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

JMP: Click here for the latest version

Carolina Lopez*

September 26, 2022

Abstract

Lack of information or cognitive biases could lead students to exert insufficient effort to obtain their secondary school diploma on time, which may have long-lasting consequences on their lives. In an experiment with high-school students in Argentina, I randomize the provision of 2 types of information: graduation rates of similar students of the previous academic year by academic status, along with tips to remedy their academic standing, and information about the returns to education by achieved level of education. Both treatments increased graduation by 10 and 20 percent, respectively. Poor-performing students at baseline respond most to the treatments and I do not find differences by gender. In addition, the probability of college enrollment increases by 38 percent in both treatments. These findings indicate that inaccurate beliefs about own future performance and labor market characteristics explain a significant part of the low graduation rates in high school in a developing context.

JEL Classifications: C93, D83, D91, I25, I26, O10

*Brown University, Department of Economics. Email: carolina_lopez@hks.harvard.edu. I am incredibly thankful to Andrew Foster, Pedro Dal Bó, Bryce Steinberg, Anja Sautmann, and Neil Thakral for their support and encouragement. I am grateful to Alex Eble, Simone Schaner, Jesse Shapiro, Alejandro Ganimian, Natalie Bau, seminar participants at the Brown Applied Micro Lunch, and the Development Tea group members at Brown University for helpful discussions and feedback. I am specially grateful to Asim Khwaja for valuable comments. Also, I would like to thank to Angélica Astorga, Felipe Brugués, Amanda Loyola Heufemann, Ignacio Lopez Erazo, Ana Pacheco, Juan Pablo Uribe, Santiago Hermo, Diego Ramos-Toro, and Natalia Sández Pernas for their invaluable support. The field experiment would not have been possible without the authorization and support of the Directorate of Secondary Education of the Ministry of Education of Salta and the authorities of each participant school. I acknowledge the financial support I received to digitized academic records from the Orlando Bravo Center for Economic Research at Brown University and CAF–Development Bank of Latin America. The experiment was approved by the Brown University Institutional Review Board and was registered with the pre-analysis plan at AEA RCT registry (RCT ID: AEARCTR-0004511). All errors are my own.

1 Introduction

Education is a key lever for both economic growth and intergenerational mobility (Krueger and Lindahl, 2001; Chetty et al., 2014; Psacharopoulos and Patrinos, 2018). Even as barriers to education have decreased over time for children in low and middle-income countries, a large educational achievement gap persists between these children and those in higher-income countries (Glewwe and Muralidharan, 2016). In Argentina, for example, while most teenagers in school age are enrolled (92.4 percent), only 50 percent of those who reach their senior year and complete their coursework ultimately receive their diploma. Potential reasons for this gap include lack of information or cognitive biases, which leads students to exert levels of effort which, unbeknownst to them, are insufficient to complete their degree. Such information gaps or cognitive biases are likely most salient for low-income households and households which do not have exposure to mentors or successful graduates who are able to provide accurate information. A key question for both policy and global welfare is therefore how to induce greater levels of education in these contexts.

Previous literature has found that incentivizing academic achievement (outcomes) often has no effect on performance (see Ganimian and Murnane (2016) for a meta-analysis), but incentives can improve educational performance when specific tasks (inputs) are targeted. Fryer (2011) and Fryer (2016) suggest that a potential explanation for students' failure to transform effort into academic achievement could be a lack of adequate knowledge about the education production function. In this paper, I study the channel through which effort is transformed into academic achievement.

I conduct a randomized controlled trial in 61 high schools in the city of Salta, Argentina, to understand whether providing information can improve high school graduation rates and to study the mechanisms behind it. Many of the students in this area are at risk of failing to convert enrollment and attendance into graduation. I estimate the impact of two interventions on the likelihood of graduation for students currently enrolled as high school seniors. The first intervention provides information on how to get a high school diploma—that is, on the intermediate steps needed to effectively transform effort into educational achievement. The second intervention is a standard provision of an estimate of the economic returns to education —used as a benchmark, and to provide a new test, given the mixed evidence of its efficacy in existing studies. In this setting, a consequence of not getting a high school diploma is drastically lower chances of obtaining a high-quality job.[1] [2]

My study has three arms: *Production function*, *Returns to education*, and *Control*. Both information treatments were introduced through a brief presentation in a single visit to each school and reinforced with reminder messages. In the *Production function* arm, the presentation contained statistics on the previous cohort's graduation rates based on their academic standing at the beginning of their senior year, along with information about the intermediate steps necessary to improve academic standing and ensure on-time graduation. This piece of information was meant to generate a mapping between each student's academic standing (known by the student at the time of the intervention) and their chance of graduation, by showing them *how* to transform *inputs* (effort) into *outputs* (high-school diploma). In the *Returns to education* arm, students were shown information containing employment levels and wages by levels of education, using the same format as in the other treatment arm. In the *Control* group no information was provided. I combine a baseline survey, hard copies of individual academic records collected from each school, and administrative data of each school to analyze the impacts of these interventions. The participants included almost 1800 senior students attending public high schools.

I find that both treatments have a positive and significant effect on graduation rates. Specifically, the *Returns to education* treatment increases the probability of graduation by 10 percentage points (almost 20 percent with respect to the control group), and the *Production function* treatment increases graduation by 5 percentage points (10 percent). The effect of *Returns to education* is two times as large as the effect found in Jensen (2010) for his subsample of less poor students; the effect of *Production function* is similar in magnitude. The students with the greatest increase in the probability of graduation in both treatment arms are those with the worst academic standing at the beginning of their senior year. In addition, an increase in observable effort —measured as the probability of attendance to retake exams and the probability of passing those exams— can be observed among those students.

Empirical evidence shows that individuals tend to overestimate the probability of important outcomes (Feld et al., 2017; Heger and Papageorge, 2018; Machado et al., 2018), leading to suboptimal decisions, especially for unskilled individuals (Choi et al., 2014). To test this

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

[2] Jensen (2010) and Nguyen (2008) show positive, Bonilla-Mejía et al. (2019) null, and Loyalka et al. (2013) negative effects.

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

My main contribution is to provide evidence of how small but powerful pieces of information, provided on time, can improve students' decisions in a high-stakes setting. Previous papers test whether students can be motivated to invest more effort in education by providing monetary or non-monetary incentives. In contrast, with a new piece of information tested in an educational setting, I study whether students' lack of knowledge of the educational production function has an impact on the probability of high school graduation.

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

In addition, I contribute to the literature seeking to understand why people do not use services, infrastructure, or adopt new technologies that can improve their wellbeing when they become available to them. This concern, known as "the last mile problem" —although the term has roots in other fields, Soman (2015)— is present in many contexts (Mullainathan and Shafir, 2013): individuals forget to submit their taxes on time, low-income students do not use financial aid programs to attend college (Bettinger et al., 2009), farmers do not adopt fertilizer (Duflo et al., 2011), among others. Consequences of these not optimal decisions are more detrimental in contexts where individuals lack family or other forms of social support (Mullainathan and Shafir, 2013) and may impede those without such resources on their way out of poverty. I show that uncertainties about how to apply effort to achieve a desired outcome play a substantial role for students attending their last year of high school.

Also, I study whether students ignore or discount new information on finishing high school because of biased beliefs about the information they already have (DellaVigna, 2009). People tend to overestimate their own abilities. In particular, overconfidence in an educational context may lead students to study less (Nowell and Alston, 2007). I show how this biased belief in their own performance is detrimental to students' chances of graduation, and I demonstrate that those negative consequences can be ameliorated by providing accurate information about how to achieve a high school diploma and the economic returns to education.

This paper is relevant for informing policy strategies to increase the demand for high school diplomas among teenagers, especially those who are disadvantaged and at risk of failing to complete high school on time.[3] I study a vulnerable population in a high-stakes setting where students' probabilities of failing to obtain high school diplomas are high. As consequence, individuals in this setting have a high chance of being classified as not in education, employment, or training (NEET), which represent an increasing concern in Latin America. Although access to the educational system is not restricted in many settings, youths' lack of information can cause them to invest less than the optimum level of effort in education, which in the medium run will limit their economic opportunities by preventing them from attending college and working in a job market that uses high school diplomas as a signal.

The remainder of this paper is divided as follows. In Section 2, I briefly describe the context in which I carried out this randomized controlled trial. In Section 3, I discuss the theoretical framework and predictions for graduation and mechanisms. Section 4 describes the experimental design, randomization, and details of the information interventions of this

[3]Discussions are currently occurring in many countries and international organizations such as UNICEF (Annual Report 2020 `https://www.unicef.org/reports/unicef-annual-report-2020`) on how to recover from the consequences of the COVID-19 pandemic and the related closure of schools and the impacts on student achievement. Low high school diploma achievement was already a concern before the pandemic in Argentina. UNICEF has reported low school achievement (UNICEF-ARGENTINA, 2017), a referent from the private sector highlighted difficulties in hiring young people with a high school diploma (Diario La Nación, August 6 2021), and civil associations, along with the current National Director of High School level, have expressed concerns related to low completion rates (Diario La Nación, August 7 2021).

paper, Section 5 shows the main results, along with their underlying mechanisms. Section 6 presents the main conclusions.

2 Context

In Argentina, education is compulsory up to the end of secondary school; there are free public schools in every district and transportation is sometimes free for students as well. Secondary education is thus accessible for most students. As a result, the share of secondary school-age youth who are attending secondary school is 91.2 percent, with 74.7 percent attending public schools (CEDLAS and World-Bank, 2018). However, high school graduation rates remain low throughout the country. Less than half of the teenagers enrolled in high school actually graduate (UNICEF-ARGENTINA, 2017). Students drop out at different points during high school, but even those who complete the senior year (and attend until the last day of classes) often do not obtain a high school diploma because they fail to fulfill all the mandatory requirements of the system. This is explained in the following subsection.

2.1 Educational System and Students' Academic Standing

Students may not graduate because they drop out at different points during high school, mainly owing to "the need to assume adult roles, such as working outside or inside the home, caring for younger or older family members, or taking care of other domestic chores... Other students drop out because they are not able to deal with school institutional guidelines."[4] But another important explanation, which has attracted less research attention and is not even mentioned by the Director of Secondary Education at the national level, is that *students who attend until the last day of high school may still not obtain a high school diploma*. This topic has remained unexplored basically because there are no digitized data at the individual level that allow making conclusions about the magnitude of this issue.

To graduate from high school, students must pass a fixed number of subjects per year (usually 10-12).[5] The academic year begins in March and classes finish by December, but the year officially ends in February. In December and February there are examination dates which allow students who failed subjects during the academic year to remedy their academic standing. Students who receive a score higher than 5 (the exams are graded on a 10-point

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

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

scale) pass subjects which they previously failed. If a student does not remedy their standing in all subjects by the beginning of the next academic year, they can still be promoted with at most two failed subjects —with a grade lower than 6 (if a student has three or more failed subjects, they must repeat the year). Those failed subjects must be passed at some point during the students' following years of high school to receive a diploma; I refer to these failed subjects as *pending subjects* going forward. All high schools have three examination dates on which students can pass pending subjects each year (July, December, and February). At any given time while in high school, students can have at most two accumulated pending subjects (for example, they can have one from the second year and another from the third year).

Each student is fully aware of the number of pending subjects they have.[6] I use this concept throughout this paper to categorize students by academic standing at the beginning of their senior year. They can be considered as "in good standing" (zero pending subjects) or "in bad standing" (one or two pending subjects). During phone interviews, school administrators said that the main driver of low graduation rates is the prevalence of pending subjects; the administrators report that students either fail the examinations that would allow them to pass pending subjects or do not attend them at all.

2.2 Educational Situation in Salta

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

According to self-reported data from an anonymous national survey of students collected at the end of the 2017 academic year (Aprender, 2017), almost 40 percent of senior students were in bad standing (had at least one pending subject). In Figure 1, I show data from the control group (cohort 2019) and I observe that at the beginning of their senior year more than 55 percent of the students had at least one pending subject. These findings indicate

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

that the chances of timely graduation for these cohorts were low, and at the same time it reveals how common it is for students to have pending subjects at the beginning of the academic year.

At the onset of this study, qualitative field work was conducted to understand why students who had already invested at least 5 years of their lives attending high school were failing to obtain a diploma in their last year. Principals, other school authorities, and teachers were in accord in reporting that students do not exert enough effort to pass pending subjects and often do not attend the examination periods to remedy their standing. They also note that these issues become worse during students' senior year.[7] Students in bad standing stated that they did not use the examination exam dates because they had other "important" matters but they would use the next one "for sure," pass the exam, and receive a diploma on time (by the end of the senior academic year). Their confidence in being able to complete this process suggests cognitive dissonance regarding what they believe about their actions and effective effort to obtain the diploma. I use this insight in the next section to develop a theoretical framework that relates beliefs to effort.

3 Theoretical Framework

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

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

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

will not change their effort.

