optimum level 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 high school diploma as a signal). This paper provides evidence of a factor that could explain underinvestment in education: the lack of knowledge of how to efficiently transform inputs into outcomes by providing accurate information on how to improve chances of graduation depending on students' 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 RCT. In Section 3 I discuss the theoretical framework

[1] In almost all countries and international organizations such as UNICEF (Annual Report 2020, `https://www.unicef.org/media/100946/file/UNICEF%20Annual%20Report%202020.pdf`) there is a current discussion about how to recover from the consequences of the COVID-19 pandemic and the related closure of schools and the impacts on students achievement. The concern for low high school diploma achievement was already present before the pandemic in Argentina. UNICEF reported the low school achievement (UNICEF-ARGENTINA, 2017), a referent from the private sector highlight the difficulties to hire young people with the high school diploma `https://www.lanacion.com.ar/sociedad/crisis-educativa-por-que-toyota-no-consigue-200-jovenes-con-el-secundario-completo-para-\trabajar-en-nid05082021/` and civil associations along with the current National Director of High School level shown their concern related to low completion rates `https://www.lanacion.com.ar/sociedad/preocupacion-por-que-la-mitad-de-los-alumnos-no-termina-el-secundario-en-el-tiempo-\esperado-nid07082021/`.

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 the mechanisms behind those results. 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 in some of them transportation is free for students. As a result, most of the teenagers are enrolled in high school (91 percent, CEDLAS and World-Bank (2018)). However, high school graduation rates remain low in the entire country. Less than half of teenagers who are enrolled graduate from high school (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) do not obtain the high school diploma because they fail to pass all the required subjects.

A possible reason for this result is that those students are not interested in pursuing an education beyond high school. However, there is another consequence of not getting the high school diploma: their chances of obtaining a quality job are drastically lowered.

2.1 Educational System and Students' Academic Standing

In Argentina, the educational system is compulsory up to the end of secondary school. As a result, the share of youths in secondary school age attending secondary school is 91.2 percent and 74.7 percent are attending public schools (CEDLAS and World-Bank, 2018). Students drop out at different points during high school, and among specialists in the area there is a consensus that the main explanations are "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 there is another important explanation which has attracted less research attention (this issue is not even mentioned by the Director of Secondary Education at the national level): students attend until the last day of high school but do not obtain the high school diploma. This topic remained unexplored basically because there is no digitized data at the individual level that allows to make conclusions about the magnitude of this issue.

In Argentina, the academic year begins in March, classes finish by December, but the

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

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, a characteristic of the system, is that 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 2 pending subjects, if they have 3 or more they must repeat the year). Each student is fully aware of the number of pending subjects they have.[4] Throughout the rest of the paper, I use this concept to define students' type at the beginning of senior year: in good standing (zero pending subjects) and in bad standing (at least one pending subject). In all high schools there are 3 examination dates to pass these pending subjects each year (July, December, and February). During phone interviews, school administrators said that the main issue related to low completion rates are the pending subjects, they observe that students do not pass them or simply they just do not show up during the examination dates.

As a consequence of not getting the high school diploma the chances of obtaining a quality job for a young individual are drastically lowered.[5] In fact, most of the teenagers and young people who did not finish high school are part of those not in education, employment, or training (NEET). The NEET issue represents an increasing concern in all Latin American countries (Tornarolli, 2016), in Argentina, approximately 1 out of 5 individuals aged 15-24 are classified as NEET. Improving graduation rates in developing countries constitute a relevant topic for the design of effective educational policies.

[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 subject are highlighted and pending subjects from previous years have a dedicated space. During the academic year, these reports are send (via students) to the parents/guardian to be signed every quarter. Although signature forgery is possible, parents are aware of the dates in which they should receive the report. To verify the knowledge of the parents/guardian of high school senior students' academic status, in interviews collected previous to the design of the intervention, they reported to be fully aware of their children academic status, that they insisted them to improve their situation but "they are not able to enforce rules."

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

2.2 Educational Situation in Salta

The intervention was carried out in the city of Salta —the capital city of the Argentinian province bearing the same name. In this setting, education and transportation are free for all students enrolled in all formal school levels. In 2018, the province of Salta had the eighth-largest sub-national secondary school system in Argentina (among 24 provinces). The province's school system is one of Argentina's worst-performing (Ganimian, 2020): in 2017, only 28.7 percent of students attending the senior year had a 'satisfactory' level in math.

According to self-reported data from a students' anonymous national survey collected at the end of the 2017 academic year (Aprender-2017, 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 indicates that the chances of timely graduation for that cohort was 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.

3 Theoretical Framework

Does a student change her effort after being informed about her probability of graduation or the economic returns of graduation? How is this behavioral change affected by her number of pending subjects? I develop a simple effort-choice problem to answer these questions.

To obtain the diploma a high school senior student needs to pass all her senior year exams and her pending subjects (if any). The low graduation rates at the end of the academic year reflects lack of knowledge of the diploma production function. The misinformation could be about the translation of study time into effective effort on exams on senior year courses or the amount of effective effort needed to pass pending subjects. In this case, student's utility (under certainty) will be:

$$u_i = g(e(b_i - a_i n))V - c_i(e)$$

Where i is a student type (under- or overconfident with respect to her chances of graduation at the beginning of the senior year (see Definition 1), $g(.)$ is the probability of graduation *(concave),* e is effort, b_i is the return to effective effort to pass subjects during the senior year, n is the number of pending subjects (endogenous variable), a_i is a penalty on effort for having pending subjects, V is the perceived return to the high school diploma, and $c_i(.)$ the cost function *(convex)* that depends on the student type. For simplicity, I assume that $c(e) = e$.

I assume that there are 2 potential states of the world for the student. In the first state the student perceives a high return to effort and a low penalty for having pending subjects; in the second one, the student perceives that there is a lower return to effort and a high penalty for having pending subjects. p is the true probability of the state of the world 1 and $(1-p)$ of the state 2. Let $\hat{p}$ be the probability assessed by the student. The student optimizes her utility considering the *expected* state of the world: $E(\tilde{g}) = [\hat{p}g(b_{i1}e - a_{i1}ne) + (1-\hat{p})g(b_{i2}e - a_{i2}ne)]$. Correspondingly, let $e^*(\hat{p})$ be the student's optimal effort given a perception of $\hat{p}$.

Let $g_1(e) = g(b_{i1}e - a_{i1}ne)$ and $g_2(e) = g(b_{i2}e - a_{i2}n)$.

Definition 1 *The student is over-confident if and only if her subjective probability of graduation is higher than her objective probability. This is,*

$$\hat{p}g_1(e^*(\hat{p})) + (1-\hat{p})g_2(e^*(\hat{p})) \quad > \quad g(e^*(p))$$

With $b_{i1} > b_{i2}$ and $a_{i1} < a_{i2}$, if [6]

$$g_1(e^*(\hat{p})) > g_2(e^*(\hat{p}))$$

and $\hat{p} > 0$, then the student is overconfident.

Assumptions

- Individuals have perfect information about their own standing at the moment of the interventions (how many pending subjects they have).
- Providing information about the state of the world consists of moving the beliefs of the students toward the true state they are in. This affects their perceived probability of graduation given effort and academic type.

The student chooses $e = e^*$ given their baseline perceptions. Then, the student should solve:

$$\underset{e}{max}\{[\hat{p}g(b_1e - a_1ne) + (1-\hat{p})g(b_2e - a_2ne)]V - e\}$$

This problem has a unique solution given by $e^* = e(\hat{p}, V)$. To obtain main prediction of this model I apply the implicit theorem function.

[6] From now on, I will omit in the equations the type of student (i) to reduce notation.

Role of the Treatment Arms

There are comparative statics of special interest to make theoretical predictions for each treatment arm. First, the *Production function* treatment arm was designed to change the subjective probabilities of graduation, it would be expected to observe a change in effort if the student update their perceptions. By using the FOC:

$$g'\left(b_1e^*-a_1ne^*\right)\left(b_1-a_1n\right)+\hat{p}g''\left(b_1e^*-a_1n\right)\left(b_1-a_1n\right)^2\tfrac{de^*}{d\hat{p}}-g'\left(b_2e^*-a_2ne^*\right)\left(b_2-a_1n\right)+\left(1-\hat{p}\right)g''\left(b_2e^*-a_2n\right)\left(b_2-a_2n\right)^2\tfrac{de^*}{d\hat{p}}=0$$

$$\left[\hat{p}g''\left(b_1e^*-a_1ne^*\right)\left(b_1-a_1n\right)^2+\left(1-\hat{p}\right)g''\left(b_2e^*-a_2ne^*\right)\left(b_2-a_2n\right)^2\right]\tfrac{de^*}{d\hat{p}}=\left[-g'\left(b_1e^*-a_1ne^*\right)\left(b_1-a_1n\right)+g'\left(b_2e^*-a_2ne^*\right)\left(b_2-a_2n\right)\right]$$

$$\frac{de^*}{d\hat{p}}=\frac{\left[-g'\left(b_1e^*-a_1ne^*\right)\left(b_1-a_1n\right)+g'\left(b_2e^*-a_2ne^*\right)\left(b_2-a_2n\right)\right]}{\left[\hat{p}g''\left(b_1e^*-a_1ne^*\right)\left(b_1-a_1n\right)^2+\left(1-\hat{p}\right)g''\left(b_2e^*-a_2ne^*\right)\left(b_2-a_2n\right)^2\right]}\tag{1}$$

Notice that the sign of this derivative cannot be determined without an explicit function with values for b_i and a_i in both states of the world.

