

Scholarship policy evaluation. The case of *Progresar* scholarships in Argentina

Sebastian Miguez

MSc in Economics, Universitat de Barcelona

Advisor: Dr. Daniel Montolio

Universitat de Barcelona

Abstract

Education can be considered as a motor for increasing productivity, subsequent growth and development. Moreover, it can also be considered the key to lower socioeconomic gaps in the society. Inequality in Latin America is a concern, with an average GINI coefficient higher than developed countries. Even though Argentina exhibits one of the lowest coefficients in the region, economical and especially educational inequality is something to improve. Therefore, public programs that aim to lower those gaps and inequality in every sense of the word are substantial for these countries. Considering this, we will analyze the *Progresar* scholarship program in Argentina in the period between 2018 and 2020 where the program muted from a conditional transfer to a merit-based scholarship. The unique administrative dataset used by the office in the Ministry of Education is a novel data for research, and even with some flaws that are considered in this work, can provide policy makers with empirical studies to make better funded decisions. We use a logit regression to measure the impact of the scholarship on the dropout of starter students for a specific university. Results suggest that the scholarship increases the odds of continuity in a career in the university and that female students have a higher chance than male students to continue in their first year. We also use an OLS estimation to test the impact of the scholarship on first-years' performance. Results suggest that there is a positive impact of the scholarship though not consistently neither of considerable size. (JEL: C21, I22, I24, I28, I38)

Keywords: Education, Scholarship, Need-based grant, Policy evaluation, University, *Progresar* scholarships, Argentina.



UNIVERSITAT DE BARCELONA
SCHOOL OF ECONOMICS

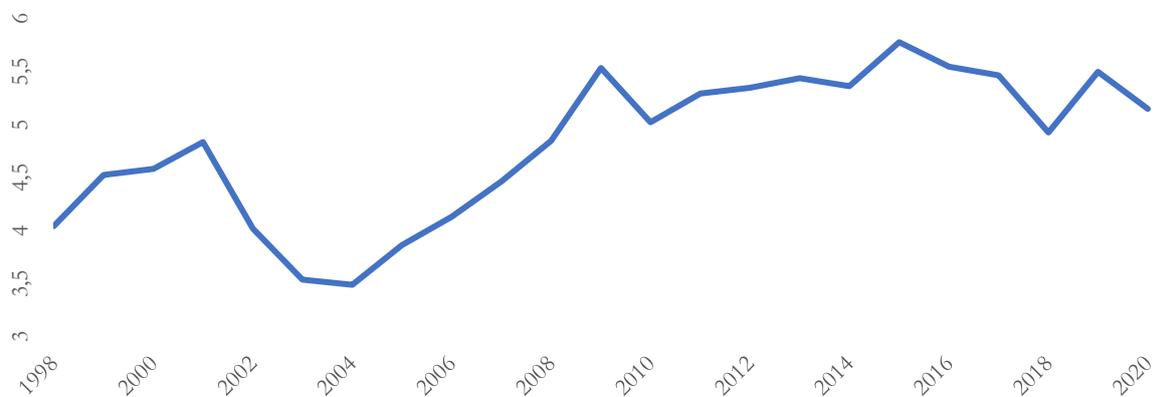
Index

1. Introduction	3
2. <i>Becas Progresar</i>: A Need-Based Scholarships Program	7
2.1. Process of giving the scholarship	12
3. Data and Variables.....	13
3.1. Micro Database Construction.....	13
3.2. Variables and Descriptive Statistics.....	17
4. Methodology and Empirical Model.....	20
4.1. Empirical Model Logit.....	21
4.2. Empirical Model OLS.....	22
5. Results.....	22
5.1. Logistic Model Results.....	23
5.2. OLS Results.....	25
5.3. Robustness Checks.....	27
6. Conclusions and Policy Implications.....	34
7. Bibliography.....	36

1. Introduction

A vast economic literature considers education to be a motor for increasing productivity, subsequent growth and development (Krueger & Lindahl, 2000; Lucas, 1988). Moreover, it is considered the key variable to lower socioeconomical gaps between members of the society and hence a strong driving force for a more equal and just economic system (Sylwester, 2002; De Gregorio & Lee, 2002). According to Jensen's (1983) definition of educational equality, equal educational opportunities will be achieved only when students who have entered the institutions of their choice are able to continue their educational career until they reached the personal limits of their capabilities and motivations. In that line, governments all over the world spend a significant part of their budget in education, both for infrastructure and transfers. The major chunk of expenditure in Argentina goes to educational staff as well as sundry transfers. Even though most of the expenditures labeled as transfers in the budget of the Argentinian Ministry of Education is destined to public universities and provinces (which are in charge of the educational system until higher education), a significant part goes to scholarships, mainly those included in the *Progresar* program¹.