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

Assumptions

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

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

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

States of the World.— Students can have uncertainty about the returns to effort in the senior year and their ability. I assume there are only two potential states that combine those beliefs: the first one has a probability p and the second one $(1-p)$. There are four potential combinations of β_i and α_i. A student could think that the return to effort is low to get the diploma but it could be compensated with high ability; or the student could think that they own ability is low, so to get the diploma a high return to effort is perceived; and so on.

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

- State 1 occurs with probability p this state is represented by β_l and α_h.
- State 2 occurs with probability $(1-p)$ this state is represented by β_h and α_l.

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

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

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

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

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

The maximization problem is the following:

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

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

Role of the Treatment Arms

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

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

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

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

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

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

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

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

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

3.0.1 Summary of Mechanisms

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

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

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

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

4 Experimental Design

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

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, the material prepared for students (contents for the online platform, survey instrument, and presentations) was approved by the Ministry of Education; officials at the Ministry of Education were not informed in advance which information treatment arm would be 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 afternoon shifts due to logistic/budget constraints. Power calculations were conducted using information from the 2018 academic year. In 2018, there were 2933 enrolled students in the senior class across 63 school-shifts. The unit of randomization is at the school-shift level given that randomization at the individual or class level would be more likely to contaminate the control group.

4.3 Timeline

At the beginning of this project, in mid-October 2018, I contacted authorities of the Ministry of Education of Salta. The office in charge of supervising my intervention was the Directorate of Secondary Education. They have overseen all the stages of the intervention. In addition to having their approval, I needed the direct approval of each school's principal and vice-principals, who were more aware of the specifics of each shift: school festivities, exams, and trips.[9]

[8] From hereon, Salta refers 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 principal per each shift. From hereon, I use the term "school" to refer to "school-shift".

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

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

4.4 Data

Baseline Survey

A description of the baseline data collection process follows. At least 2 days before the intervention date, the research team visited and delivered to the school administrators envelopes containing consent forms for parents of senior students. At a date and time agreed on with the school administrators, the team met with all students of the school in a single room.[10] A description of the activities conducted during each visit day is shown in Figure 3.

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

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

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

At the meeting with students, school administrators introduced the research team. Then, tablets were given to students, a short presentation (containing slides with pictures) was shown to instruct students on their use, and the students were asked to fill out the questionnaire. At the same time, a brief explanation of the questionnaire was provided.[11] Afterward, the research team showed a presentation introducing the online platform. If applicable, the information treatments were then conducted. After the presentation, the research team asked students to answer an additional question about their perceptions of their own graduation (the same question as in the beginning of the questionnaire). This question was intended to test for any changes in students' perceptions after hearing the information presented, and is the only experimental outcome included in the survey.

Given that a single presentation, including statistics and unknown facts for the students, could not have been enough to change the students effort, I sent an SMS and/or email two weeks before the December examination period (senior students' chance to pass pending subjects and failed subjects) to briefly reinforce the information treatment received (excluding students attending schools in the control group).[12] As was shown in previous papers, reminders can help to boost information interventions (Damgaard and Nielsen, 2018).

School Academic Records

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

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

[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 2020 academic year —the academic year immediately after the graudation of my treated cohort— 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 set up jointly by the social security administration, ANSES (Administración Nacional de Seguridad Social), and the national tax authority, AFIP (Administración Federal de Ingresos Públicos).

4.5 Experimental treatments

The treatment assignment was randomly determined at the school level stratifying by the number of students and geographic area of Salta. Information interventions considered in this study are described below.

> *Control*: No information treatment was provided. As in the other arms, this group received the presentation about the free online platform and its use is not part of this analysis.
>
> *Production Function*: Using data from a subset of students of the previous cohort (2018), I computed the mean of a dummy variable that indicates the rate of on-time graduation (by December 2018, after the December examination period) for students with and without pending subjects at the beginning of the 2018 cohort's senior year. The overall on-time completion rate for this subsample was 50 percent. Having pending subjects is not necessarily the main cause of failure to obtain a diploma—students can fail to pass additional subjects in their senior year—but providing this information would highlight the role of pending subjects in getting a diploma and the importance of using examination periods. The provision of this information should highlight aspects of the production function of high school graduation that students do not fully know or understand, such as how much effort should be devoted to passing pending subjects and subjects taken during students' senior year. A full description of the treatment is in Appendix A.
>
> Suggestions about *how* to improve academic standing were provided to all students (because at the time of the visit the status of each student was unknown). All of these suggestions were *intermediate steps* to effectively transform inputs into outputs. The information provided included the following: request mock exams

(*modelos de examen*) from teachers,[13] ask 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), talk with teachers in advance to ask them for studying recommendations, or ask which teachers will be a part of the committee in each subject.[14]

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

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

Figure 4 shows that students' participation differed between the intervention treatment arms. A higher percentage of students and parents decided not to participate in the *Production function* treatment. This selection into participation could have had detrimental impacts on the analysis of this treatment arm, but the protocol of the visits to the schools allowed me to discard selection in participation. No school authorities knew beforehand which treatment was assigned to their school. The research team itself only knew which treatment should be implemented 30 minutes before the arrival to each school. To test for the reason of participation differences, Figure C2 in Appendix C shows that the difference

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

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

is driven by a single school with low participation rate, as it can be observed in Panel B. The main results of this paper are robust to the exclusion of that school (see Table C1 in Appendix C).

4.6 Measuring Students' Confidence in Graduation

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

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

Figure 5 shows in Panel A the distribution of the estimated probabilities for students with zero pending subjects, and in Panel B the distribution of the difference with respect to the self-estimation of students' graduation probabilities. Figure 6 shows the same distributions for students with at least one pending subject. According to my definition of confidence, students with a positive difference are classified as underconfident (the objective measure is higher than the subjective one) and those with a negative difference as overconfident. Figure 7 shows that there are no differences across treatment arms.

5 Results

5.1 Description of the Control Group and Balance Checks

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

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

where y_{is} is the outcome of interest for student i who attends school-shift s, the dummy variables $ProductionFunction_s$ and $ReturnsEducation_s$ indicate which information treatment school s received, δ_s indicates the strata fixed effects (Bruhn and McKenzie, 2009). Errors are clustered at the school level. To control for previous differences in graduation, I add graduation rates at the school level from the previous cohort (senior students in 2018). Each row shows results from a separate regression. Columns 4 and 5 show p-values of the tests of $PF = RE$ and $PF = RE = 0$, given that the comparison of the two information treatments is of special interest.

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

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

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

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

5.2 Empirical Strategy and Main Results

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

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

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

Table 3, column 1, shows that graduation for all students who were selected to participate in either treatments arm increases and the effects are statistically significant: (1) students in the *Production function* treatment arm are 5 percentage points more likely to graduate (10 percent with respect to the control group) and (2) those in the *Returns to education* are 10 percentage points more likely to obtain a diploma (20 percent with respect to the control group). I find that the differences associated with these treatments are statistically significant.

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

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

5.3 Mechanisms for Production Function and Returns to Education

Perceptions on Graduation and Updating

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

Under the theoretical framework shown above, perceptions of graduation should only change if students update their beliefs about the level of effort needed to obtain their diploma. This is only possible if they receive information about the actual probabilities, the effort that is required, and all the intermediate steps needed to successfully transform that effort into graduation. Table C2 shows the change in the *subjective* probability of graduation. Individuals who received the *Production function* treatment became more accurate with respect to their own chances of graduation: the variable decreases by 2 percentage points compared to the baseline response.[15] I observe that in the experimental outcome they become

[15]Notice that the students in the control became less accurate (more optimistic about their chances of graduation). A reasonable explanation for this result is that a visit to the school by a student at an American

more accurate, but this result could not be transmitted into effective effort to remedy their standing. As expected by the design of the treatments, the most striking and significant differences are observed in the *Production function* arm.

I analyze graduation by academic standing and its relationship to my definition of confidence in Table 4. I interact the treatment received with the level of confidence (under- or overconfident) and I show the results for the entire sample in column 1 and then by academic standing at the beginning of the students' senior year (columns 2 and 3). Overall, the results show that none of the treatment arms caused a discouragement effect. Although differences in the probability of graduation exist between the under- and overconfident students (in both treatment arms), the differences with the largest magnitude are observed in the *Production function* arm for students in bad standing (column 3). There are positive and statistically significant effects (at the 5 percent level) for both under- and overconfident students, with a difference of 20 percentage points (but nonsignificant) in favor of underconfident students.

This result indicates that while the *Production function* treatment made the perceptions of overconfident students statistically more accurate (Table C2) that effect fades away until the end of the academic year.

Effort

I analyze the effect of the information treatments on variables that indicate direct measures of effort to pass pending subjects: (1) enrollment in the examination period (December 2019 and/or February 2020) and (2) attendance at the examination period. The first variable indicates the degree of effort because according to high school rules, only students who explicitly register for the examination date are allowed to take the exam.[16] The second variable indicates whether students actually attended the examination. I did not restrict this variable to enrollment.

Table 5 Panel A shows positive impacts of the information treatments on these outcomes, but only for those who received the *Returns to education* treatment. For these students, the effect is statistically significant at the 1 percent level in columns 2. Panel B shows the effect of the information treatments by confidence level at baseline. As discussed above, underconfident students are those who respond to the treatment by increasing their effort more than overconfident students; the difference in the *Production function* between the two types of students is more than 40 percentage points (significantly different at the 1 percent level). The *Returns to education* treatment arm also has differences in favor of the

university and students at UNSa could have generated an optimistic response among students given that there is almost no formal connection between secondary and post-secondary levels.

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

underconfident students, but they are lower; only in column 2 is the difference with respect to the overconfident students significant (at the 10 percent level).

Performance

To understand how the information treatments impact students' performance during the academic year, I separate the analysis by considering what happens with the mandatory senior subjects and pending subjects by February 2020 (the end of the academic year). Both of these variables determine if a student receives a high school diploma: if they pass ALL the senior subjects and have no pending subjects, then they graduate. To understand my results better, I split the sample considering the number of pending subjects, but these results should be considered with caution because of the small sample sizes.

Table 6 shows the impact of the information treatments on a dummy variable that indicates if the student passed all the senior subjects for the entire sample and by separating the analysis by the number of pending subjects that the students had left after July 2019 (after the visit of the research team). Column (1) shows that the *Returns to education* treatment increases the probability of passing all the senior subjects by 5 percentage points (7 percent, statistically significant at the 5 percent level). The *Production function* arm has a small and non-significant impact. Conditional on the number of pending subjects the students had at the moment of the school visit, it can be observed that the positive impacts of the *Returns to education* are driven by those in good academic standing (zero pending). In Table C3 column 1, I include the interaction of the treatment arms with the level of confidence at baseline. Results indicate that the positive impacts of the *Returns to education* are driven by those underconfident students; they are 6 percent more likely to pass all the senior subjects with respect to the overconfident ones, although the difference is not statistically significant.

To analyze the performance of those with pending subjects, Table 7 shows the impact of the treatments on a dummy variable that indicates if the student has at least one pending subject left by the end of the academic year. Both treatments decrease the probability of having at least one pending subject: those students who receive the *Production function* arm are 7 percentage points less likely to have pending subjects (8 percent, significant at the 5 percent level) and those who receive the *Returns to education* are 12 percentage points less likely to have pending subjects (16 percent, significant at the 1 percent level). Splitting the analysis by the number of pending subjects they had at the moment of the visit, higher impacts are found for those students who are at the margin of being in good standing (column 2). In Table C3, I include the interaction by level of confidence: underconfident at baseline are those who respond most to both treatment arms. Column 2 shows that among those who receive the *Production function* arm, underconfident students are 34 percentage points less likely to have pending subjects left with respect to the overconfident ones (difference is statistically significant at the 5 percent level), and among those who receive the *Returns to education* underconfident are 20 percentage points less likely to have a pending subject left with respect to the overconfident ones (difference significant at the 10 percent level).

Perceptions of Labor Market Outcomes

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

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

Table 8 shows the impact of these misperceptions at baseline on graduation, considering the returns to two levels of education: completed secondary and completed college. I focus here on the students who received the *Returns to education* treatment. Both those who misperceived expected earnings (for completed secondary and completed college) in a negative way and those whose misperceptions were positive at baseline have positive magnitudes. The magnitude of the effects is higher for students with a positive misperception, although the difference in coefficients is not statistically significant.

When I provide information about the true returns to education, students weight their prior beliefs according to the new information, and they can subsequently decide which piece of information to assign a higher weight. Based on previous results in the literature, students with a negative misperception are expected to update their beliefs upward and graduation will increase. However, the aggregated result depends on the percentage of students who assign a higher weight to their prior beliefs versus the new information.

5.4 Heterogeneous Effects

Time Preferences

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

Following this argument, I consider the role of time preferences on timely graduation. By using a set of questions in the baseline questionnaire following a standard Becker DeGroot Marschak procedure (Bursztyn and Coffman, 2012), I computed the discount factor for each student. I then took the median and separated students based on whether they were above or below the median. Results are shown in Table 9. As expected, the effect in the *Returns to education* treatment arm is greater and statistically significant for students above the median. Although the difference with respect to students under the median value is not statistically significant, it shows that this is a relevant individual characteristic to consider when providing information like this to teenagers.