The *Returns to education* arm was conducted to analyze the impact of labor market statistics by level of education on students effort, so the relevant comparative static in this case is with respect to V:

$$\left[\hat{p}g'\left(b_1e^*-a_1ne^*\right)\left(b_1-a_1n\right)+\left(1-\hat{p}\right)g'\left(b_2e^*-a_2ne^*\right)\left(b_2-a_2n\right)\right]+\left[\hat{p}g''\left(b_1e^*-a_1ne^*\right)\left(b_1-a_1n\right)^2\tfrac{de^*}{dV}+\left(1-\hat{p}\right)g''\left(b_2e^*-a_2ne^*\right)\left(b_2-a_2n\right)^2\tfrac{de^*}{dV}\right]V=0$$

$$\left[\hat{p}g''\left(b_1e^*-a_1ne^*\right)\left(b_1-a_1n\right)^2+\left(1-\hat{p}\right)g''\left(b_2e^*-a_2ne^*\right)\left(b_2-a_2n\right)^2\right]V\tfrac{de^*}{dV}=-\left[\hat{p}g'\left(b_1e^*-a_1ne^*\right)\left(b_1-a_1n\right)+\left(1-\hat{p}\right)g'\left(b_2e^*-a_2ne^*\right)\left(b_2-a_1n\right)\right]$$

$$\frac{de^*}{dV}=\frac{-\left[\hat{p}g'\left(b_1e^*-a_1ne^*\right)\left(b_1-a_1n\right)+\left(1-\hat{p}\right)g'\left(b_2e^*-a_2ne^*\right)\left(b_2-a_2n\right)\right]}{\left[\hat{p}g''\left(b_1e^*-a_1ne^*\right)\left(b_1-a_1n\right)^2+\left(1-\hat{p}\right)g''\left(b_2e^*-a_2ne^*\right)\left(b_2-a_2n\right)^2\right]V}>0\tag{2}$$

3.1 Predictions of the Model

I show how the provision of information could have an impact on beliefs and probabilities of graduation. The *Production function* arm can change perceptions ($\hat{p}$) 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 is going to vary depending 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 n is endogenous, meaning that students have some knowledge about the cost of effort they have to implement to pass subjects. I assume that the perceived cost of effort is negatively correlated with the academic standing of students (notice that my running variable is effort, not productivity or 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.

- Effects of the Production Function Treatment Arm

  - **Prediction 1** *Students with* $n = 0$ *(zero pending subjects) are less likely to update their beliefs, because the differences between states will be lower (*$a_i n = 0$*) and as a consequence their effort under the* Production function *treatment arm (see details in Subsection 4.5).*

    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 a higher weight was put on remedying the academic standing of students with pending subjects. In Figure 1, Panel A, it can be observed that the difference between the two states of the world is not extremely different, so the provision of information is not very useful to them.[7]

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

    That treatment arm 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. How they will adjust the weight in each potential state of the world? n depends on past effort and some knowledge about the return to effort (correct or incorrect value of a_i). Under different cost structures, different results can be observed. To explain this, in Figure 1 Panel B, the difference in marginal benefits will be more different according to the parameters in a good o bad state of the world. First, notice that the dashed grey line (expected self-perception of graduation) is going to 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 $\hat{p}$, this means an overconfident student ($\hat{p} = 1$) and an underconfident student ($\hat{p} = 0$) before my intervention.

[7]Notice that there is another solution if $b_1 \sim b_2$ (but still $b_1 > b_2$) $\Rightarrow \frac{de^*}{d\hat{p}} \approx 0$.

* **Prediction 3** *Underconfident students will increase their effort after receiving the Production Function treatment arm.*

  According to my definition of confidence, the true state of the world for an underconfident student is higher than her perception. This student thinks she is in a bad state, when in fact is in a good one. In the graph, for this type of students the marginal benefit of graduation is on the black line, before the intersection of curves. At baseline their equilibrium will be at e_{hc}^{u*}, where hc indicates that they perceived higher costs, and $u*$ indicates the baseline equilibrium. With the intervention they will learn that they are in 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 will depend on the magnitude they update $\hat{p}$.

* **Prediction 4** *Onderconfident students will increase their effort after receiving the Production Function treatment arm.*

  According to my formal definition of confidence, they think that $\hat{p} = 1$ (they perceive they are in the good state of the world). Following the previous explanation, this type of students think they are in the right hand side of the grey line (after the intersection of states). Notice that a potential explanation of why these students believe are high types could be because they perceive a low cost of effort, so the initial equilibrium is e_{lc}^{o*}, where lc indicates that they perceived lower 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, equilibrium will move to the right, meaning that their effort is going to increase under this scenario too.

- Effects of the Returns to Education Treatment Arm

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

    Under this treatment arm, previous literature, Jensen (2010) and Nguyen (2008), indicates that students with low perceived returns to education should update the 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 at baseline perceived high returns to education should decrease their effort.

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

Mechanisms

- *Production function:* Students update their beliefs about the right state of the world they are, and correct the level of effort to obtain the high school diploma.

- *Returns to education:* Students receive truthful information, update their priors on perceived returns to education, which motivates students to achieve the diploma.

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 2. 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 also 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 of pending subjects for those students is 1.6 subjects.

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 informing in advance which information treatment arm was randomly assigned to each school.

4.2 Sample

The eligible population for this study is students attending their senior year at public high schools in Salta.[8] While some schools can have more than one shift, I only considered the morning and the 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 school/shift levels given that randomization at the individual or class level would be more likely to contaminate the control group.

4.3 Timeline

At the beginning of this project, 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. Besides 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, trips.[9]

This process finished in the first quarter of 2019 (see Figure 2). At the same time I requested from the Directorate access to 5 "representative" schools to collect individual data about school performance and graduation. This administrative data was not available, so I followed their recommendation to collect data which was stored in secured rooms at each school building. The main intention was to compute statistics at the individual level that were used for the production function treatment arm and confirm that in fact graduation rate is approximately 50 percent, in large part due to the pending subjects issue (see more details in Appendix A).

In 2 out of those 5 schools I tested the survey instruments on groups of 11th graders to consider time required and to change questions if it was necessary to facilitate the understanding of students. Several edits were made to the survey instruments at this point; revision was crucial as school principals allotted just one hour at each school to avoid disruptions to usual curricula. The day of the visit was coordinated with the vice school principal. 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. 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 outbreak

[8] From now on, I will use Salta to make reference to the capital city and not the province.

[9] Each school has one principal and if the school has more that one shift there is a vice school principal per each shift. From now on, I will use the term "school" to make reference to "school/shift".

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**.
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 of extremely importance, given that the issue with the pending subjects and connections with 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-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 COVID outbreak, and delayed the entire process to analyze results. In the meantime, I started negotiations with the main post secondary institutions in Salta, to obtain information about universities enrollment.

4.4 Data

Baseline Survey

To get access to all schools to collect baseline questionnaire data and to implement the interventions, the research team visited all the schools in the sample to show how to get access to a free online platform with math contents (designed by professors at Universidad Nacional de Salta - UNSa). This works as a "placebo" for those schools in the control group.

Before that presentation, 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 the consent forms for parents of students attending the senior year. At a date and time agreed with the school administrators, all students of a school/shift were reunited in a room.[10] A description of the activities conducted during each visit day is shown in Figure 3.

The research team was first introduced by school administrators. Then, tablets were delivered to students and they were asked to fill out the questionnaire, a short presentation (containing slides with pictures) was shown to instruct students on the use of the tablets. At the same time, a quick explanation of the questionnaire was provided.[11] After that, 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 perceptions of own graduation to test for any changes in those 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, the information interventions were shown (if applicable). Given that a single presentation, including statistics, and unknown facts for the students, could not have been enough to change students effort, before the examination dates (to pass pending subjects and failed subjects during the senior year) I have sent SMS and/or email two weeks before each date to remind the treatment received in short sentences.[12] As was shown in previous papers, reminders could 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 (from February 2020 to March 2021). Those individual records contained data on performance during the entire school year and graduation. Also contains information about the pending subjects (if any) and attendance to examination dates. An example of an individual record is in Figure C1, Appendix C.

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

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

[12]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 assignment to treatments was randomly determined at the school/shift level stratifying by number of students and geographic area of the capital city of Salta. All the schools' visits included the presentation of a free online platform with math contents (see Appendix B) so their 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 graduation rate on time (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 completion rate on time for this subsample was 50 percent. Having pending subjects are not necessarily the main cause of failure to obtain a diploma —students can fail in passing 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 the academic standing were provided to all students (because at the moment 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[13], asking for study material from classmates or students from younger cohorts (given that the teachers employed by the schools and the required academic material can change over time), talking with teachers in advance to ask them for studying recommendations or asking which teachers will be a part of the committee in each subject. [14]

Returns to Education: Students might not be aware of the disadvantages of not finishing high school and its impacts on their labor market prospects. The provision of this information should incentivize students to obtain the 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 (2nd 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, besides 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 2 shifts (out of 64 school-shifts) 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 **?**).

Notice the difference in students' participation among the intervention treatment arms. A higher percentage of students, their parents, or both, decided not to participate in the *Production function* treatment arm. This selection into participation could have had detrimental impacts on the analysis of this treatment arm, but the protocol of the visits to the schools allow me to discard selection in participation. No school authorities knew before hand which treatment was assigned to them. The research team only knew about the arm that should be implemented 30 minutes before the arrival to each school.

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

[14]Usually, the committee for each subject/year is formed by 3 to 5 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 some dispute arises.

4.6 Measuring Students' Confidence in Graduation

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

Given the difference in graduation I observe at baseline for those student 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 2 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 in the household there is a computer, a washing machine, an AC, 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-shift level, along with shift and strata fixed effects.

Figure 5 shows in Panel A the distribution of the estimated probabilities for those 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 those students with at least one pending subject. According to my definition of confidence, those 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 3 shows 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 consented their participation). Columns 2 and 3 present coefficients from the following regression specification:

$$y_{is} = \beta_0 + \beta_{PF} ProductionFunction_s + \beta_{RE} ReturnsEducation_s + \delta_s + \epsilon_{is} \qquad (3)$$

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. In addition, and 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 both information treatments is of special interest.

Table 3 Panel A shows that the average number of students that participate in each school-shift visit is almost 31 and there are no significant differences among treatment arms. Panel B, shows students characteristics. On average they are 18 years old. Sixty percent of participants are female, 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 states that they have some internet access (via their household, cellphones, school, or public places). On average, 1.74 persons per room live in students' households. 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. Twenty percent of them state that they take care of a family member. There are no statistically significant differences in these measures across 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 of them had at least one pending subject at the moment of the visit.

Panel E shows 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 moment of the school visit, students perceived that their chances of on-time graduation at 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:

$$y_{is} = \beta_0 + \beta_{PF} ProductionFunction_s + \beta_{RE} ReturnsEducation_s + \delta_s + x'_{is}\omega + \eta_{is} \qquad (4)$$

this equation is the same as equation (3) 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 4, column 1, shows that graduation for all students who were selected to participate in both treatments arms 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 like to obtain the diploma (20 percent with respect to the control group). I find that the difference among these treatments is statistically significant.