Figure 1 - Education expenditure as a % of total government expenditure



Source: self-elaboration based on UNESCO and Ministry of Economy for values of 2019 and 2020

Considering government expenditure in education, Argentina is no exception with an average level of expenditure over GDP of 5,34% for the last decade². Evolution of educational expenditure as a percentage of the total government expenses is detailed in Figure 1. For Argentina to return to a path of growth requires not only capital accumulation, innovation, and job creation, but also

¹ *Progresar* represents a 5,11 percentage of the total expenditure of the Ministry, but if we exclude the transfers to universities (budget that is not administrated directly by the Ministry) the percentage of *Progresar* ascends to 17. *Source: self-elaboration using data from the Ministry of Education and the Ministry of Economy in 2019.*

² Calculated as expenditure in education over total GDP. *Source: UNESCO and Ministry of Economy for values of 2019 and 2020 (according to the law of budget of each year).*

increasing productivity. In this line, reforms and programs that efficiently boost human capital are central for long-term growth and shared prosperity. Therefore, an analysis of the programs in place is a must.

The aim of this paper is to provide a first insight into a unique administrative data set of scholarships in an Argentinian university. The unique dataset with detailed student information will be analyzed with two main objectives: i) understand the type of administrative information available to Argentinian education institutions to assess if it can be used to evaluate public scholarship programs and ii) to present a first attempt on the impact that the scholarship *Progresar* has on dropout rates and student performance. Both objectives will help laying out a foundation for future research that should allow policymakers to make evidence-based decisions.

Dropout rates are considerable in university level, especially for students in their first year of the degree, and even higher for individuals from the low-income quartiles of the population (Fanelli, 2017). Policies such as *Progresar* in a National level or scholarships and mentoring programs in a university level are implemented to lower the dropout rates but empirical evidence on the impact of those programs is not available (Celada, 2020). Therefore, a first approach to study the impact of *Progresar* on first-year students with a unique database is of significant importance, even though it might present some flaws considered in this paper.

Public policies in Argentina - and *Progresar* is no exemption – are volatile in the sense that the reforms of one government are prone to be rescinded by their successors. The program was launched in 2014 as a conditional transfer system aimed to continue the Universal Child Allowance (AUH - *Asignación Universal por Hijo*) and covers children up until they come out of age at eighteen. The AUH scholarship required an up-to-date vaccination plan as well as enrollment in school. The admission requirements of the *Progresar* transfer were likewise based on education, demanding individuals themselves (not their parents as in AUH) to be enrolled in higher education or in technical professional training (e.g., electrician, bakery, etc.) in order to receive 80% of the transfer. The remaining 20% was distributed at the end of each year in which the respective individual certifies that he/she keeps studying³.

In 2018, under Mauricio Macri's presidency, the government attempted to redesign the program into a based-on merit and performance scholarship. Therefore, an academic requirement for receiving the grant paired with an increase in the benefit amount (and even a more significant increase for students from careers considered strategic) was introduced. To analyze *Progresar* new

³ For the cases of beneficiaries of basic education, technical professional training scholarships and new students of higher education, the program still works with that 80-20 condition.

design it is substantial to highlight the difference between merit and performance-based grants, since unlike the former, performance-based scholarships are paid to students based on their academic performance in the current term, regardless of what happened in previous terms (Patel et al., 2013). This connects to recent controversial debates in literature with some scholars suggesting that merit and performance-based grants provide a positive impact on students (Richburg-Hayes et. al., 2009; Angrist, Lang and Oreopoulos, 2009; Angrist, Oreopoulos and Williams, 2014). However, according to Montalbán (2019) it remains unclear to what extent grants tied to academic requirements are more effective than those without. Several studies provide empirical support for the claim that stronger monetary incentives tend to lead to higher levels of effort, but the effect of monetary compensation on performance does not seem to be monotonic (Gneezy and Rustichini, 2000). Finally, there is suggestive evidence that merit aid has the greatest adverse effect on income for the bottom percentile of earners and benefits the top quartile of earners. Overall, findings imply that merit-based grants may exacerbate wealth inequality by crowding out minorities and low-income students from in-state campuses, and in some cases, diverting resources away from government assistance programs (Young, 2018).