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

Socioeconomic Status and Gender

In the baseline questionnaire, I did not include a question about family income due to that question's low response rate in the pilot survey. To generate a proxy for economic status, I use an index constructed by using variables indicating the ownership of goods including air conditioning, heating, a washing machine, and a personal computer, whether the student's family lives in an overcrowded dwelling,[17] and whether at least one parent or guardian has some post-secondary education. If the index is less than or equal to 3, I classified the student as "poor" and otherwise, as "least poor".[18]

Table 10 shows that in the control group, students classified as poor have a lower graduation rate at 45 percent, which is 14 percentage points lower than the least poor students. In column 1, I demonstrate that contrary to previous findings (Jensen, 2010), less poor students are positively affected by both treatments: students in the *Production function* treatment arm are 8 percentage points more likely to graduate than the control group, and those in the *Returns to education* treatment arm are 14 percentage points more likely to graduate than the control group. Both results are statistically significant at the 5 percent level, and the

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

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

difference of the magnitudes is also statistically significant at the 5 percent level.

Table 10 also shows the impacts by gender. Columns 3 and 4 show that female students are more likely to graduate than male students in the control group. However, both information treatments have a positive impact on both genders, with higher impacts observed for male students. I observe positive results of both treatments for both genders, and the differences are not statistically significant.

5.5 Other outcomes

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

University enrollment

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

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

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

Table 11 column 1 shows that only 13 percent of the students in the control group are enrolled in university, and both treatment arms increase the probability of enrollment by

5 percentage points (almost 40 percent). These effects are statistically significant at the 10 percent level. The difference between treatments is not statistically significant. Bonilla-Mejía et al. (2019) present an experiment aimed to improve college enrollment in Colombia by providing information on returns to education for senior students and no effects were found. A potential explanation for my results is that the settings are different regarding access to post-secondary education: in Argentina there are no examination entrance exams for colleges, public post-secondary institutions are free, and in many districts public transportation for all students is free.

Formal Employment

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

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

Column 2 of Table 11 shows the results for both treatment arms. As expected, the level of formal employment for the control group is small; only 3 percent of the students in that group have a formal job at the considered time. However, both treatment arms generate a negative and statistically significant impact on formal employment. A potential, but not conclusive, explanation is that students' reservation wage increased after receiving the treatments.

One key caveat is that the sample size in this analysis is lower than the original sample because I did not find information for all students in the administrative data—there were errors in IDs in the data I received from the high schools. To test for potential issues of attrition, I created a dummy variable equal to 1 if a student was not found and 0 otherwise. Then I run the main specification and I do not find differences across treatment arms (see Table C4 in Appendix C).

[19] See Subsection 4.4.

6 Conclusions

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

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

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

The findings of this study have substantive policy importance: graduation rates can be improved in low-income settings using an inexpensive intervention that fills information gaps that are more likely to be present in low-income households. Small bureaucratic hurdles, which those with substantial parental or other forms of social support can easily negotiate, may trip up those without such resources. In these contexts, the provision of small pieces of information offers an excellent opportunity to improve graduation rates, as shown in this paper. Students who are positively affected by this intervention now have a previously unavailable chance to achieve economic mobility.

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Fryer, R. G. (2011). Financial Incentives and Student Achievement: Evidence from Randomized Trials. *The Quarterly Journal of Economics 126*(4), 1755–1798.

Fryer, R. G. (2016). Information, Non-Financial Incentives, and Student Achievement: Evidence from a Text Messaging Experiment. *Journal of Public Economics 144*, 109–121.

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

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

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

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

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

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

Loyalka, P., C. Liu, Y. Song, H. Yi, X. Huang, J. Wei, L. Zhang, Y. Shi, J. Chu, and S. Rozelle (2013). Can Information and Counseling Help Students from Poor Rural Areas Go to High School? Evidence from China. *Journal of Comparative Economics 41*(4), 1012–1025.

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

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

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

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

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

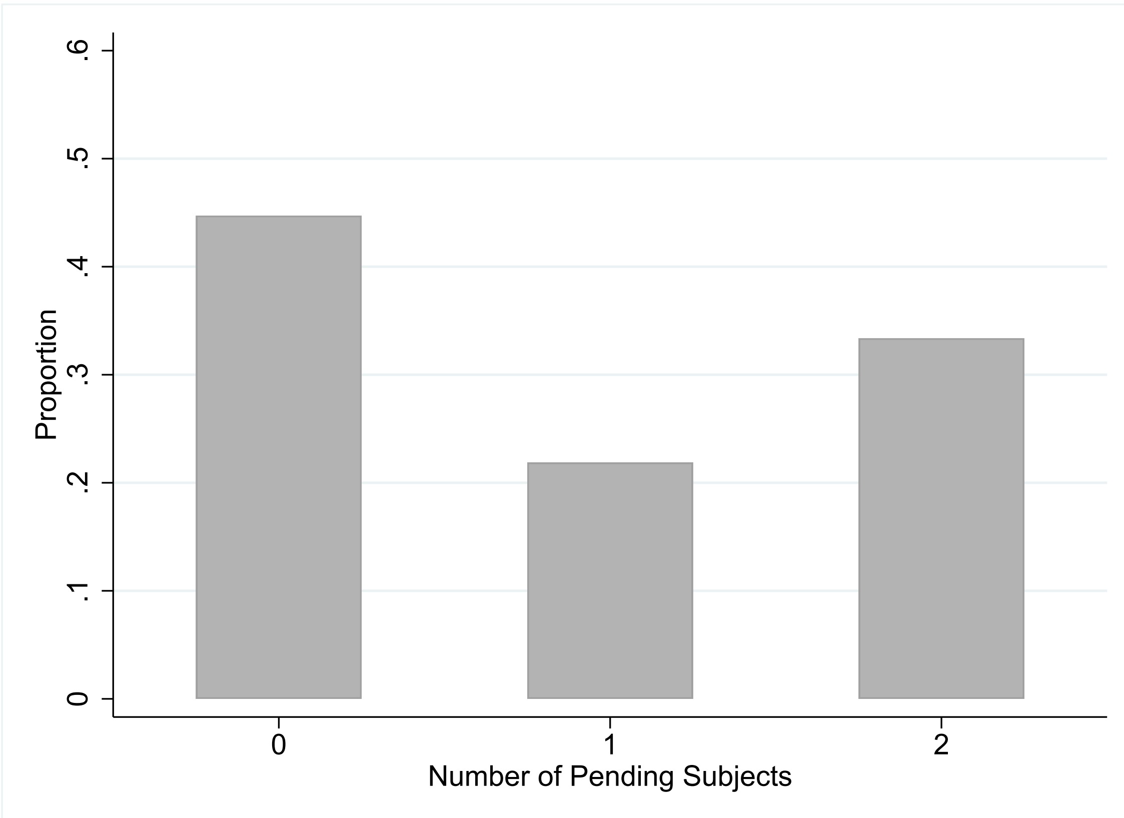
Soman, D. (2015). *The Last Mile: Creating Social and Economic Value from Behavioral Insights*. UNIV OF TORONTO PR.

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

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

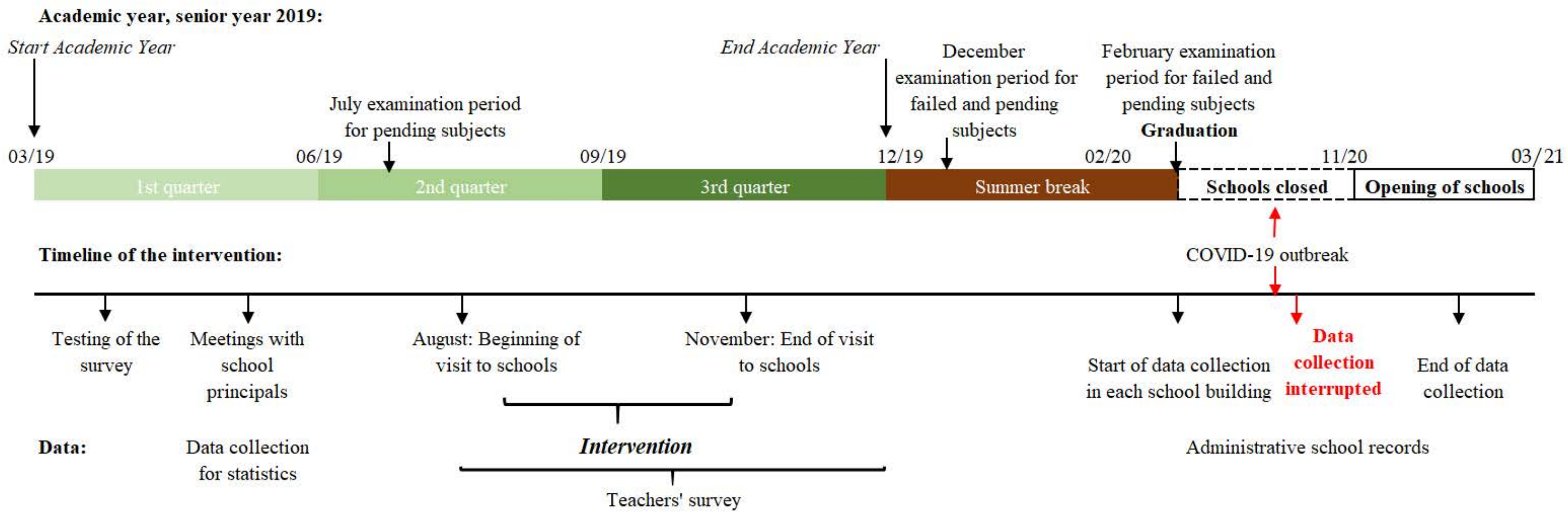
Figures

Figure 1: Senior Students and Pending Subjects, Control Group



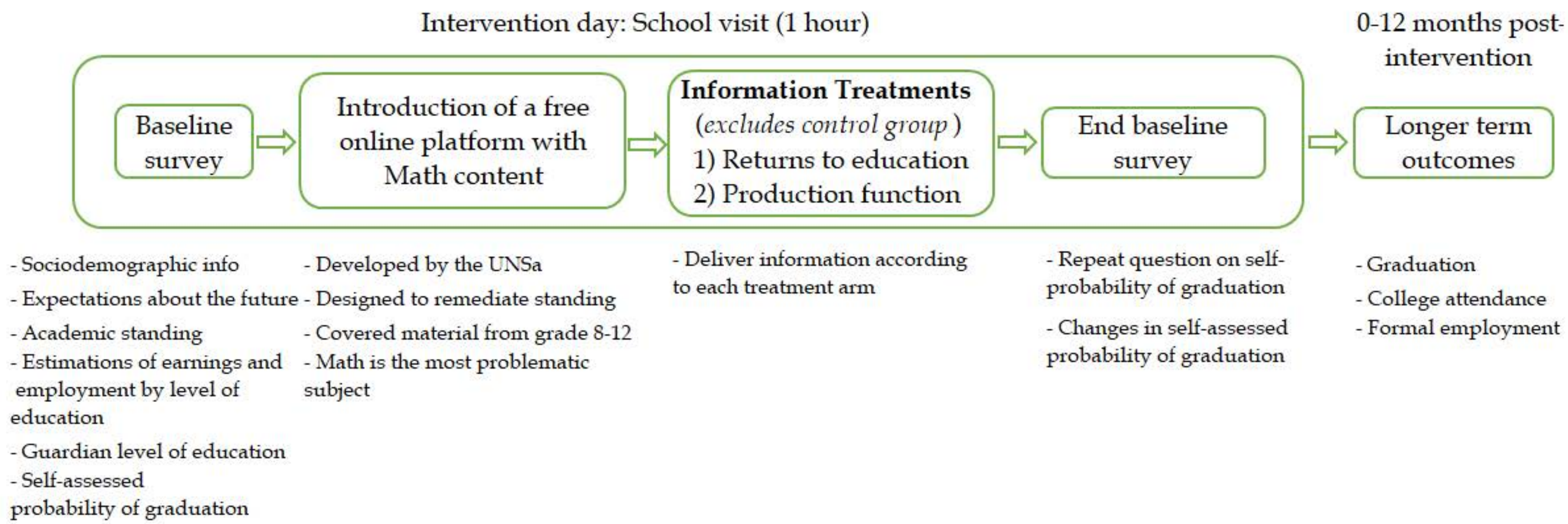
Notes: Number of pending subjects at the beginning of the senior year. Schools' administrative records.