The effect of *Returns to education* is double than the effect found in Jensen (2010) for the subgroup of less poor students (the author does not find an impact for poor students). A potential explanation for this higher impact could be related to the fact that in this paper, the target population were students who were closer to receive the high school diploma. Additionally, in my setting, there are fewer economic barriers: enrollment and transportation to school are free. The *Production function* effects are the same in magnitude as in Jensen (2010) but in my case they apply to the entire sample. This shows that this 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 be impacted in the same way by the *Production function* treatment arm. In Table 4, column 2 and 3 show the treatment effects by academic standing separating the group into those students in good standing (zero pending subjects) and those in bad standing (at least one pending subject). As expected, I observe there is no significant effect on those students in good standing and the magnitude is close to zero. This is *Prediction 1.* A likely reason is that they 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 the 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, it can be observed that both groups have positive and statistically significant effects (*Prediction 5*).

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 in graduation on actual graduation (Table 5) by using the answers of the questions about chances of graduation before and after the interventions. An important part of the *Production function* treatment arm was to make students aware of the correct shape of the production function of the high school diploma given academic standing at the beginning of the senior year. As previously mentioned, at the moment of intervention the standing of the students was their private information and the goal was to allow students to create a *mapping* of their situation to graduation rates of similar students from the previous year.[15]

The question about the self perception of timely graduation was included in the baseline survey, conducted before the intervention treatments, and then the question was 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 and 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 get more accurate with respect to their own chances of graduation: the variable decreases by 2 percentage points from the baseline response.[16] 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 their relationship with my definition of confidence in Table 5. Although there are differences between the underconfident and overconfident students (for both treatment arms), the largest differences in magnitude are observed in the *Production function* arm for those in worst standing. *Prediction 3* and

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

[16]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 of the US and members from Universidad Nacional de Salta (UNSa) could have per se generated a optimistic response among students, given that almost there is not formal connection between secondary and post secondary levels.

Prediction 4 of the theoretical framework are verified. There are positive and significant effects (at the 5 percent level of statically significance) for both under- and overconfident students, with a difference of 20 percentage points (but non significant) in favor of those underconfident students.

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

Effort

Table 6 measures the role of effort devoted to the study time of pending subjects. I analyze the effect of the information treatment arms on 3 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 one indicates a degree of effort because according to high school rules, only those who explicitly register for the examination dates with a secretary are allowed to take the exam.[17] The second one, indicates if in fact the students attend to the examination date and the third one shows a dummy variable 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 6 Panel A shows positive impacts of the information treatments on these outcomes, specifically, for those who received the *Returns to education* treatment the effect is statistically significant at the 1 percent level of confidence for column 2 and 3.

Panel B shows the effect of the information treatment arms by confidence. As discussed above, those underconfident students are those who respond more to the treatment by exerting more effort, the difference in the *Production function* between both type of students is more than 40 percentage points (difference significant at the 1 percent level of confidence). In the *Returns to education* treatment arm there are also differences in favor of the underconfident students but they are lower and in only column 2 the difference with respect to the overconfident students is significant at the 10 percent level of confidence.

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 can have a positive misperception (meaning they overestimate the returns to education, relative to the true values) or a negative one

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

(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 a 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 arm. 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 7 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 and positive way at baseline have positive magnitudes, confirming in this way *Prediction 5* but not *Prediction 6*. The magnitude of the effects is higher for those with 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, they could decide to which piece of information assign a higher weight. According to the previous results in the literature is expected that those with negative misperception, will update their beliefs up and graduation will increase. But the aggregated result depends on which percentage of students assign a higher weight to their prior beliefs or to the new information. This is a potential explanation of why I cannot verify *Prediction 6* in my results.

Time Preferences

The *Returns to education* treatment arm 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, and then I took the median and then I separated students by those above and below median. Results are shown in Table 8. As expected, the effect for the *Returns to education* treatment arm is greater and also statistically significant for students above the median. Although the difference with respect to those under the median value is not statistically significant, it shows that this is a relevant individual characteristic to consider when information like this are provided to teenagers.

Also, it can be observed that the magnitudes for both groups of students that received the *Production function* are lower, similar in magnitude, and non significant, and 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 survey pilots. In order to generate a proxy for economic status, I use an index constructed by using variables indicating the ownership of AC, heating, washing machine, a desktop or laptop, whether the student's family lives in an overcrowded dwelling[18] and if at least one parent or guardian has some post secondary education. If the index is less or equal than 3, I classified that student as "Poor" and "Least poor" otherwise.[19]

Table 9 shows that in the control group, students classified as poor have a lower graduation rate, 45 percent, 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 treatment arms: those 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 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 9 also shows the impacts by gender. Column 3 and 4, show that female students are more likely to graduate than males in the control group. However, both information treatment arms have a positive impact on both genders, with higher impacts observed for men. Although I observe positive for both treatments arms for both genders, the differences are not statistically significant.

5.5 Other outcomes

One of the objectives of this paper was to analyze the effects of information treatments beyond secondary school. Given some data limitations explained below I only consider if the

[18]This variable indicates that in average students live in a household with less than 2 persons per room.
[19]For the control group, the median value of this variable is 3 and the mean 3.12.

student is enrolled in a university in the academic year after my interventions were conducted (2020) or enters in 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.

Something important to highlight is fact 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 do not impact his or her selected degree. It is important to stress that the only requirement is the high school diploma, although students with pending subjects can enroll in a provisional way. It was not possible to obtain information on other tertiary educational centers, so my measure only includes universities.

Another fact to mention, is that it is not very likely that students from Salta (attending a public high school) move to another province to attend college because even if they were to attend a public university, they would have to consider the cost of moving and housing, which are very expensive compared to the tuition at UCASAL. There are not available data at the national level that allow me to test the percentage of students who move to another province to study in the post secondary level. Given these facts, my results will represent a lower bound of the effect of the information treatments on a superior level of education.

Table 10 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 in 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 huge 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 participant students (and parents, if the student was minor) gave me consent to check their employment status.

The system only allows to access to information from the 6 previous months at the moment of the inquiry.[20] Given the strict quarantine that was imposed by the government in Argentina to respond the COVID-19 outbreak, 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 formal employee at least one month out of those 6 months.

Column 2 of Table 10 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 in the considered time. But it can be observed that both treatment arms generate a negative and statistically significant impact on formal employment. One key caveat is that the sample size in this analysis is lower than the original due to the fact that 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. A potential, but not conclusive explanation, is that students' reservation wage increased after receiving the treatments.

6 Conclusions

This paper analyzes the effect of information interventions to improve high school graduation by correcting mistaken perceptions of students on different aspects by using a novel intervention and a traditional one. The first one, and the main contribution of this paper, is aimed to make students aware of their chances of graduation by academic standing at the beginning of the senior year and teaches how to effectively transform inputs into outputs (*Production function*). The second one shows information about the returns to education by 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 noticed in previous papers, overconfidence could be a detrimental personality trait in an educational setting. Overconfidence in graduation is widespread in my sample, and the most significant

[20]See Subsection 4.4.

impacts seem to be driven to those classified as underconfident students.

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 there is high economic instability. I observed positive and significant effects for both treatment arms on timely graduation, and the magnitudes are more important than those found in other studies. Also, I found positive and significant impacts on college enrollment, while previous studies aimed to drive demand for post-secondary education did not find this effect.

The findings of this study are of substantive policy importance: graduation rates can be improved in low-income settings by using a cheap intervention, collecting data from a previous academic year at a low cost. In contexts where there is no feasible way to compute returns to education, using the available information is a great advantage to improve graduation rates, as proved in this paper. Students who are positively affected by this intervention now have a previously unavailable chance to achieve economic mobility.

References

Aprender-2017 (2017). *Aprender 2017: Informe de resultados, Salta, 5to a˜o de secundaria.*

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

Bonilla-Mejía, L., N. L. Bottan, and A. Ham (2019, dec). Information policies and higher education choices experimental evidence from colombia. *Journal of Behavioral and Experimental Economics 83*, 101468.

Bruhn, M. and D. McKenzie (2009, sep). 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, jun). 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/.

Cotton, C., B. Hickman, J. List, J. Price, and S. Roy (2020, oct). 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, jun). Nudging in education. *Economics of Education Review 64*, 313–342.

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

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

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

Ganimian, A. J. (2020, jul). 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, jul). 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, aug). 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.

Loyalka, P., C. Liu, Y. Song, H. Yi, X. Huang, J. Wei, L. Zhang, Y. Shi, J. Chu, and S. Rozelle (2013, nov). 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.

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, apr). I thought i got an a! overconfidence across the economics curriculum. *The Journal of Economic Education 38*(2), 131–142.

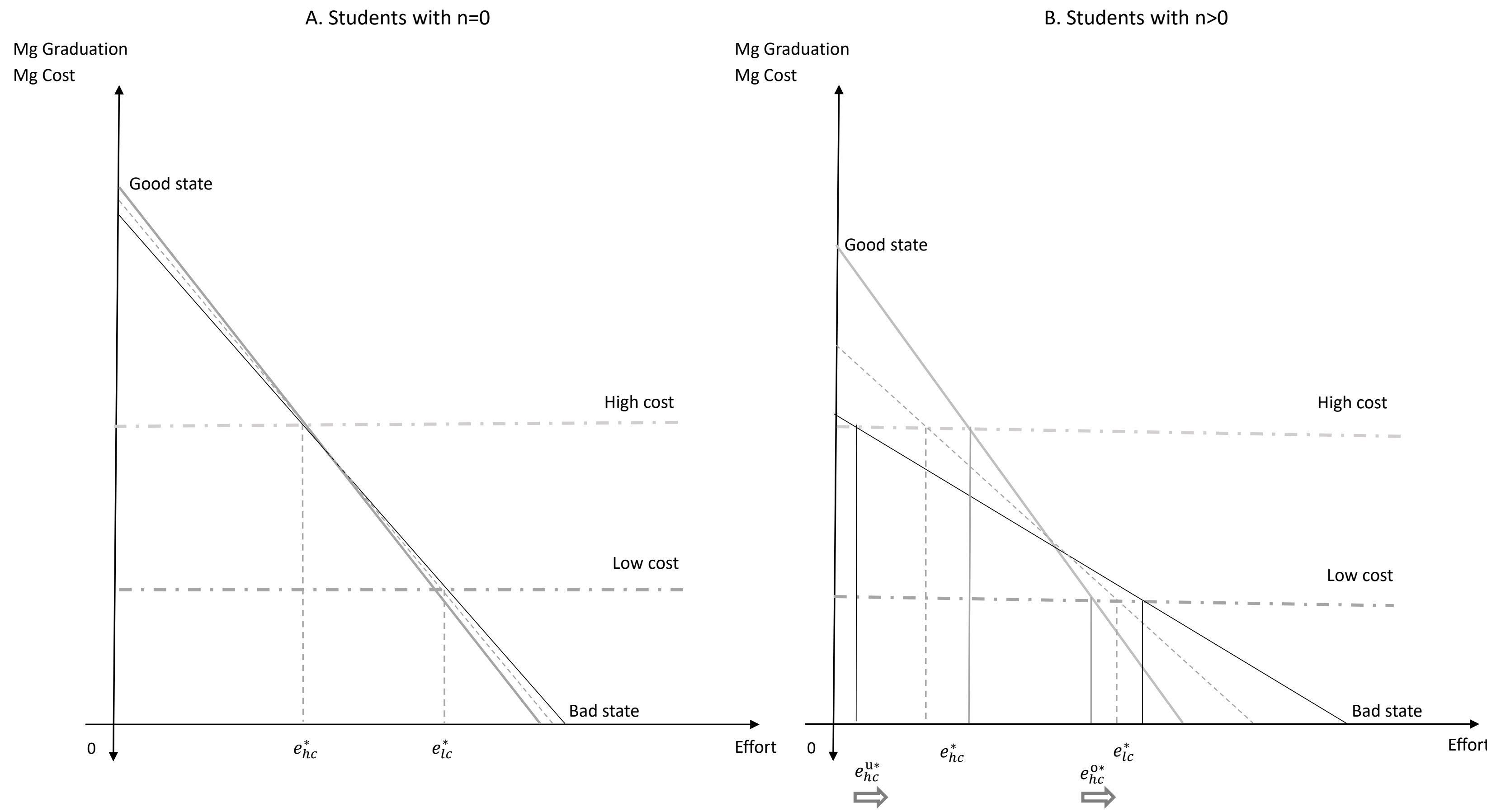
Spence, M. (1973, aug). Job market signaling. *The Quarterly Journal of Economics 87*(3), 355.