Besides the changes in design, the administration of the program was shifted from the National Social Security Agency (ANSES – *Administración Nacional de la Seguridad Social*) to the Ministry of Education. Furthermore, a process of communication between the national government - through the mentioned Ministry - and the universities and other higher education institutes was initiated to exchange information upon students. This process took circa a year to become efficient so that the databases used in this paper is based on the 2018-2020 period. Unfortunately, the period investigated in this dataset is unlikely to be extendable in the future since the mentioned volatility and inconsistency of Argentinian politics has once again reoccurred in that the presidency of Alberto Fernández removed the academic requirement in an attempt to bring back the original setup in which a far-reaching coverage was the main objective of the policy⁴. This change means that the current design of the scholarships can only be analyzed in students starting a career since for them the academic requirement does not *apply*⁵, making them the objective population of this paper.

The objective of this paper is, therefore, to study the impact of the scholarship on students accessing university that received the grant over the years 2018-2020. We need to emphasize that

⁴ To learn more about the new normative about the redesign of the scholarship check Terms and Conditions, art. 20 and 21 of the program in <https://www.argentina.gob.ar/educacion/progresar>.

⁵ The academic requirement applies to every applicant of the scholarship but for starting students this requirement is equal to 0, since the 50% of the study plan so far is null.

both programs (AUH and *Progresar*) target the low-income segment of the population since they both require that the applicant's household income does not surpasses a certain threshold⁶. Hence, this study is particularly relevant since most of the literature has focused on merit-based grants with certain academic requirements which typically do not target low-income students (Montalbán, 2019).

The paper is premised on the database obtained from the Ministry of Education, ANSES, and the National University of the North-East. The data stems directly from the Ministry's selection process and is not published which is why the raw data is not always consistent. In order to maintain the integrity of the study these inconsistencies were not included in the sample. The micro dataset is still worth analyzing albeit its methodological and data issues, as a result the aim of this study is not to present suggestive results on the impact of the scholarship, and to encourage policymakers, universities, and researchers to study the educational policies and to give them a starting point for future studies and decision-making processes in the form of first preliminary results based on the data currently available. If more educational institutions provide much needed data and an exhaustive research approach is set up by the Ministry of Education, the preliminary results of this paper can be refined and then become more precise, more robust, more conclusive, and determinant in the efficient assignation of the educational budget.

Using a logit model, we analyze the impact of the scholarship on the dropout decision in a subset of students using data from the National University of the North-East. In this work we will refer to dropout decision as the non-continuity of a student in the university considered. Then, we analyze the impact on academic performance for those students that received the scholarship compared with similar students that did not obtain it. Therefore, we present a primer on the effects of the scholarships on students.

The rest of the paper is going to be divided into 6 sections. In section 2 the scholarship program is explained since it is complex to understand how the assignation process works and how the databases from different sources interact with one another. In section 3 the variables selection is explained and a first exploratory analysis with descriptive statistics is provided. Sections 4 and 5 describe the methodology, the model used, and the results. Finally, section 6 concludes and comments on policy implications and future research steps.

⁶ In the years studied in this paper the limit was always to have a household income lower than 3 minimal wages. In 2019 the monthly minimal wage was \$12.500 (314 USD), therefore the sum of 3 of them (3SMVM) was \$37.500 (942 USD).

2. *Becas Progresar*. A Need-Based Scholarships Program

The Government of Argentina has engaged in core reforms to transform and modernize the economy, while aiming to improve fiscal management and advance social protection. Between 2004 and 2008, Argentina was the regional champion in reducing poverty, but deteriorating fiscal conditions limited further improvements. By 2011-2015, wide macroeconomic imbalances, paired with economic distortions, led to high inflation, low investment, and lack of access to external financial markets. During 2016 and 2017, the government launched reforms that aimed to substantially change the country's economic structure, which included the unification of the exchange rate, the modernization of the import regime, cuts in government spending, debt agreements with foreign creditors, and a reform of the national statistics system. Moreover, social programs were revisited to improve targeting and efficiency while increasing coverage. The simultaneous and gradual adjustments on these multiple fronts posed pressures to growth and led to a recession in 2016, followed by a slow recovery in 2017. (World Bank, 2019)