Figure 2: Timeline, Intervention and Data Collection



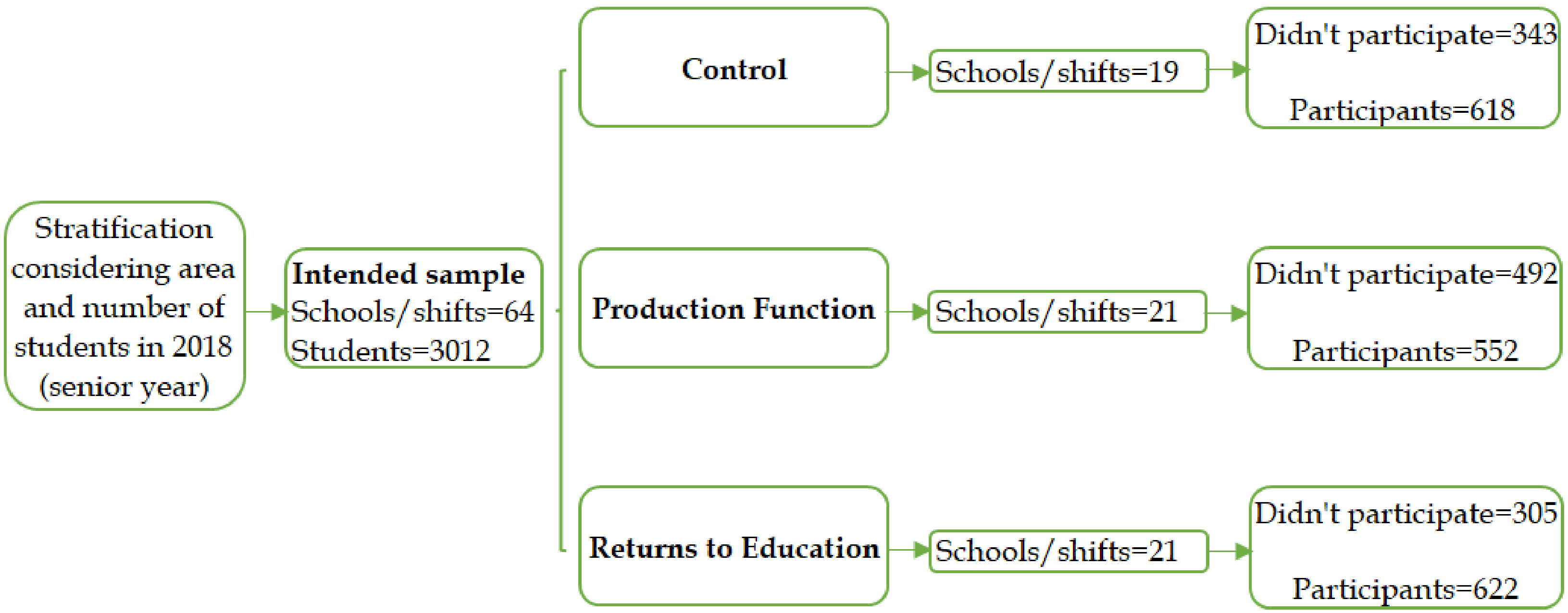
Notes: The intervention was designed for students who were seniors in 2019.

Figure 3: The Intervention Day



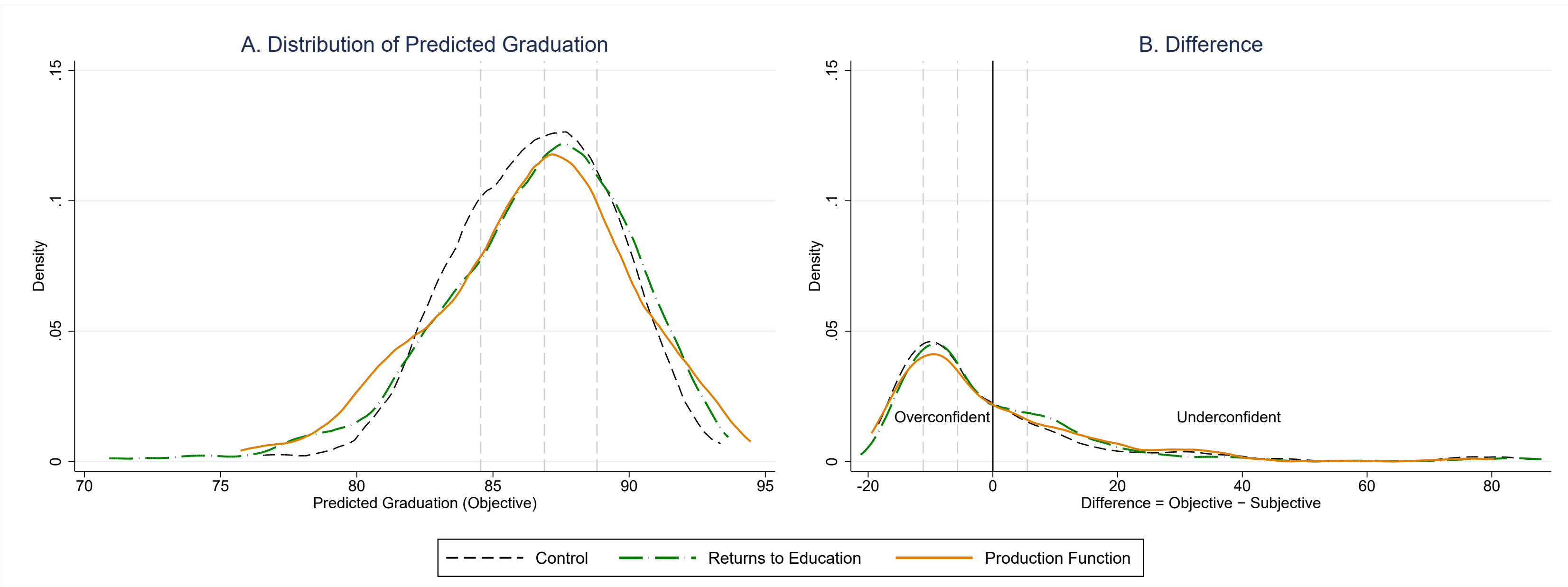
Notes: At the start of this intervention the questionnaire was tested in several rounds. Several corrections were made to improve students' understanding. The main change was related to the question used to ask probabilities of own graduation. Higher variability in responses was found using Figure C3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

Figure 4: Randomization Design and Sample



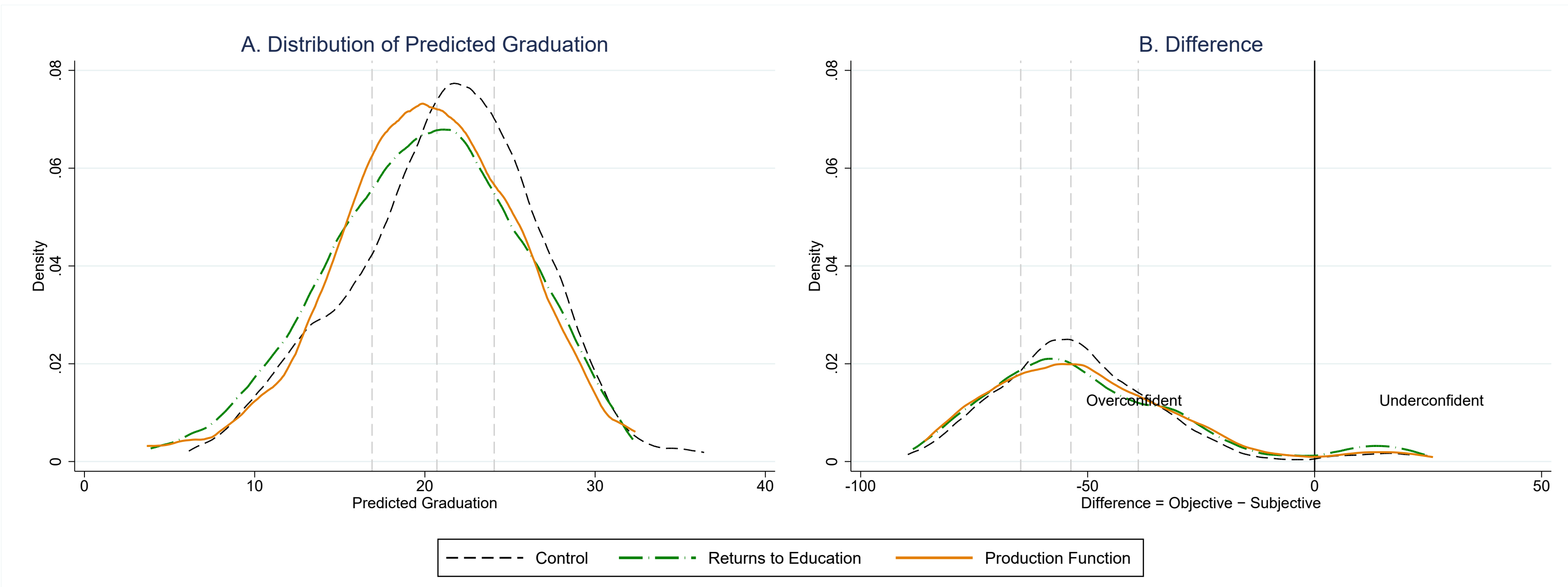
Notes: Baseline survey and schools' administrative records only for participant students.

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



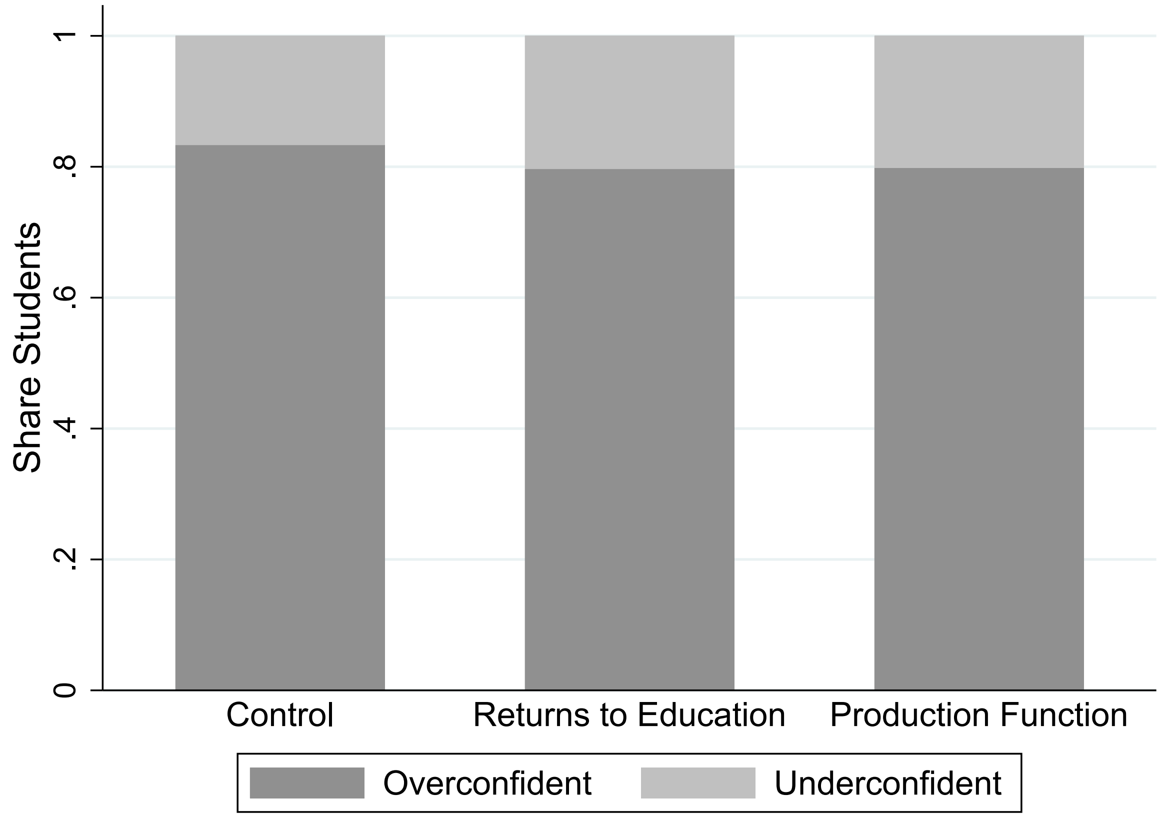
Notes: Kernel density estimates. Vertical dashed lines indicate 25th, 50th, and 75th percentiles of overall distribution, respectively.

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.

Figure 7: Overconfidence by Treatment Arm



Notes: Proportions of overconfident students computed according the classification shown in Figures 5 and 6.

Tables

Table 1: Descriptive Statistics from Control Group

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

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

Table 2: Randomization Verification

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

Notes: Column 1 reports the number of non-missing observations of variables among all students in the control group. *, **, and *** denote statistical significance at the 10, 5, and 1 percent levels, respectively.