Tornarolli, L. (2016, dec). 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: The number of pending subjects is endogenous, the treatment only changes the belief in which state of the world a student thinks she is. In Panel A, students with zero pending subjects, it can be observed that a change in $\hat{p}$ will marginally move the equilibria under the two different cost structures. In Panel B, students with pending subjects, the marginal benefits of graduation are more salient in each state ($a_i n > 0$). It can be observed that the equilibria (dashed grey lines) will change according to the update in $\hat{p}$.

Figure 2: Timeline, Intervention and Data Collection

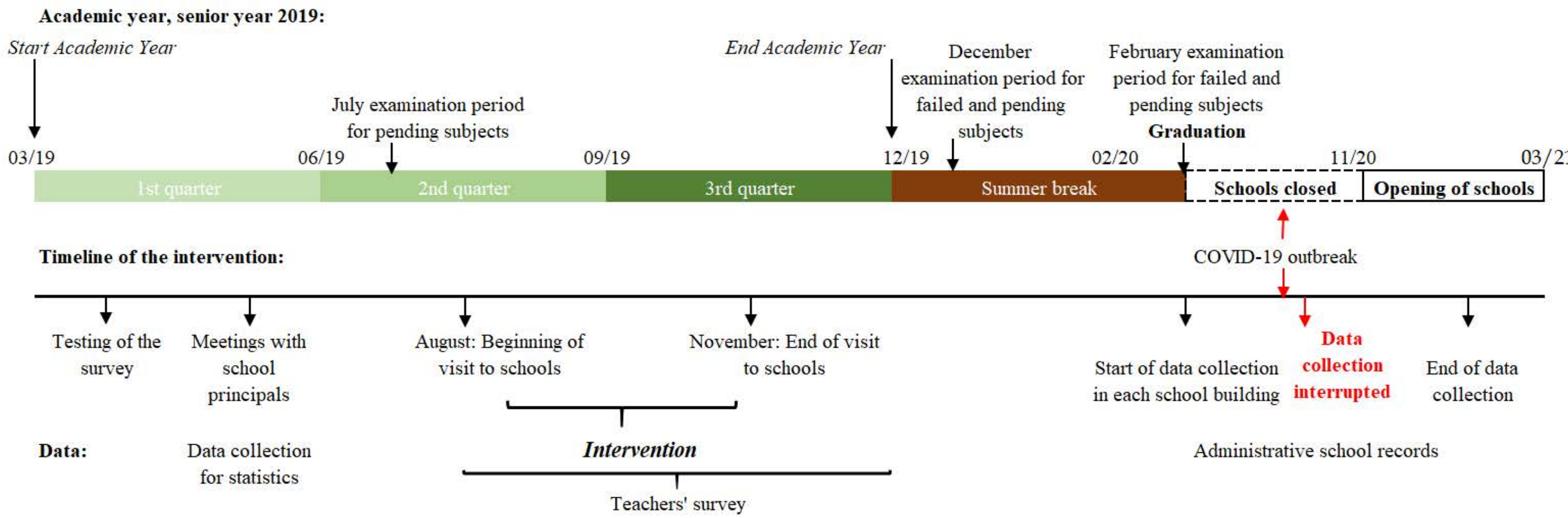


Figure 3: The Intervention Day

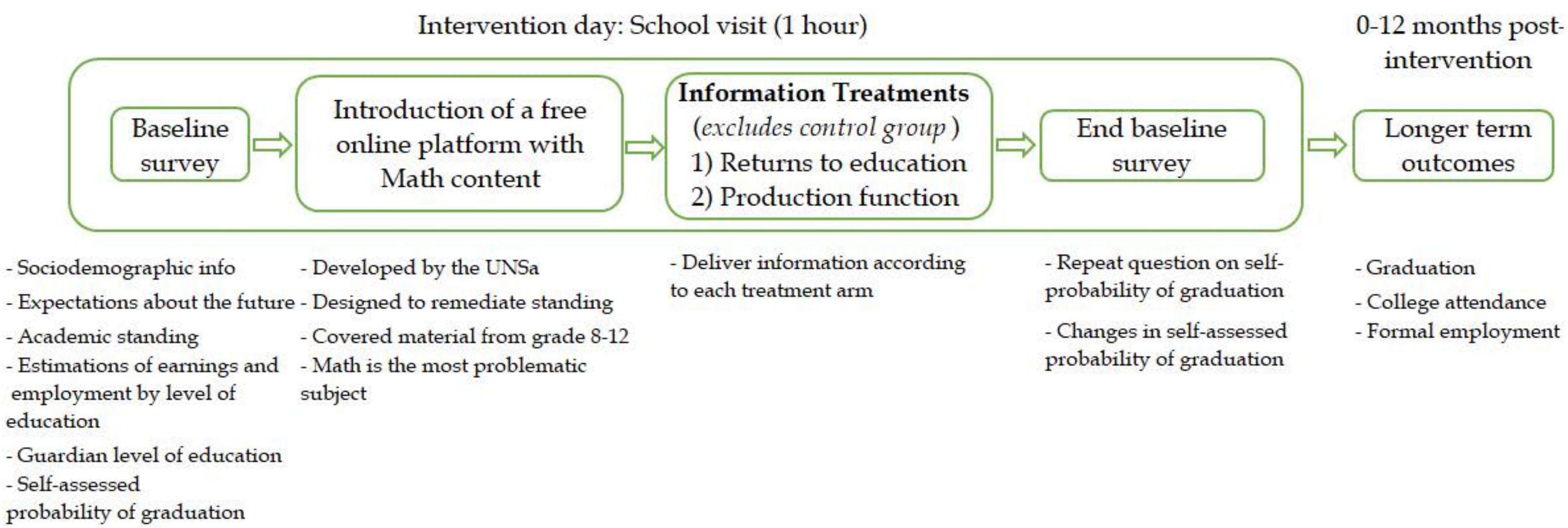


Figure 4: Randomization Design and Sample

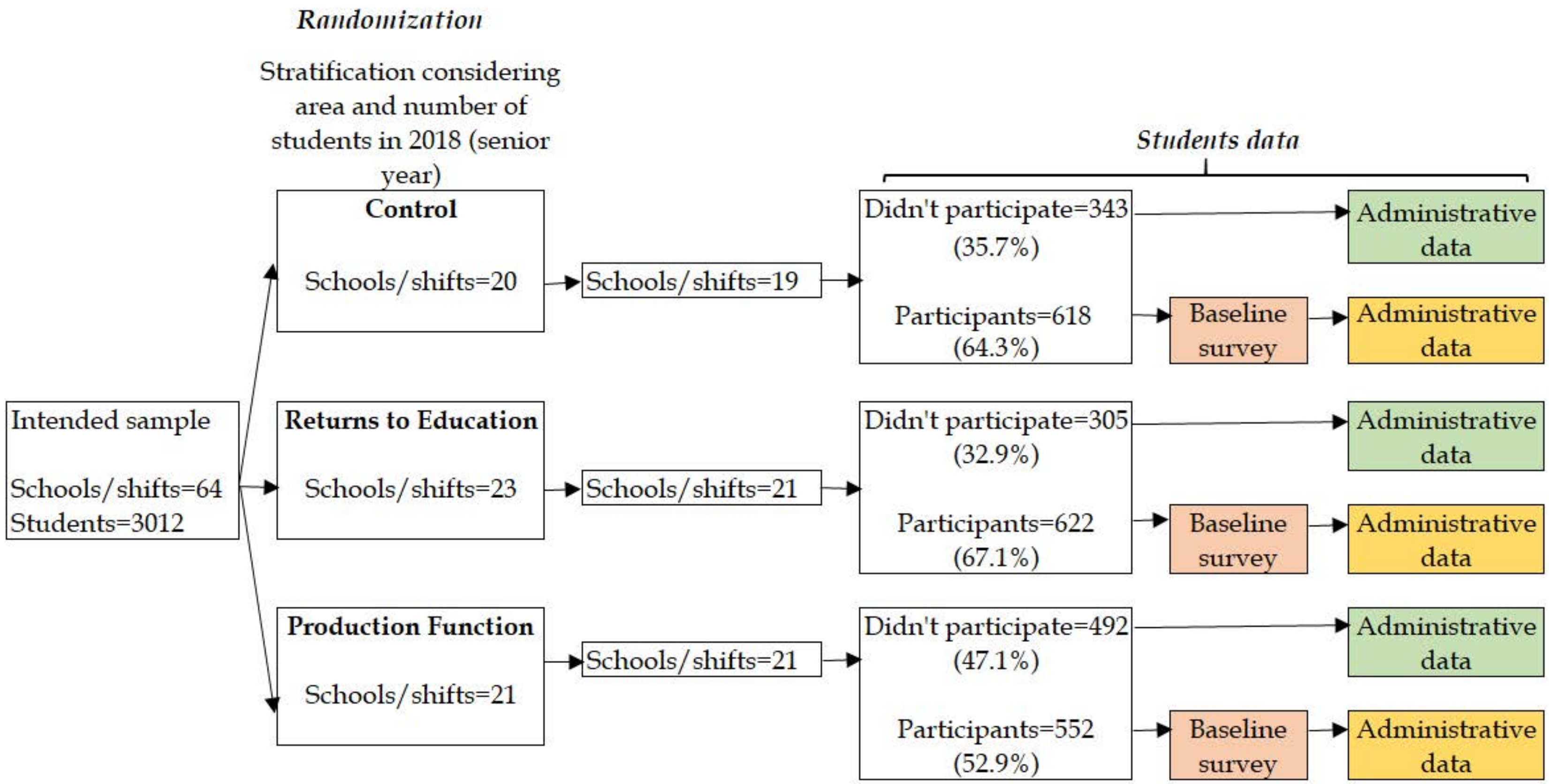
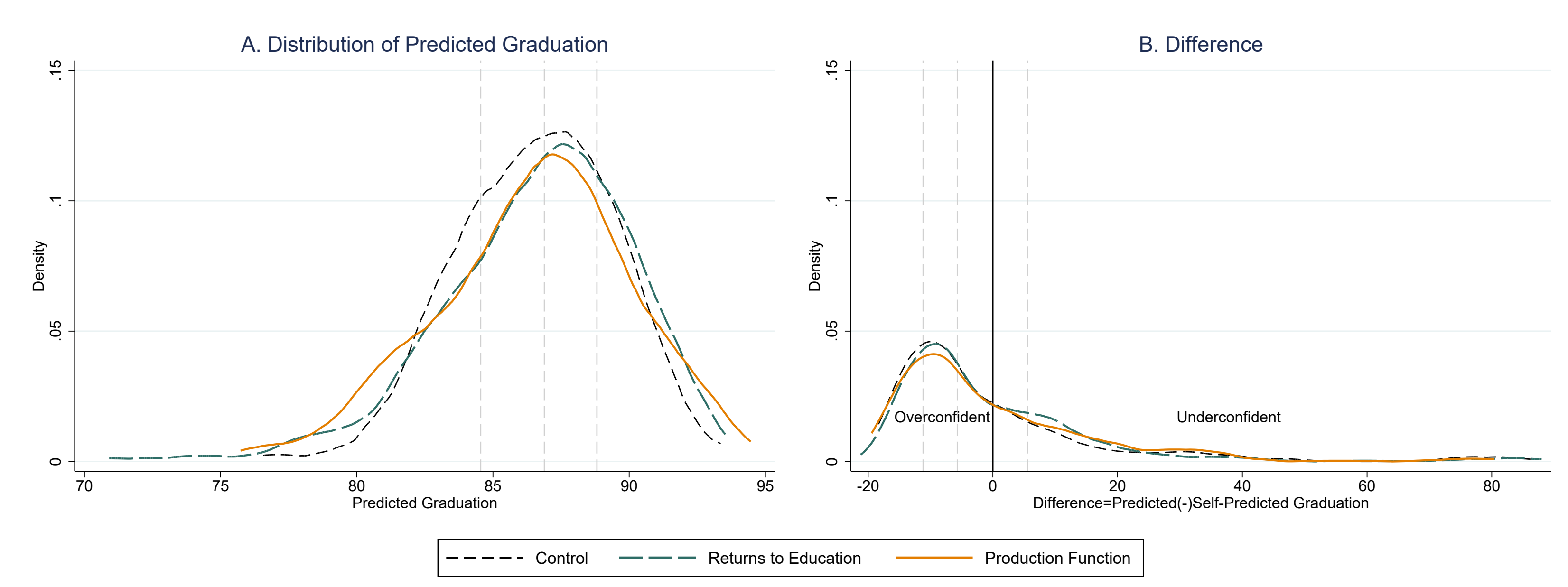
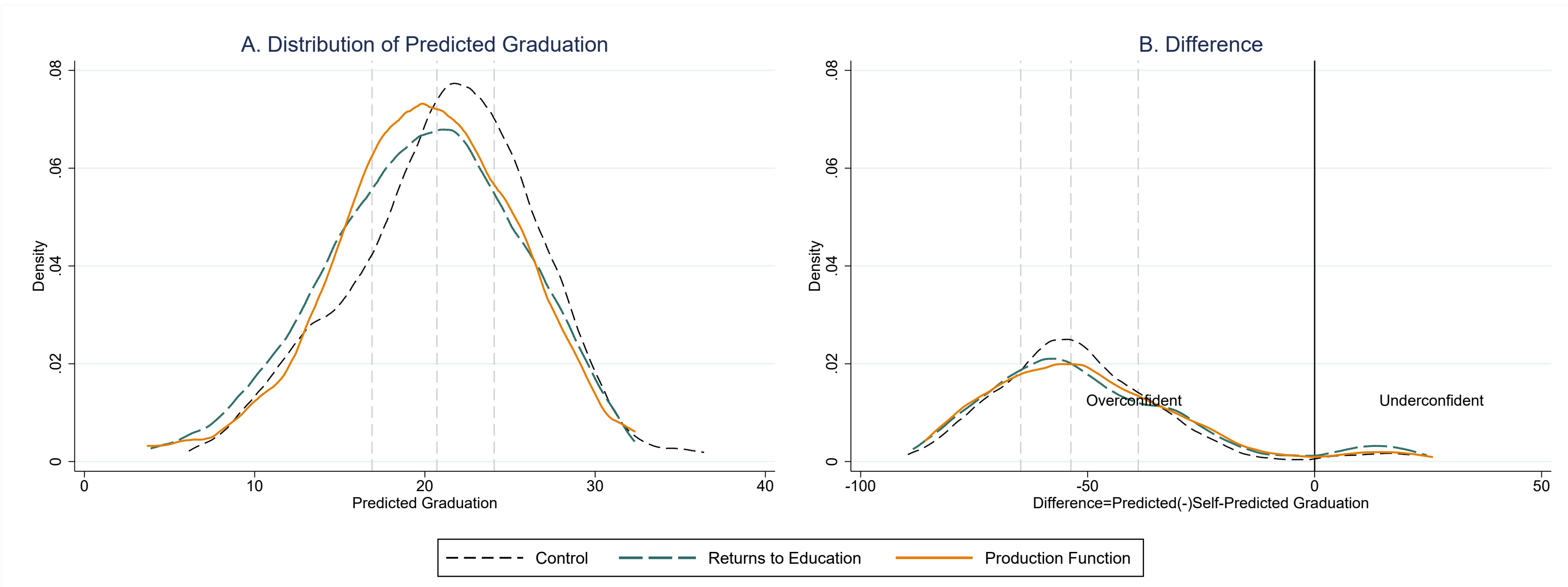


Figure 5: Distribution of Predicted Graduation and Difference with Self-estimation by Treatment Group: Students with Zero Pending Subjects



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 2: Descriptive Statistics from Control Group

	(1) N	(2) Full Sample
Panel A. Graduation Beliefs and Treatment Outcomes		
Graduation (by February 2020)	617	0.504
Math Teachers' Graduation estimation	479	0.735
Students' Graduation estimation at baseline	615	0.784
Students' Graduation estimation at endline	601	0.842
Number of pending subjects at the beginning of the senior year	617	0.887
Number of pending subjects at the moment of the intervention	617	0.781
Number of passed pending subjects by the end of the academic year	341	0.507
Ratio passed subjects by February 2020	617	0.926
Panel B. Students Characteristics		
Gender Student =1 female	617	0.598
Age Student	612	18.000
Student works or takes care of family member	617	0.454
Discount factor 1 (today vs. one week)	532	0.872
Discount factor 2 (one week vs. two week)	522	0.860
Time inconsistency: Discount factor 1/Discount factor 2	496	1.043

Notes: Column 1 reports the number of non-missing observations of variables among all students in the Control group. I elicited time preference using a standard Becker DeGroot Marschak procedure.

Table 3: 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	0.2	-4.61	0.295	0.445	61
	[16.8]	(5.34)	(4.54)			
B. Students Characteristics						
Age	17.9	0.093	0.069	0.909	0.881	1792
	[1.88]	(0.184)	(0.232)			
Gender	0.597	-.017	0.006	0.533	0.813	1792
	[0.491]	(0.042)	(0.042)			
Pregnancy/Has children	0.06	-.002	-.003	0.965	0.978	1705
	[0.237]	(0.017)	(0.014)			
Has email	0.726	-.007	0.046	0.175	0.372	1773
	[0.446]	(0.044)	(0.047)			
Has cellphone	0.858	-.02	-.01	0.696	0.766	1777
	[0.35]	(0.027)	(0.027)			
Lives with mother	0.871	0.001	-.014	0.523	0.802	1792
	[0.336]	(0.022)	(0.026)			
Lives with father	0.581	0.009	-.036	0.153	0.325	1792
	[0.494]	(0.026)	(0.029)			
C. Households Characteristics						
Has computer	0.762	0.026	0.004	0.477	0.633	1783
	[0.426]	(0.032)	(0.038)			
Has internet access	0.845	0.017	0.025	0.77	0.713	1783
	[0.362]	(0.029)	(0.03)			
Persons per room	1.74	-.066	-.009	0.366	0.43	1765
	[0.919]	(0.058)	(0.074)			
Parent has some superior educ.	0.335	-.044	-.032	0.826	0.751	1792
	[0.472]	(0.058)	(0.062)			
Student works or helps in the family business	0.453	-.038	-.026	0.742	0.479	1792
	[0.498]	(0.032)	(0.039)			
Student takes care of family members	0.196	0.029	0.009	0.405	0.549	1792
	[0.397]	(0.029)	(0.029)			
D. Students Academic Performance						
Has repeated a year in high school	0.383	-.012	-.059	0.554	0.694	1792
	[0.487]	(0.081)	(0.073)			
At least one pending subject from previous years	0.552	-.008	-.05	0.306	0.537	1792
	[0.498]	(0.049)	(0.055)			
E. Expectations						
Wants to attend college	0.951	-.03*	-.022	0.667	0.126	1792
	[0.215]	(0.016)	(0.016)			
Wants to work after school	0.872	-.03	-.028	0.949	0.394	1792
	[0.334]	(0.028)	(0.023)			
Perceived probability of obtaining the diploma	0.784	-.014	-.003	0.472	0.686	1789
	[0.22]	(0.018)	(0.017)			

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 4: 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: R = PF	0.038**	0.010**	0.124
P-value: R = PF = 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 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.