Table 3: Impacts of Information on Graduation by Pending Subjects

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

Notes: Robust standard errors clustered at the school-shift level in parentheses. All regressions include graduation from the cohort 2018 at the school-shift level, and strata fixed effects. Eligible controls include area of the city dummies, student age, student gender, if the student has children or is pregnant, average grades of classes during the first 2 quarters of the senior year, if the student has a job or takes care of a family member dummy, if the student repeated at least one year in secondary school, if her/his parent/guardian has some superior education, if the student does not live in a crowded dwelling, if in the household there is a computer, a washing machine, an AC, heating, and pairwise interactions between all previously-listed students. Missing values are recoded to the sample mean and separately dummied out. These missing dummies are also used to construct pairwise interactions. *, **, and *** denote statistical significance at the 10, 5, and 1 percent levels respectively.

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

	(1)	(2)	(3)
		Graduation	
	All	Zero Pending	At least One Pending
Production Function × Overconfidence	0.0300	-0.0372	0.0630**
	(0.0287)	(0.0234)	(0.0276)
Production Function × Underconfidence	0.0820*	0.0184	0.262**
	(0.0450)	(0.0591)	(0.131)
Returns to Education × Overconfidence	0.0920***	0.0184	0.123***
	(0.0298)	(0.0260)	(0.0346)
Returns to Education × Underconfidence	0.115**	0.0786	0.182**
	(0.0461)	(0.0544)	(0.0836)
Overconfidence	-0.109**	0.0975**	0.155***
	(0.0478)	(0.0410)	(0.0579)
P-value: PF × Overconfident = PF × Underconfident	0.381	0.376	0.139
P-value: RE × Overconfident = RE × Underconfident	0.696	0.358	0.549
P-value: PF × Overconfident = RE × Overconfident	0.020**	0.025**	0.089*
P-value: PF × Underconfident = RE × Underconfident	0.406	0.301	0.579
Mean (Control, Underconfident)	0.61	0.72	0
N	1786	833	953

Notes: Robust standard errors clustered at the school-shift level in parentheses. All regressions include graduation from the cohort 2018 at the school-shift level, and strata fixed effects.. See notes in Table 3 for a list of potential controls. *, **, and *** denote statistical significance at the 10, 5, and 1 percent levels respectively.

Table 5: Impacts of Information on Pending Subjects (in December 2019) Behavior

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

Notes: Robust standard errors clustered at the school-shift level in parentheses. All regressions include graduation from the cohort 2018 at the school-shift level, and strata fixed effects. See notes in Table 3 for a list of potential controls. *, **, and *** denote statistical significance at the 10, 5, and 1 percent levels respectively.

Table 6: Impacts of Information on Performance — Senior Subjects

	(1)	(2)	(3)	(4)
	Passed all senior subjects			
	All	Zero pending July 2019	One pending July 2019	Two pending July 2019
Production Function	0.0127	-0.0266	0.0255	0.00137
	(0.0240)	(0.0236)	(0.0462)	(0.0487)
Returns to Education	0.0489**	0.0494**	-0.00392	0.0532
	(0.0218)	(0.0207)	(0.0520)	(0.0483)
P-value: PF = RE	0.152	0.000***	0.647	0.357
P-value: PF = RE = 0	0.074*	0.001***	0.848	0.513
Mean (Control)	0.65	0.86	0.55	0.38
N	1786	933	413	440

Notes: Robust standard errors clustered at the school-shift level in parentheses. All regressions include graduation from the cohort 2018 at the school-shift level, and strata fixed effects. See notes in Table 3 for a list of potential controls. *, **, and *** denote statistical significance at the 10, 5, and 1 percent levels respectively.

Table 7: Impacts of Information on Performance — Pending Subjects

	(1)	(2)	(3)
	At least one pending left		
	All	One pending July 2019	Two pending July 2019
Production Function	-0.0674**	-0.0523	-0.0122
	(0.0279)	(0.0442)	(0.0432)
Returns to Education	-0.124***	-0.166***	-0.0383
	(0.0323)	(0.0554)	(0.0340)
P-value: PF = RE	0.095*	0.029**	0.564
P-value: PF = RE = 0	0.000***	0.011**	0.528
Mean (Control)	0.79	0.66	0.90
N	853	413	440

Notes: Robust standard errors clustered at the school-shift level in parentheses. All regressions include graduation from the cohort 2018 at the school-shift level, and strata fixed effects. See notes in Table 3 for a list of potential controls. *, **, and *** denote statistical significance at the 10, 5, and 1 percent levels respectively.

Table 8: Impacts on Graduation by Perceptions on Expected Earnings by Level of Education

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

Notes: Robust standard errors clustered at the school-shift level in parentheses. All regressions include graduation from the 2018 cohort at the school-shift level, and strata fixed effects.. To compute the dummy variable Misperception (−) by level of education (level shown at the top of each column), I consider whether a student is accurate or is underestimating employment and earnings are being underestimated. See notes in Table 3 for a list of potential controls. *, **, and *** denote statistical significance at the 10, 5, and 1 percent levels respectively.

Table 9: Impacts on Graduation by Time Preferences

	(1) Graduation
Production Function × Above Median	0.0349
	(0.0364)
Production Function × Below Median	0.0394
	(0.0371)
Returns to Education × Above Median	0.117***
	(0.0347)
Returns to Education × Below Median	0.0438
	(0.0487)
Above Median Discount Factor	-0.0208
	(0.0402)
P-value: R × Very Patient = R × Not Very Patient	0.238
P-value: PF × Very Patient = PF × Not Very Patient	0.928
Mean (Control, Not Very Patient)	0.56
N	1562

Notes: Robust standard errors clustered at the school-shift level in parentheses. All regressions include graduation from the cohort 2018 at the school-shift level, and strata fixed effects.. To compute the dummy variable Above Median Discount Factor I classified the students under that category if the discount factor was higher than the median value of the variable discount factor today vs. one week . See notes in Table 3 for a list of potential controls. *, **, and *** denote statistical significance at the 10, 5, and 1 percent levels respectively.

Table 10: Impacts of Information on Graduation by Poverty Level and Gender

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

Notes: Robust standard errors clustered at the school-shift level in parentheses. All regressions include graduation from the cohort 2018 at the school-shift level, and strata fixed effects. To classify students as Poor or Less Poor I created an index variable that includes ownership of household items and a dummmy variable that indicates if at least one parent or guard has some college education. In total the index includes 6 dummy variables, if the score is lower or equal to 3 the student is classified as poor. See notes in Table 3 for a list of potential controls. *, **, and *** denote statistical significance at the 10, 5, and 1 percent levels respectively.

Table 11: Impacts of Information on Other Main Outcomes

	(1) College Enroll-ment	(2) Formal Employ-ment
Panel A. No Interactions		
Production Function	0.052*	-0.014*
	(0.027)	(0.0087)
Returns to Education	0.054**	-0.022***
	(0.024)	(0.0076)
P-value: PF = RE	0.909	0.227
P-value: PF = RE = 0	0.059*	0.012**
Mean (Control)	0.13	0.032
Panel B. Interactions with Students' Confidence		
Production Function × Overconfidence	0.035	-0.0080
	(0.027)	(0.010)
Production Function × Underconfidence	0.092*	-0.040**
	(0.049)	(0.016)
Returns to Education × Overconfidence	0.047*	-0.026***
	(0.024)	(0.0088)
Returns to Education × Underconfidence	0.074	-0.0086
	(0.046)	(0.022)
Overconfidence	0.024	-0.00091
	(0.033)	(0.018)
P-value: PF × Overconfident = PF × Underconfident	0.160	0.098*
P-value: RE × Overconfident = RE × Underconfident	0.556	0.485
P-value: PF × Overconfident = RE × Overconfident	0.606	0.021**
P-value: PF × Underconfident = RE × Underconfident	0.637	0.064*
Mean (Control, Underconfident)	0.13	0.035
N	1786	1348

Notes: Robust standard errors clustered at the school-shift level in parentheses. All regressions include graduation from the cohort 2018 at the school-shift level, and strata fixed effects. College is a dummy variable equal to 1 that indicates if the student is formally enrolled in at least one college of Salta during 2020 (Universidad Nacional de Salta and Universidad Catolica de Salta). Formal employment is a dummy variable equal to one if the student was employed in the formal sector at least one month during the last quarter of 2020 and the first quarter of 2021. See notes in Table 3 for a list of potential controls. *, **, and *** denote statistical significance at the 10, 5, and 1 percent levels respectively.

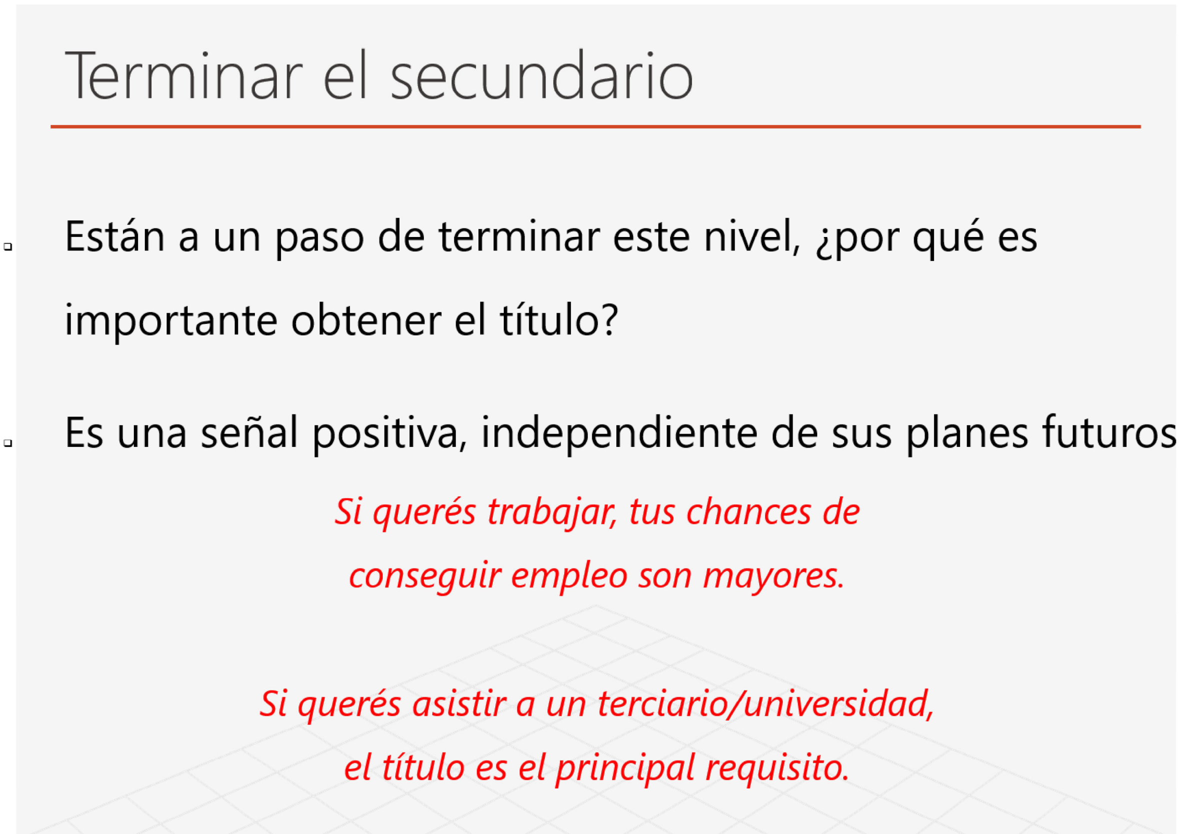
A Appendix: Information Treatment Arms

Information Interventions

I show the specific content introduced to the senior students that participated in each treatment arm. For both treatment arms, I discussed why it is important to finish high school, highlighting the fact that they already spent almost 5 years attending this level and that only a small fraction of the students that enter their senior year drop out at some point during the year (Anuarios Estadísticos, Ministerio de Educación de la Nación). See Figure A1.