Table 5: 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: R × Overconfident = R × Underconfident	0.696	0.358	0.549
P-value: PF × Overconfident = PF × Underconfident	0.381	0.376	0.139
P-value: R × Overconfident = PF × Overconfident	0.020**	0.025**	0.089*
P-value: R × Underconfident = PF × 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 shift and strata fixed effects.. See notes in Table 4 for a list of potential controls. *, **, and *** denote statistical significance at the 10, 5, and 1 percent levels respectively.

Table 6: 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: R = PF	0.859	0.048**	0.041**
P-value: R = PF = 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: R × Overconfident = R × Underconfident	0.449	0.058*	0.431
P-value: PF × Overconfident = PF × Underconfident	0.958	0.002***	0.001***
P-value: R × Overconfident = PF × Overconfident	0.931	0.031**	0.018**
P-value: R × Underconfident = PF × 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 shift and strata fixed effects. See notes in Table 4 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 Perceptions on Expected Earnings by Level of Education

	(1)	(2)
	Graduation	
	Complete Secondary	Complete College
Production Function × Misperception (+)	0.0545*	0.0949**
	(0.0296)	(0.0383)
Production Function × Misperception (-)	0.0317	0.0230
	(0.0426)	(0.0285)
Returns to Education × Misperception (+)	0.123***	0.131***
	(0.0320)	(0.0391)
Returns to Education × Misperception (-)	0.0677	0.0983***
	(0.0455)	(0.0334)
Misperception (+) by Level of Education	-0.0314	-0.0226
	(0.0330)	(0.0341)
P-value: R × Misperception (+) = R × Misperception (-)	0.269	0.473
P-value: PF × Misperception (+) = PF × Misperception (-)	0.652	0.108
P-value: R × Misperception (+) = PF × Misperception (+)	0.016**	0.249
P-value: R × Misperception (-) = PF × Misperception (-)	0.428	0.020**
Mean (Control, Misperception (-))	0.52	0.52
N	1610	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 shift 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 underestimating expected earnings if the perceived expected earnings are lower than actual expected earning by level of education, See notes in Table 4 for a list of potential controls. *, **, and *** denote statistical significance at the 10, 5, and 1 percent levels respectively.

Table 8: Impacts on Graduation by Time Preferences

	(1) Graduation
Production Function × Very Patient	0.0349
	(0.0364)
Production Function × Not Very Patient	0.0394
	(0.0371)
Returns to Education × Very Patient	0.117***
	(0.0347)
Returns to Education × Not Very Patient	0.0438
	(0.0487)
Very Patient	-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 shift and strata fixed effects. To compute the dummy variable Very Patient 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 4 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 Graduation by Poverty Level and Gender

	(1)	(2)	(3)	(4)
	Graduation			
	Poor students	Least 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: R = PF	0.020**	0.726	0.112	0.238
P-value: R = PF = 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 shift and strata fixed effects. To classify students as Poor or Least 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 4 for a list of potential controls. *, **, and *** denote statistical significance at the 10, 5, and 1 percent levels respectively.

Table 10: 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: R = PF	0.909	0.227
P-value: R = PF = 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: R × Overconfident = R × Underconfident	0.556	0.485
P-value: PF × Overconfident = PF × Underconfident	0.160	0.098*
P-value: R × Overconfident = PF × Overconfident	0.606	0.021**
P-value: R × Underconfident = PF × 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 shift 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 4 for a list of potential controls. *, **, and *** denote statistical significance at the 10, 5, and 1 percent levels respectively.

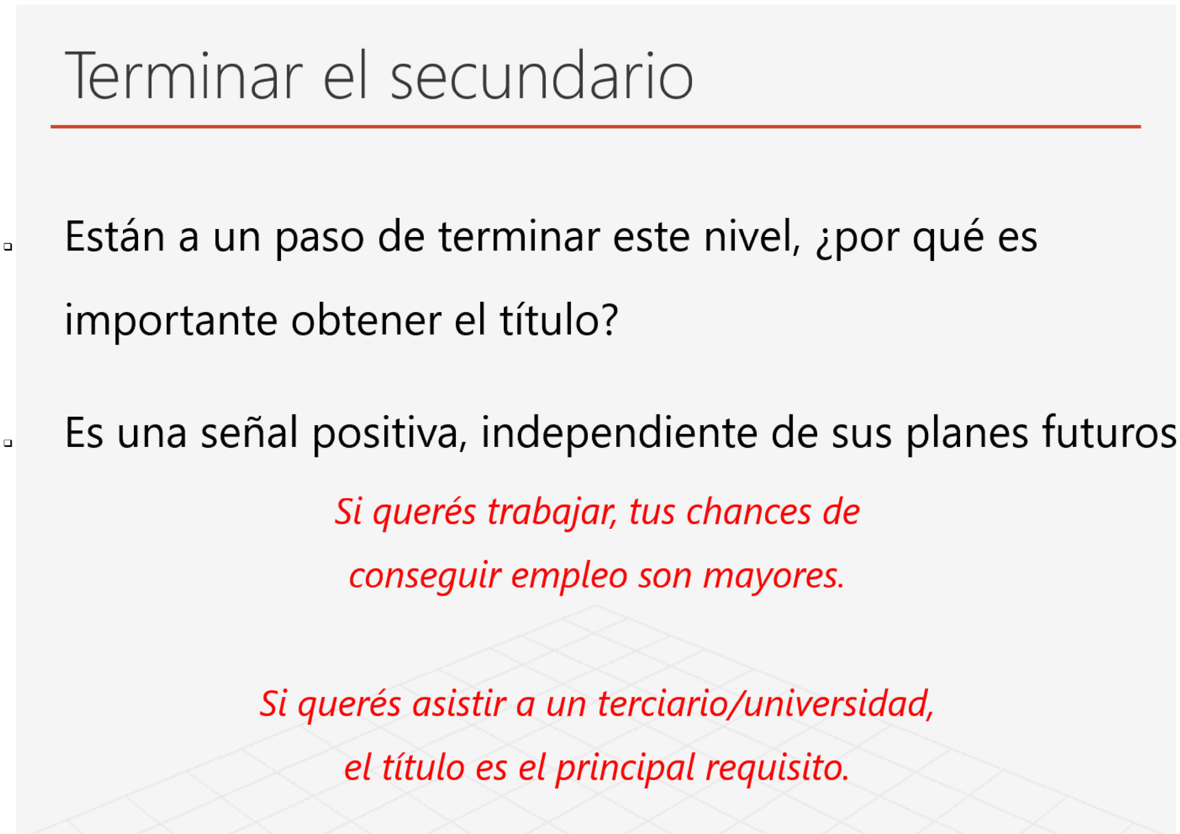
A Appendix: Information Treatment Arms

Information Interventions

I show the specific content introduced to the senior students that participated in each treatment arm. In 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



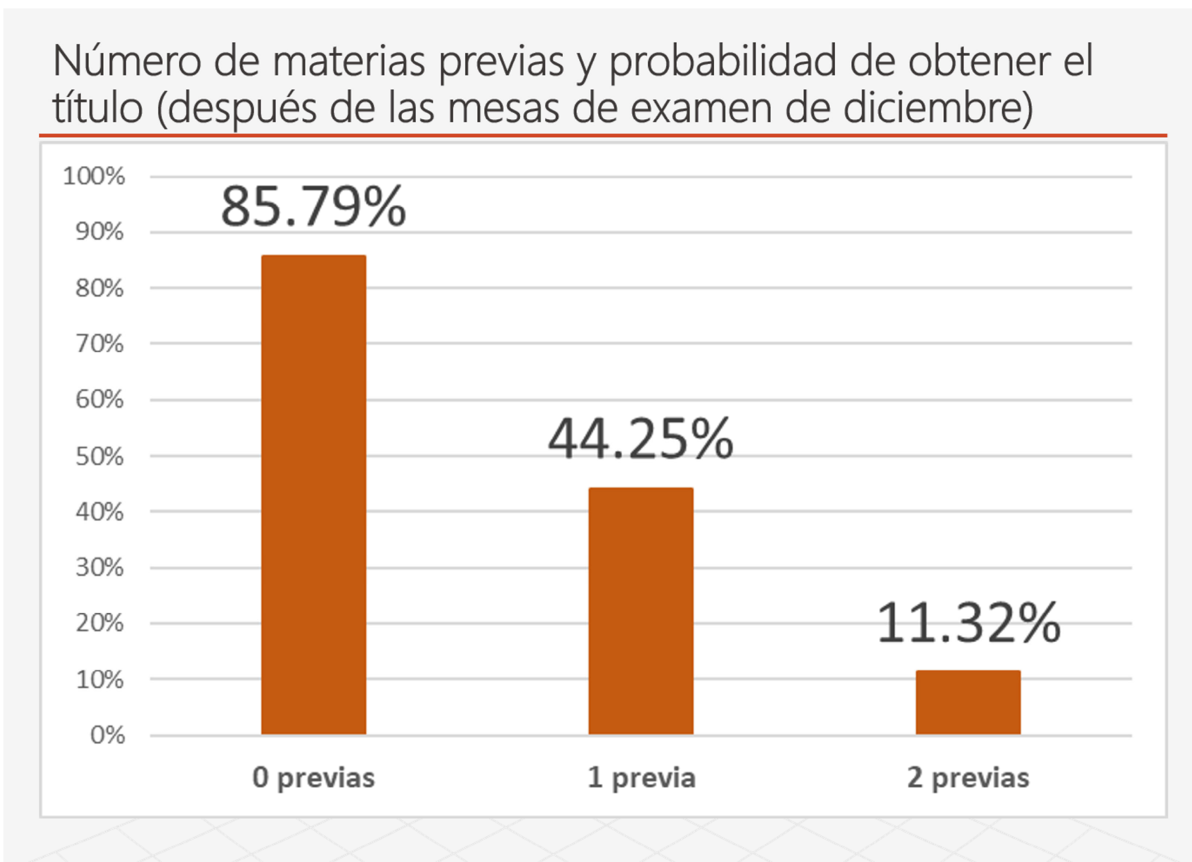
Notes: Common slide showed to all the students who received any of the intervention treatments. Translation: Finish high school, you are really close to finish this level of equation, but why it is important? It is positive signal that does not depend on your future plans: If you want to work, your chances 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 the diploma and also how important was to pass the subjects during the senior year. In order 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 to create a mapping of their situation at the beginning of the senior year and how similar students performed in terms of graduation. Given this could have been shocking news for the students in any standing, I talked about the intermediate steps to transform inputs into outputs and I discussed how to remedy their situation: first, I opened the conversation to discuss the options together (Figure A3), and then I showed a summary of the most relevant tips to effectively obtain a diploma on time.

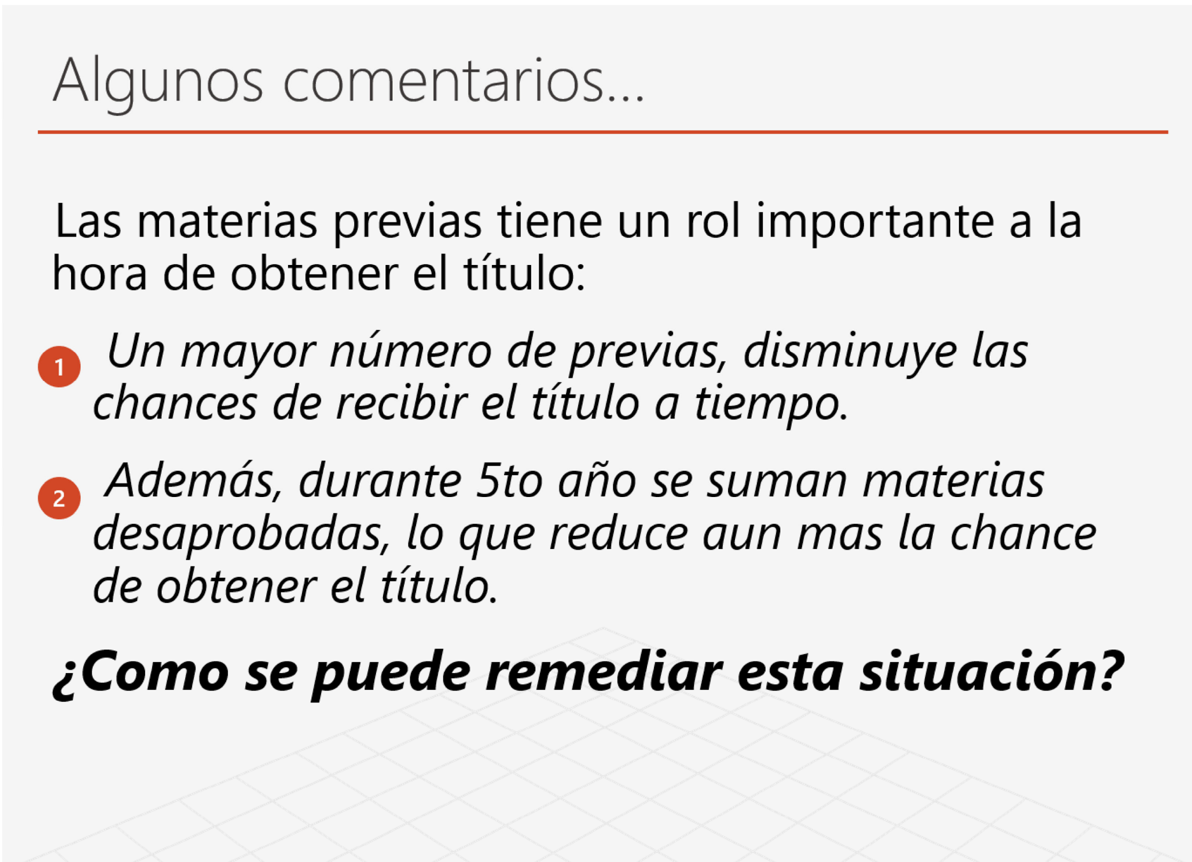
Figure A2: Statistics Shown to the Students



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

The key messages were to devote more time and effort to study the senior year subjects and attend to the examination periods for those with pending subjects. During the senior year there are several social activities (prom night, private parties, graduation trip, etc.). In interviews with the school principals and some focus groups with students from the previous cohorts, they mentioned these activities as a major distraction from academics.

Figure A3: The Role of Pending Subjects

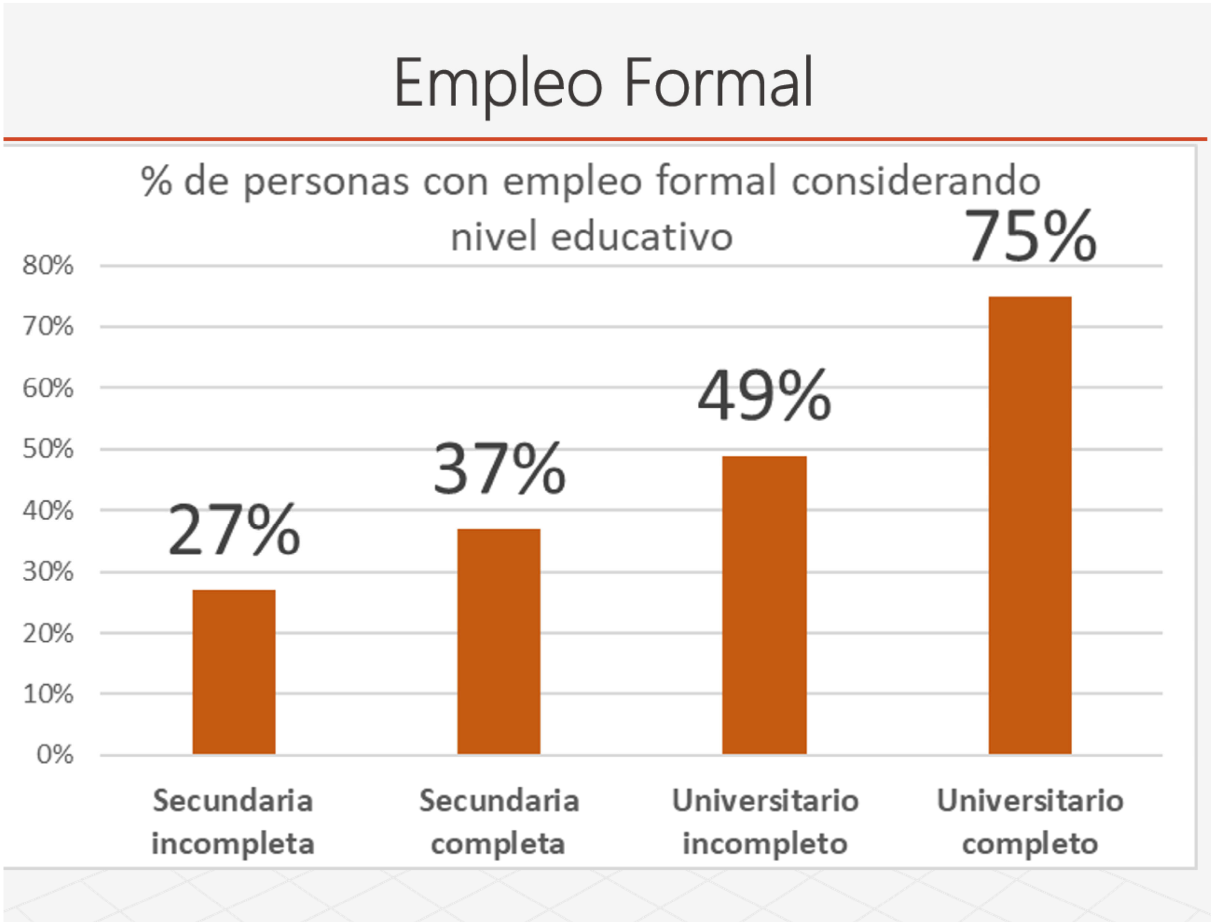


Notes: In this part of the presentation I highlighted the role of the pending subjects and pass subjects of the senior year on timely graduation. Then I opened the discussion "How can this situation be remedied?".

Returns to Education

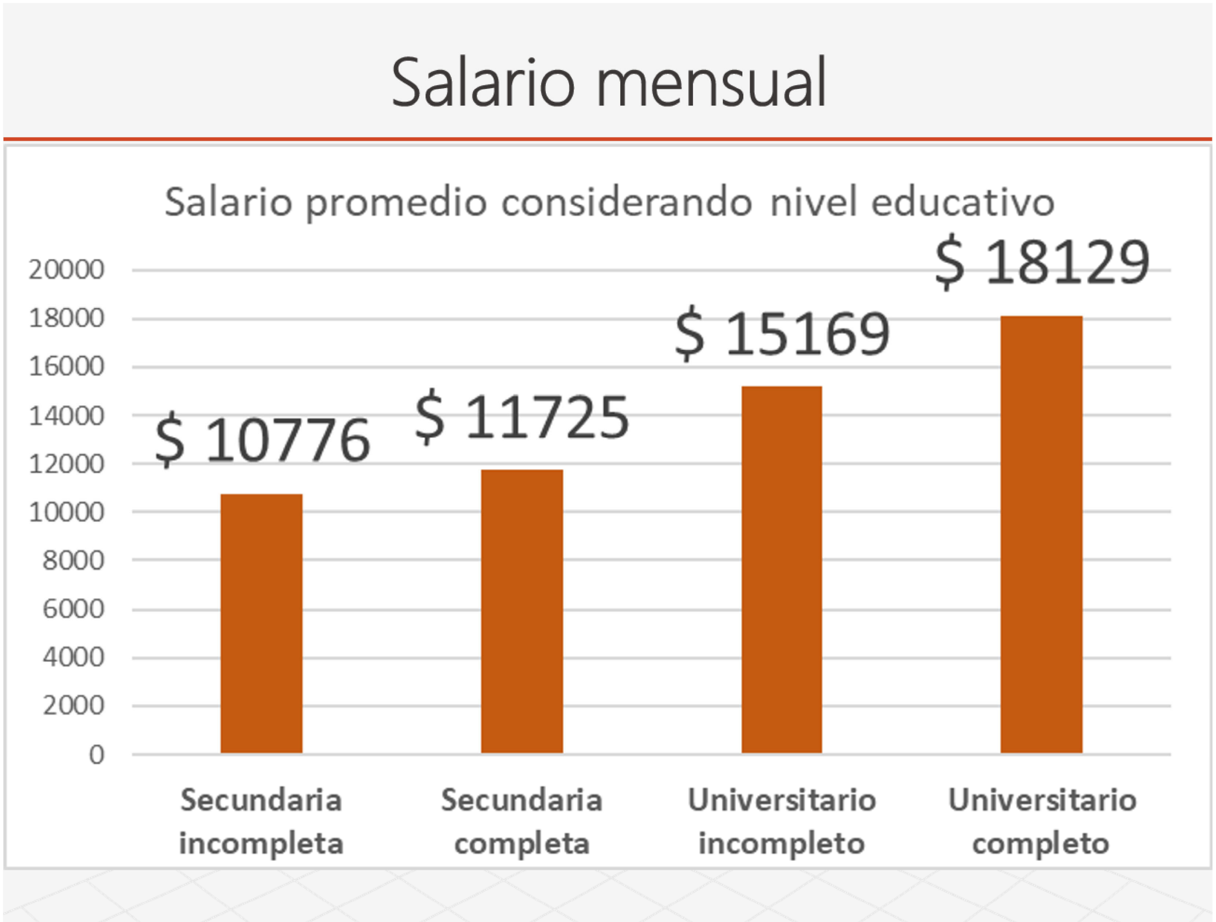
In this presentation I used data from the National Household Survey 2018 (Encuesta Permanente de Hogares) to compute the averages of formal employment and earnings to be shown to the students, I only considered individuals from the province of Salta, between 18-30 years old. The statistics were computed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

B Appendix

Statistical Power

To compute the statistical power, I used data from the previous cohort (2018, subsample of 5 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 3 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 2 main treatments by estimating an effect of 3.5 percentage points in graduation rate with a statistical power of 76%.