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

Figure A1: Why to Obtain the Diploma



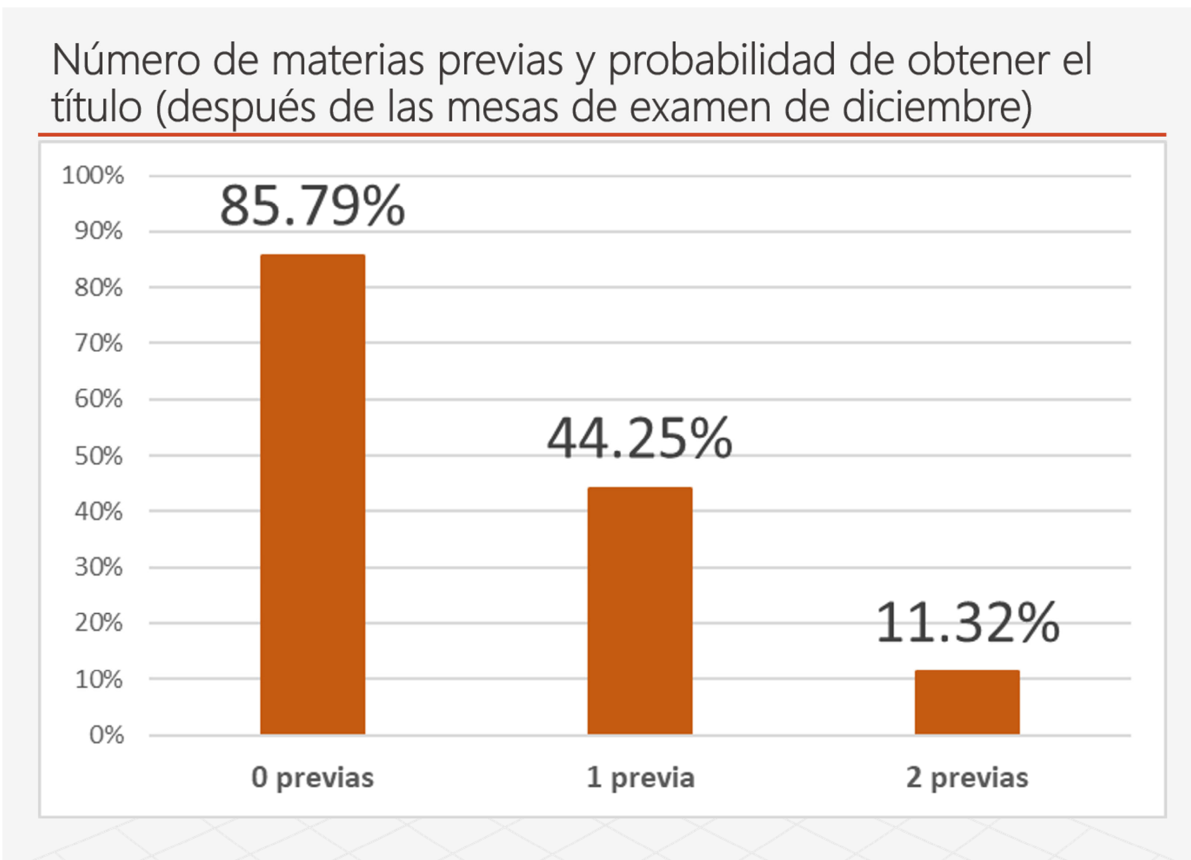
Notes: Common slide showed to all the students who received any of the intervention treatments. Translation: Finish high school. You are a step away from finishing this level of education, why is it important to get a diploma? It is a positive signal that does not depend on your future plans: if you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

Production Function

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

Based on the sample I collected, I elaborated the statistics that were shown to the students (see Figure A2). Each student was aware of their own situation, but during the presentation, I could not observe their academic standing (number of pending subjects). The idea of showing these numbers was to help them create a mapping of their situation at the beginning of the senior year and how similar students performed in terms of graduation. Given that this could have been shocking news for students regardless of standing, I talked about the intermediate steps needed to transform inputs into outputs and I discussed how to remedy their situation: first, I opened a discussion of the options together (Figure A3), and then I showed a summary of the most relevant tips to effectively obtain a diploma on time.

Figure A2: Statistics Shown to the Students

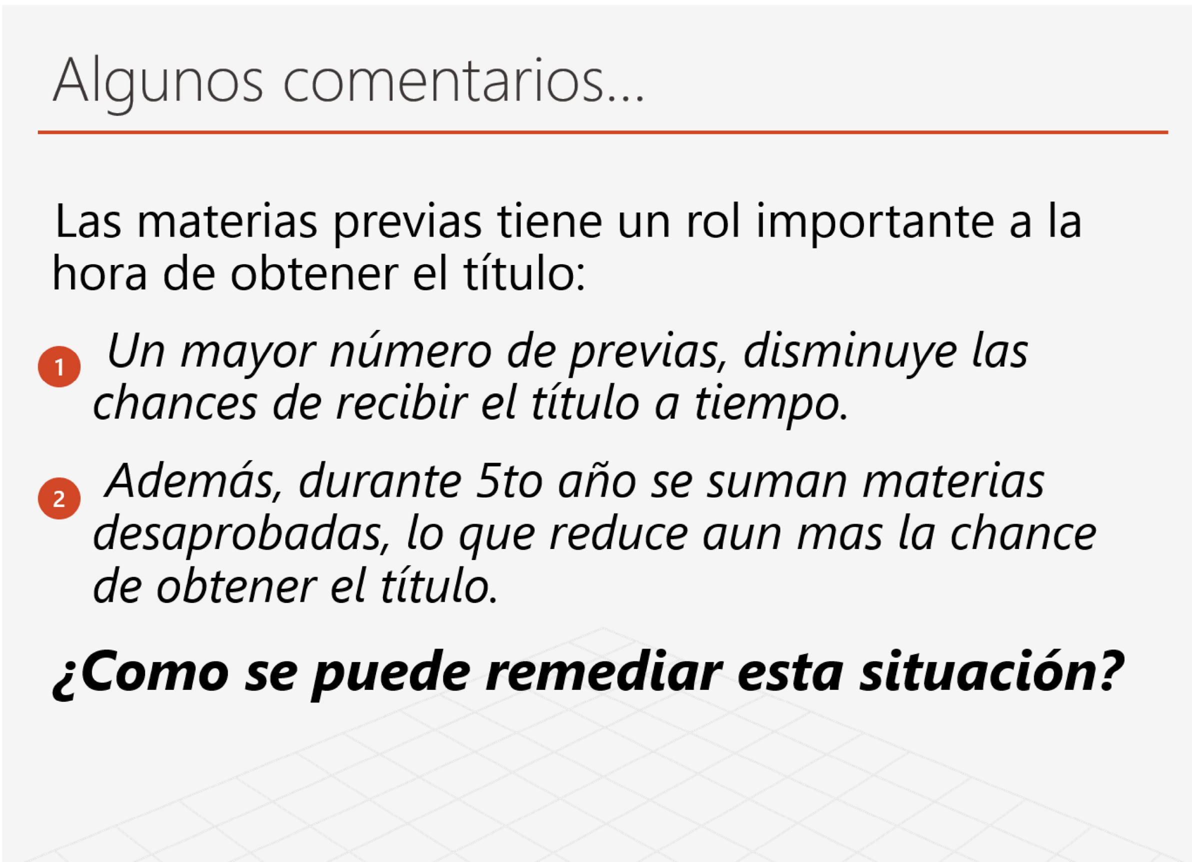


Notes: Own estimations based on a sample of representative schools in the capital city of Salta including students who were seniors during the 2018 academic year.

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

Students' senior year includes several social activities (prom night, private parties, graduation trip, etc.). In interviews with the school principals and in some focus groups with students from the previous cohorts, these activities were mentioned as major distractions from academics.

Figure A3: The Role of Pending Subjects

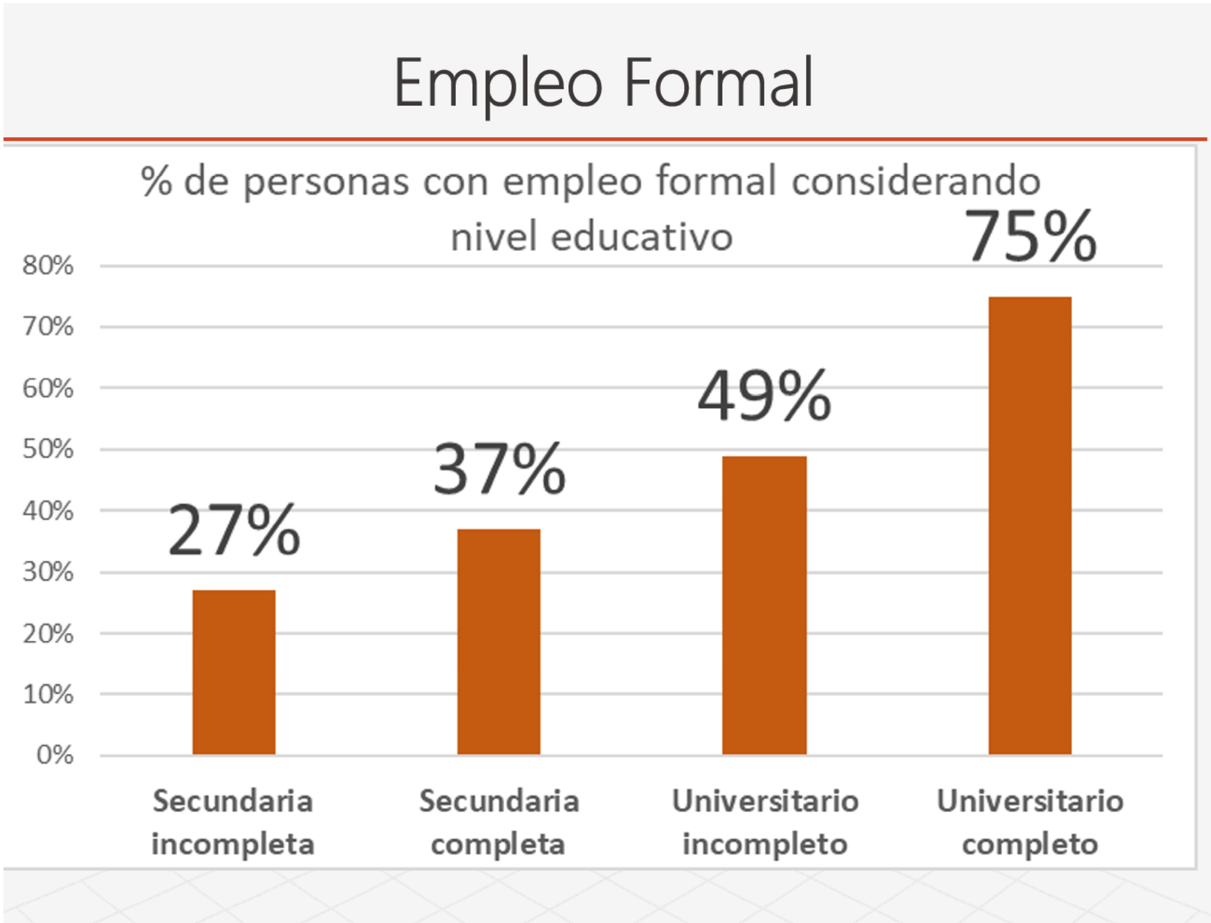


Notes: In this part of the presentation, I highlighted the role of the pending subjects and passing senior year subjects in timely graduation. Then I opened the discussion with a question, "How can this situation be remedied?"

Returns to Education

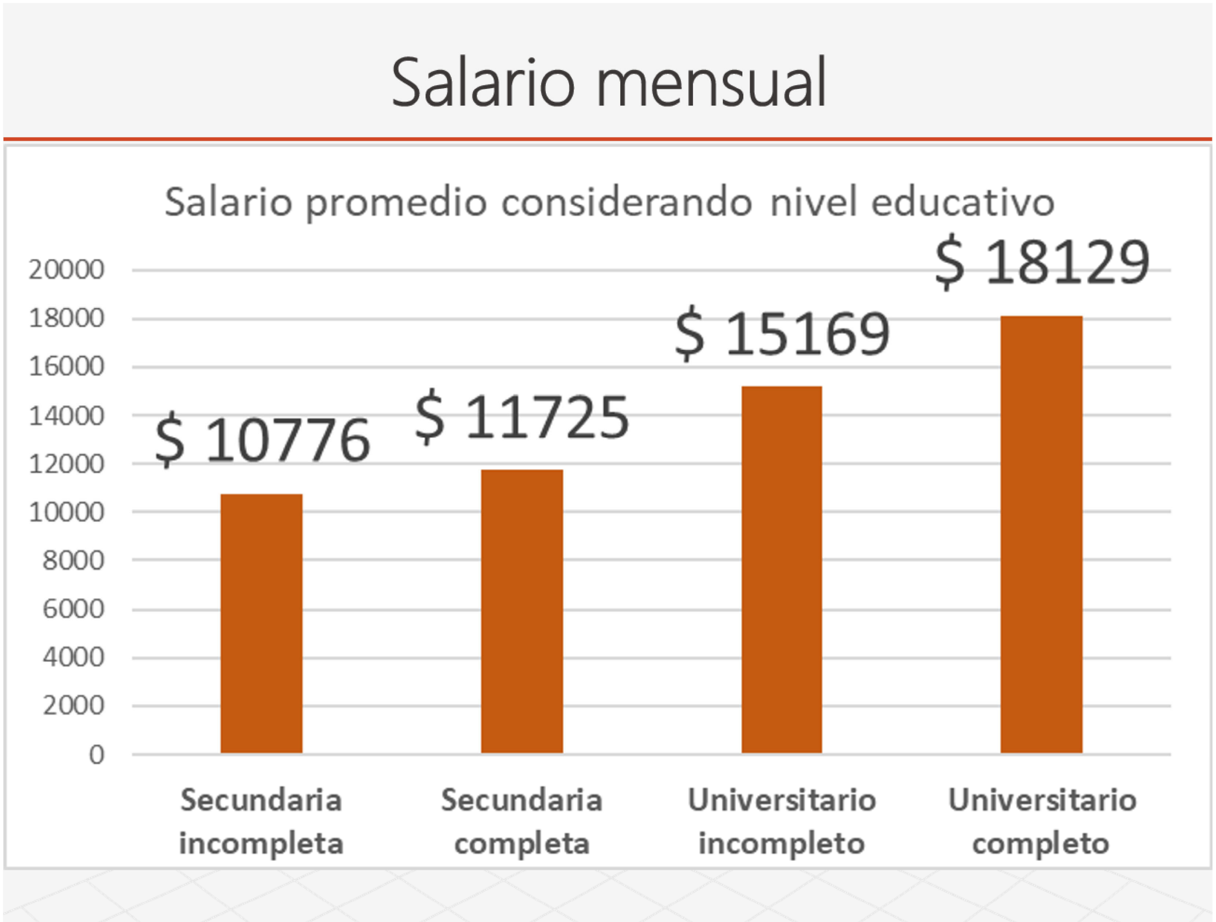
In this presentation I used data from the National Household Survey 2018 (Encuesta Permanente de Hogares) to compute the averages of formal employment and earnings to be shown to the students. I only considered individuals from the province of Salta, between 18 and 30 years old. The statistics were computed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

Figure A4: Formal Employment by Level of Education



Notes: Own estimations based on Encuesta Permanente de Hogares, 2018 (this survey only covers urban areas). Mincer equation was estimated considering age, gender, and marital status.