Free Online Platform: MOODLE

The Directorate of Secondary Education of Salta required me to provide some useful information to all the students otherwise, I would face issues with school principals reluctant to let me in their schools. So, in order to provide something in exchange, I designed a free online platform with math contents for all the years of high school that could help to remedy the academic standing of students in at least that subject.

At the onset of the project I had 2 rounds of meetings with principals, vice-principals, and senior-level math teachers to hear their opinions about my agreement with the Directorate and 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).

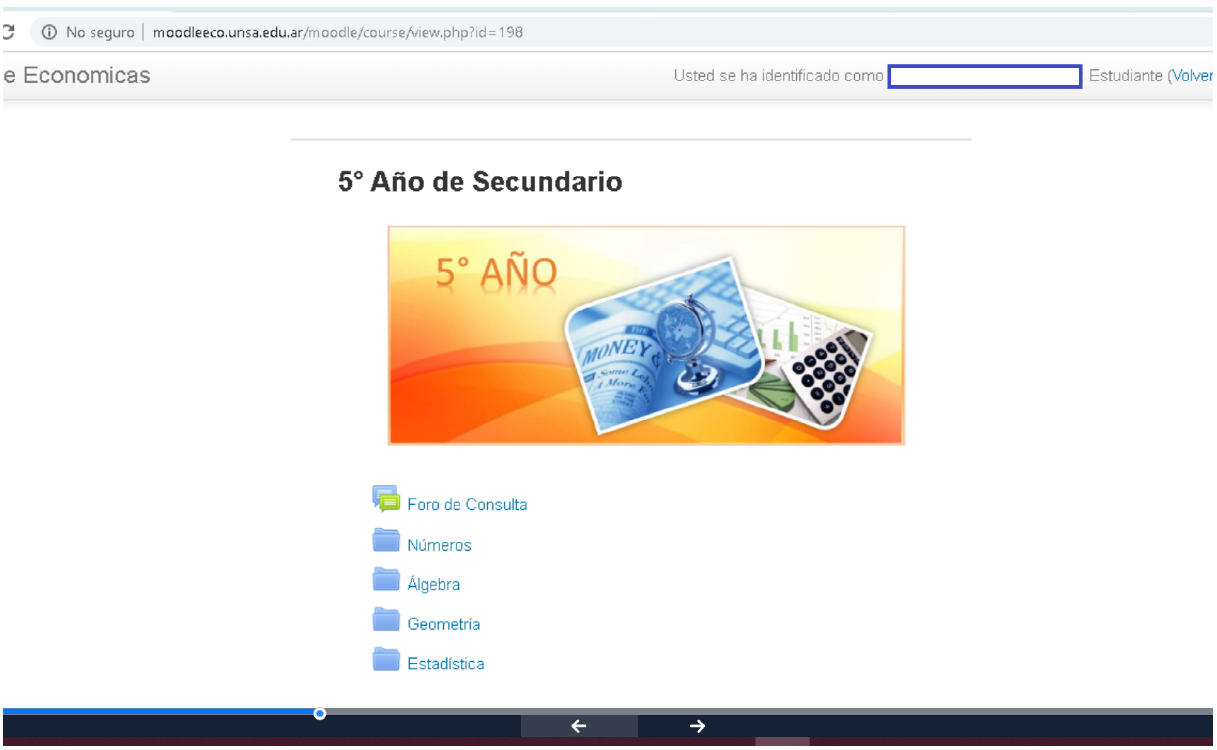
As mentioned above, the platform is not a part of the intervention, but rather allowed me to conduct the baseline surveys in all schools. After being introduced, we explained first the contents of the platform and then how to obtain free access (with a code that was determined for each school, for security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

We also show how to post questions (public or privates) with the commitment on our side to reply to each question within 48 hours. They 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



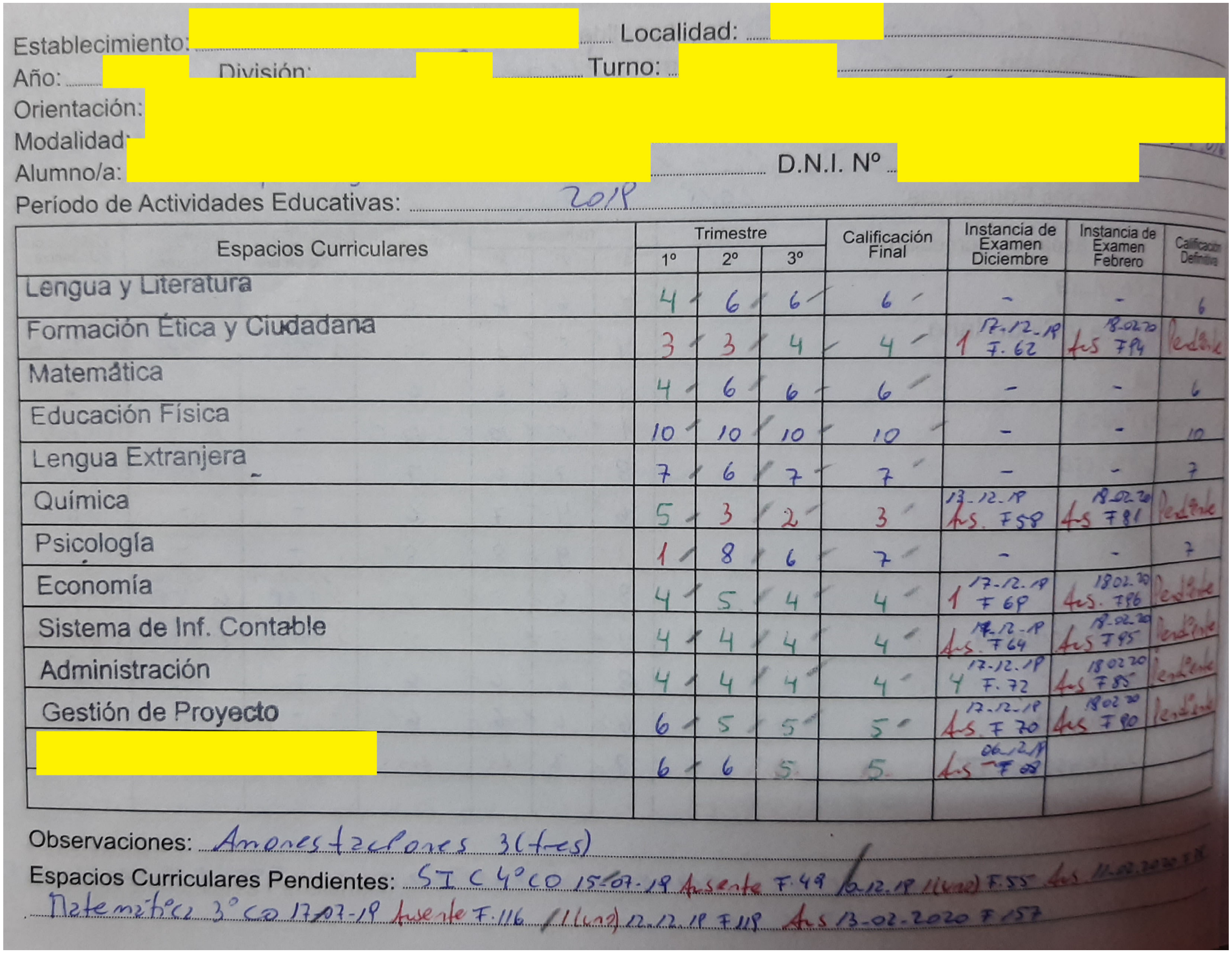
Notes: Screenshot of the platform designed by the Department of Mathematics at Faculty of Economics (UNSa).

Figure B3: MOODLE Platform: Senior year specific content

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

C Appendix: Supplementary Figures and Tables

Figure C1: Student Academic Report



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

Espacios Curriculares	Trimestre 1º	2º	3º	Calificación Final	Instancia de Examen Diciembre	Instancia de Examen Febrero	Calificación Definitiva
Lengua y Literatura	4	6	6	6	-	-	6
Formación Ética y Ciudadana	3	3	4	4	1 17.12.18 F. 62	Aus 18.02.20 F94	Pendiente
Matemática	4	6	6	6	-	-	6
Educación Física	10	10	10	10	-	-	10
Lengua Extranjera	7	6	7	7	-	-	7
Química	5	3	2	3	13.12.18 Aus. F58	18.02.20 Aus F81	Pendiente
Psicología	1	8	6	7	-	-	7
Economía	4	5	4	4	1 17.12.18 F 69	18.02.20 Aus. F86	Pendiente
Sistema de Inf. Contable	4	4	4	4	17.12.18 Aus. F64	18.02.20 Aus F95	Pendiente
Administración	4	4	4	4	4 17.12.18 F. 72	18.02.20 Aus F85	Pendiente
Gestión de Proyecto	6	5	5	5	17.12.18 Aus. F 70	18.02.20 Aus F 90	Pendiente
[redacted]	6	6	5	5	06.12.18 Aus F 68		

Observaciones: Amonestaciones 3 (tres)
Espacios Curriculares Pendientes: SIC 4º CO 15-07-19 Ausente F.49 12.12.19 1 (uno) F.55 Aus 11.02.2020 F.[illegible]
Matemática 3º CO 17-07-19 Ausente F.116 1 (uno) 12.12.19 F.119 Aus 13-02-2020 F.157

Notes: Example of an individual school record. The format is similar in all secondary schools.

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 4 for a list of potential controls. *, **, and *** denote statistical significance at the 10, 5, and 1 percent levels respectively.