Figure A5: Monthly Wages by Level of Education



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

Reminders

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

Returns to Education Reminders

- SMS

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

  Team UNSa-Brown

- e-mail

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

  Team UNSa-Brown

Production Function Reminders

- SMS

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

  Team UNSa-Brown

- e-mail

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

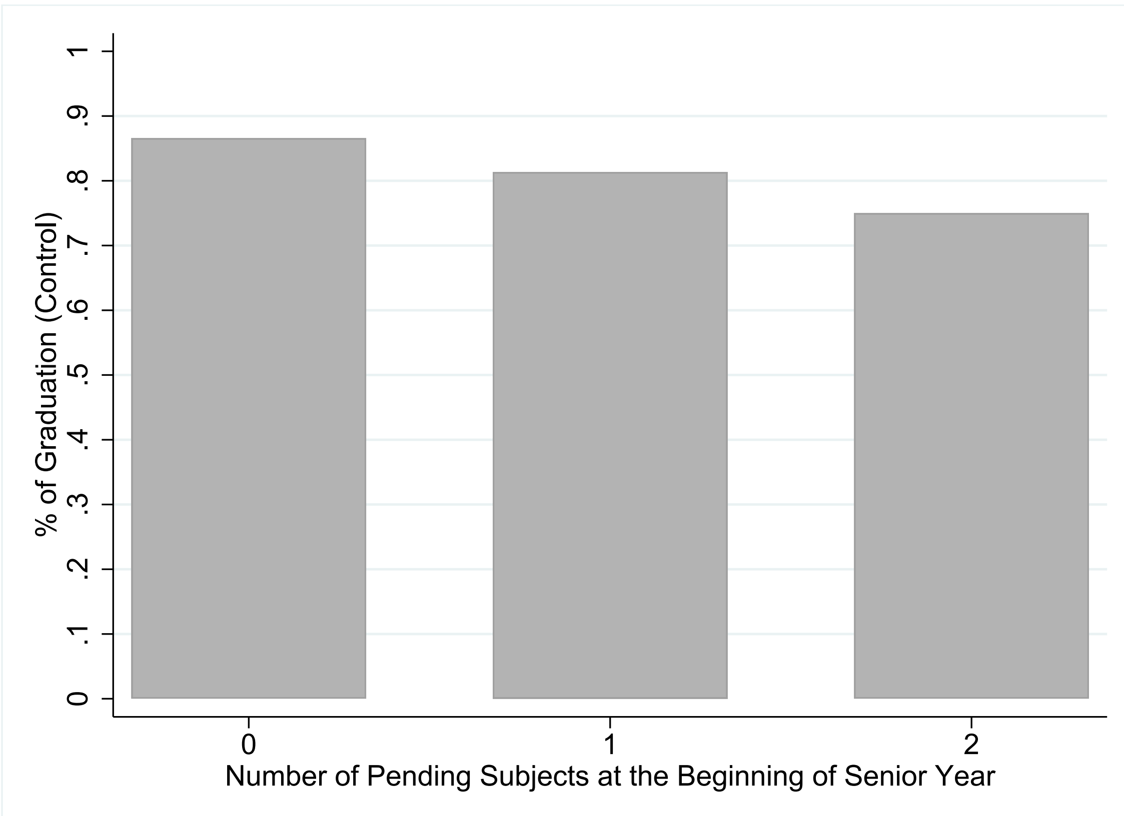
  Team UNSa-Brown

Discussion about the Production Function

A potential concern on the design of the *Production function* treatment is that it could make students believe that moving from two to zero pending subjects will increase their probability of graduation by 74 percentage points (Figure A2). In this context, deception will be present if passing the subjects is not enough to graduate, but passing those subjects is one requirement besides passing the senior subjects.

I use the control group to observe changes in the probability of graduation, considering the subset of students who had pending subjects but passed them by the end of the academic year. Figure A6 shows the graduation conditional on the number of pending subjects the students had at the beginning of the senior year. This subset of students passed their pending subjects and now moved to the “good standing bin.”

Figure A6: Graduation of Students who Passed All their Pending Subjects by the End of Academic Year. Control Group



Notes: Students moved to the bin of 0 pending subjects, but still could have failed senior subjects.

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

B Appendix

B.A Statistical Power

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

B.B Free Online Platform: MOODLE

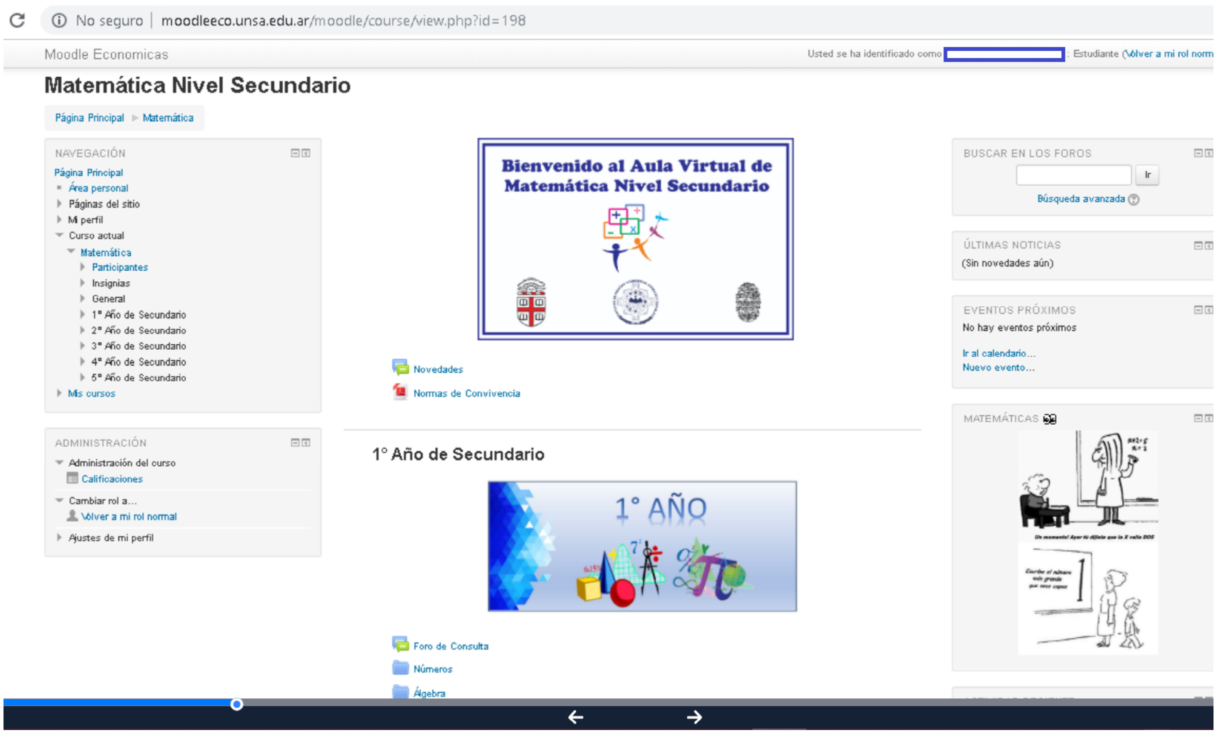
The Directorate of Secondary Education of Salta required that I provide some useful information to all students; otherwise, I would encounter resistance from school principals reluctant to give me access to their schools. So, to provide something in exchange for their participation, I designed a free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

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

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

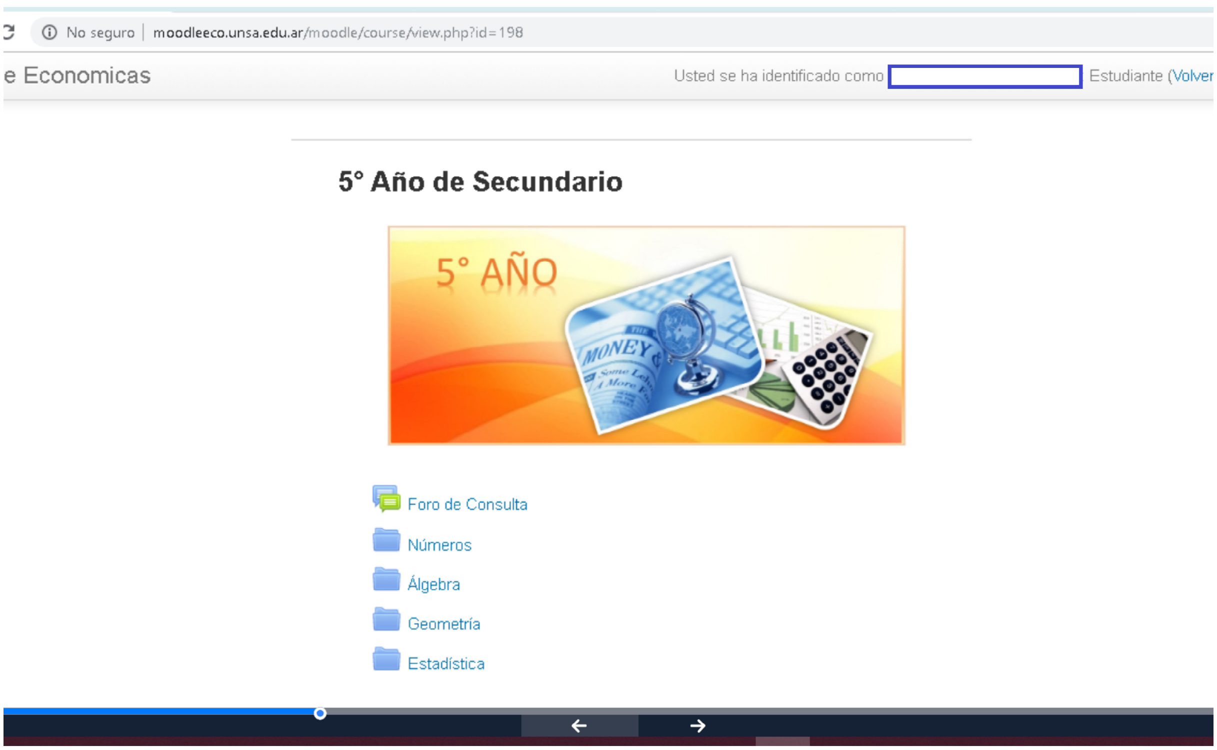
We also showed how to post questions (public or private) with the commitment on our side to reply to each question within 48 hours. Students were allowed to upload pictures for assistance with exercises involving mathematical notation.

Figure B1: MOODLE Platform: Homepage



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

Figure B2: MOODLE Platform: Senior year overview



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

Figure B3: MOODLE Platform: Senior year specific content

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

$\boldsymbol{B}$.C Full Derivatives: Model with Uncertainty

The maximization problem the student faces is:

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

with FOC:

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

Proof. Production Function

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

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

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

Proof. Returns to Education

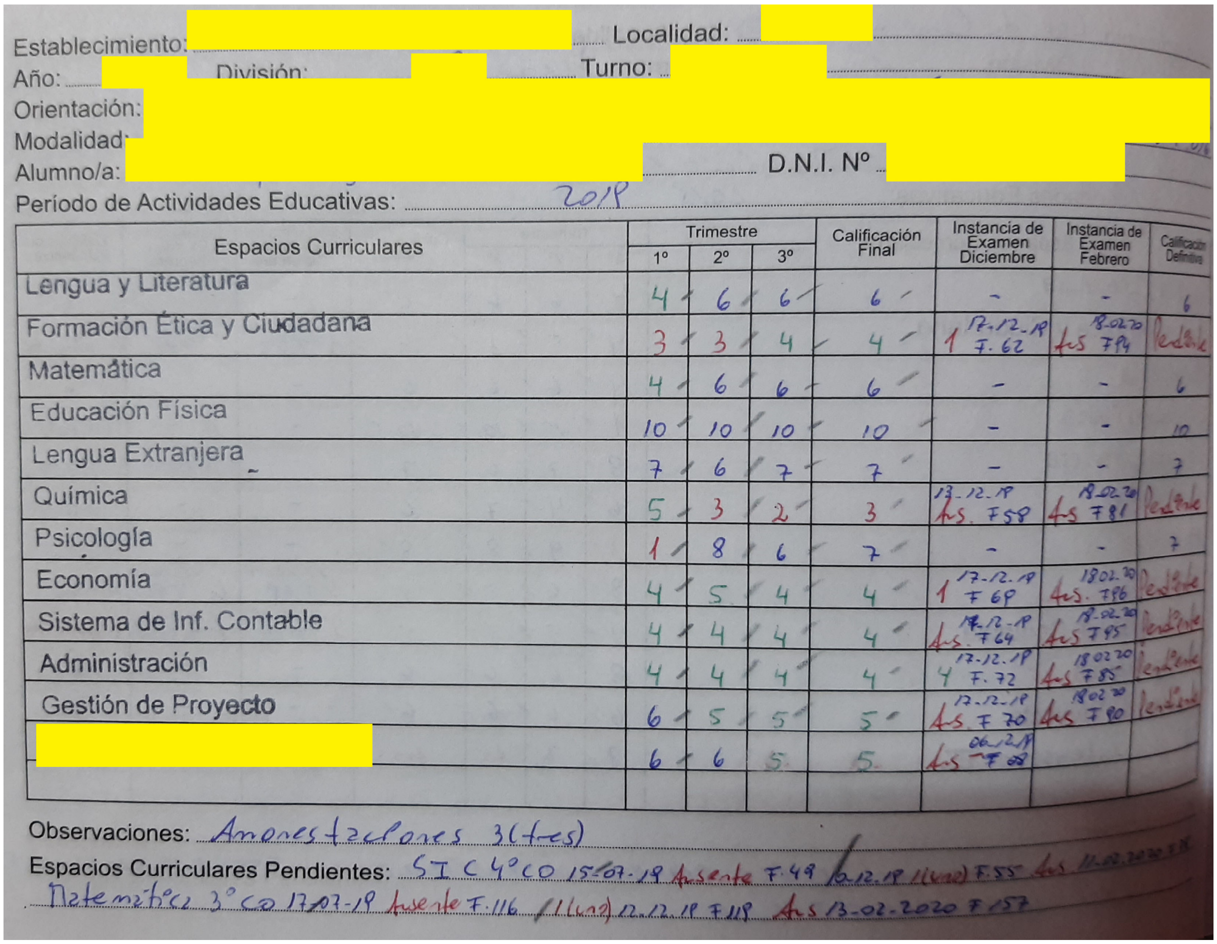
$$\begin{aligned}\hat{p}g'\left(\hat{\beta}_l e+\hat{\alpha_h}\right)\hat{\beta}_l+(1-\hat{p})\,g'\left(\hat{\beta}_h e+\hat{\alpha_l}\right)\hat{\beta}_h+\\ \hat{p}g''\left(\hat{\beta}_l e+\hat{\alpha_h}\right)\left(\hat{\beta}_l\right)^2\frac{de^*}{d\hat{V}}+(1-\hat{p})\,g''\left(\hat{\beta}_h e+\hat{\alpha_l}\right)\left(\hat{\beta}_h\right)^2\frac{de^*}{d\ha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

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

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

C Appendix: Supplementary Figures and Tables

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

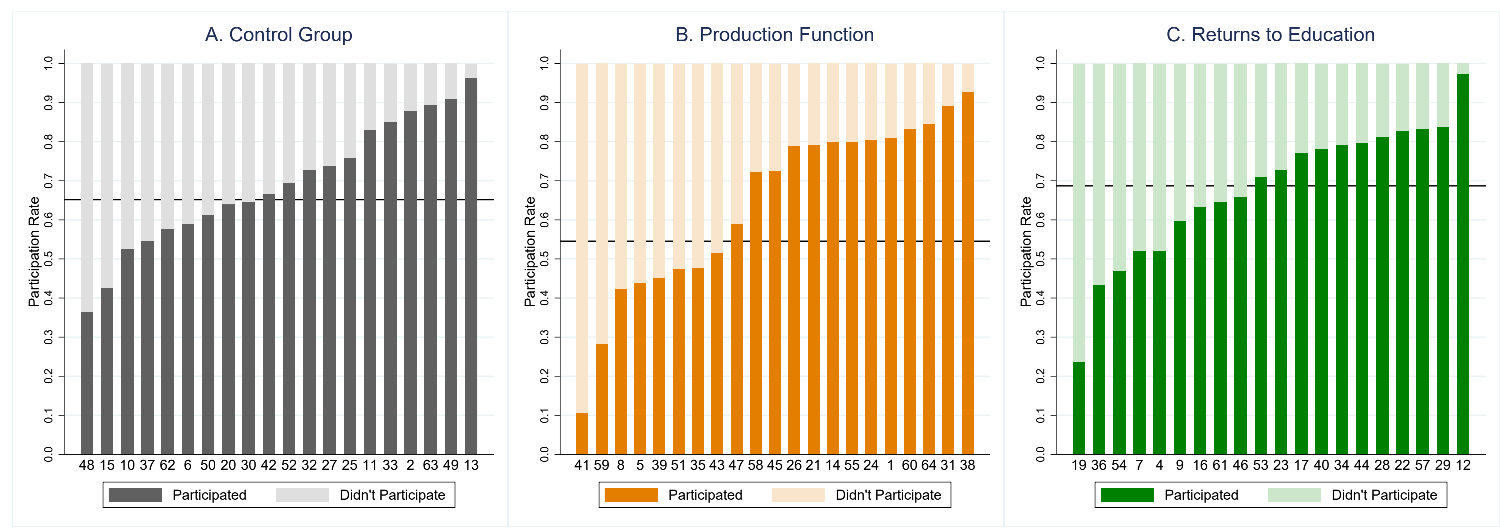


Establecimiento: [redacted] Localidad: [redacted]
Año: [redacted] División: [redacted] Turno: [redacted]
Orientación: [redacted]
Modalidad: [redacted]
Alumno/a: [redacted] D.N.I. Nº [redacted]
Período de Actividades Educativas: 2019

Espacios Curriculares	Trimestre 1º	Trimestre 2º	Trimestre 3º	Calificación Final	Instancia de Examen Diciembre	Instancia de Examen Febrero	Calificación Definitiva
Lengua y Literatura	4	6	6	6	-	-	6
Formación Ética y Ciudadana	3	3	4	4	1 17-12-19 F. 62	18-02-20 Aus F84	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	13-12-19 Aus. F58	18-02-20 Aus F81	Pendiente
Psicología	1	8	6	7	-	-	7
Economía	4	5	4	4	1 17-12-19 F 68	18-02-20 Aus. F86	Pendiente
Sistema de Inf. Contable	4	4	4	4	17-12-19 Aus. F64	18-02-20 Aus F85	Pendiente
Administración	4	4	4	4	4 17-12-19 F. 72	18-02-20 Aus F85	Pendiente
Gestión de Proyecto	6	5	5	5	17-12-19 Aus. F 70	18-02-20 Aus F 90	Pendiente
[redacted]	6	6	5	5	06-12-19 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.118 Aus 13-02-2020 F 157

Figure C2: Participation Rates at the School Level

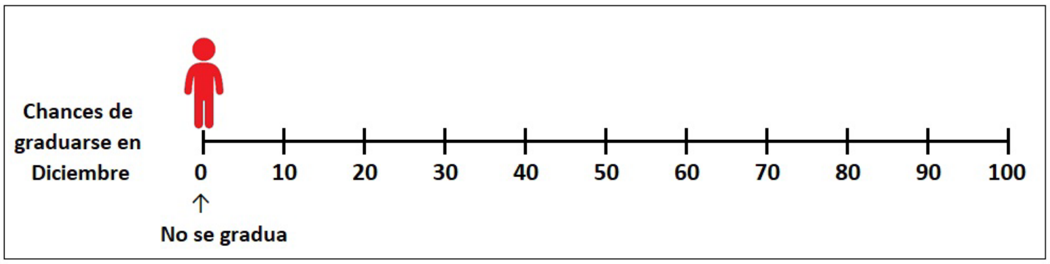


Notes: Horizontal axis shows random numbers assigned to each school. In each panel, the horizontal black lines indicates the participation rate for the entire treatment arm.

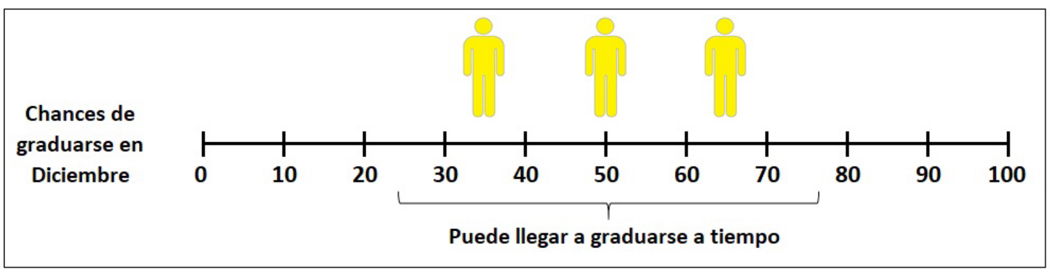
Figure C3: Prompts used to ask own probability of graduation

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

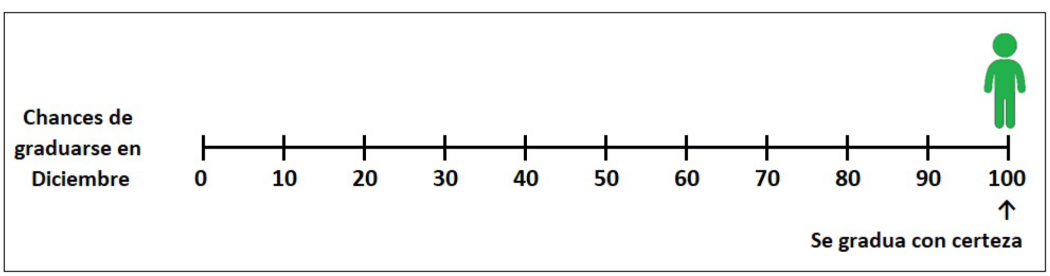
Example 1: A student who does not study, frequently skips classes. Has pending subjects and does not attend the exam periods, who does not pass all the subjects this year, has a 0% probability of receiving the diploma in December.



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



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



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

Notes: First, a concept of probability was shown, and then I asked about their perceptions of their own probabilities of graduation.

Table C1: Impacts of Information on Graduation by Pending Subjects

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

Notes: Robust standard errors clustered at the school-shift level in parentheses. All regressions include graduation from the cohort 2018 at the school-shift level, and strata fixed effects.. See notes in Table 3 for a list of potential controls. *, **, and *** denote statistical significance at the 10, 5, and 1 percent levels respectively.

Table C2: Impacts of Information on Self-estimated Probability of Graduation (after-before intervention)

	(1)	(2)	(3)
		Difference	
	All	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: PF = RE	0.004***	0.075*	0.265
P-value: PF = RE = 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 strata fixed effects. See notes in Table 3 for a list of potential controls. *, **, and *** denote statistical significance at the 10, 5, and 1 percent levels respectively.

Table C3: Impacts of Information on Performance

	(1) Passed all senior subjects	(2) At least one pending subject left
Production Function × Overconfidence	-0.00550	-0.0513*
	(0.0303)	(0.0290)
Production Function × Underconfidence	0.0502	-0.393***
	(0.0505)	(0.138)
Returns to Education × Overconfidence	0.0346	-0.114***
	(0.0241)	(0.0357)
Returns to Education × Underconfidence	0.0926**	-0.317***
	(0.0463)	(0.0996)
Overconfidence	0.00277	-0.210***
	(0.0377)	(0.0550)
P-value: PF × Overconfident = PF × Underconfident	0.378	0.017**
P-value: RE × Overconfident = RE × Underconfident	0.257	0.078*
P-value: PF × Overconfident = RE × Overconfident	0.183	0.090*
P-value: PF × Underconfident = RE × Underconfident	0.405	0.620
Mean (Control, Underconfident)	0.64	1
N	1786	853

Notes: Robust standard errors clustered at the school-shift level in parentheses. All regressions include graduation from the cohort 2018 at the school-shift level, and strata fixed effects. See notes in Table 3 for a list of potential controls. *, **, and *** denote statistical significance at the 10, 5, and 1 percent levels respectively.

Table C4: Difference by Missing Employment Data

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

Notes: Robust standard errors clustered at the school-shift level in parentheses. All regressions include graduation from the cohort 2018 at the school-shift level, and strata fixed effects. See notes in Table 3 for a list of potential controls. *, **, and *** denote statistical significance at the 10, 5, and 1 percent levels respectively.