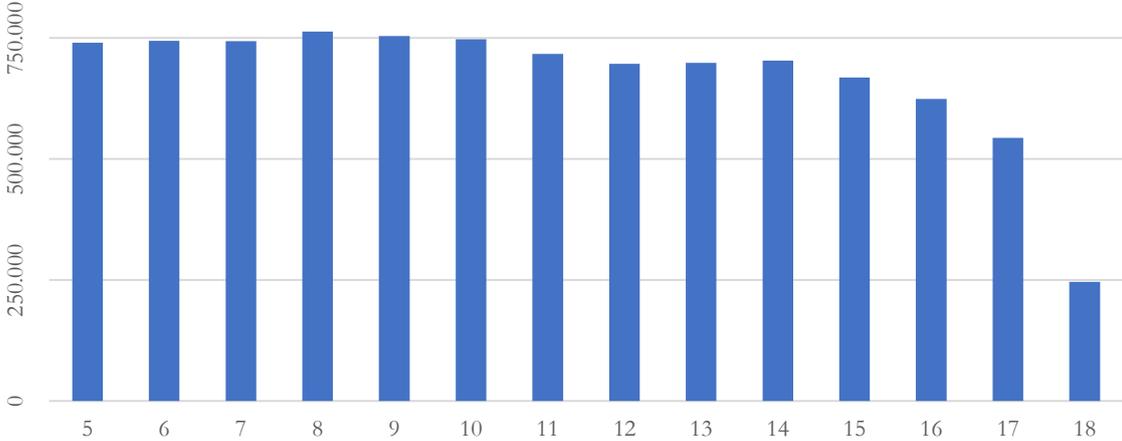
A federal education system shapes education policy in Argentina: provinces are responsible for the financing and management of initial, primary, secondary, and tertiary non-university levels, while the federal government is in charge of tertiary university level education and formulates general education guidelines and norms as well as it provides financial and technical assistance to the provinces to support certain policies, reforms and compensatory programs. In 2016-2017, 11.389.209 students were enrolled in the *Educación Común* modality (Initial, Primary, Secondary and Tertiary levels), 1.232.470 in adult education, and 1.939.419 in universities. Primary school, compulsory for all children, comprises 6 to 7 grades depending on the jurisdictions⁷. Since 2006, secondary education is compulsory and lasts 5 to 6 years according to the length of primary school. Secondary school is divided into lower (3 years, *Ciclo Básico*) and upper secondary (2 or 3 years depending on the jurisdiction, *Ciclo Orientado*). Basic education comprises primary and secondary levels. The tertiary non-university level is composed by institutes for teacher training (*Institutos de Formación Docente*), and technical-professional institutions. By 2015, there were 2.239 institutions in the tertiary non-university level (47% of which were public), with 59% offering teacher training courses. Argentina's public tertiary education is free, and institutions grant unrestricted admission to high school graduates. Public universities are also characterized by a high level of institutional autonomy. In terms of quality assurance, the well-established National Commission for University

⁷ In the provinces of Buenos Aires, Formosa, Tucumán, Catamarca, San Juan, San Luis, Córdoba, Corrientes, Entre Ríos, La Pampa, Chubut and Tierra del Fuego primary education comprises 6 years while in Ciudad de Buenos Aires, Río Negro, Neuquén, Santa Cruz, Mendoza, Santa Fe, La Rioja, Santiago del Estero, Chaco, Misiones, Salta and Jujuy it spans 7 years.

Evaluation and Accreditation (CONEAU - *Comisión Nacional de Evaluación y Acreditación Universitaria*) is responsible for external evaluation and accreditation of programs and institutions at the university level.

A significant share of the youth drops out of school without basic skills. As the provinces provide sufficient secondary school facilities and employ enough teachers, dropouts are caused by demand-side factors and not by supply constraints. Although access to secondary education is almost universal, the enrollment rate starts to fall significantly after 15 years of age (c.f. Figure 2)—as the opportunity cost of staying in school are growing— especially for boys and low-income students. Low-income students are thrown to the labor market sooner due to the necessities they face, postponing or dropping completely education. By this same age, the learning deficits are deep with more than two-thirds of youth not being able to solve simple math problems, and a half being unable to interpret basic texts (World Bank, 2019). Around half of the students that start secondary education do not complete it, and only one third of the poorest do so, leaving the educational system lacking key competences. One in five youth is neither employed, nor in education, or training, and one in ten studies and works simultaneously (National Institute of Statistics – INDEC).

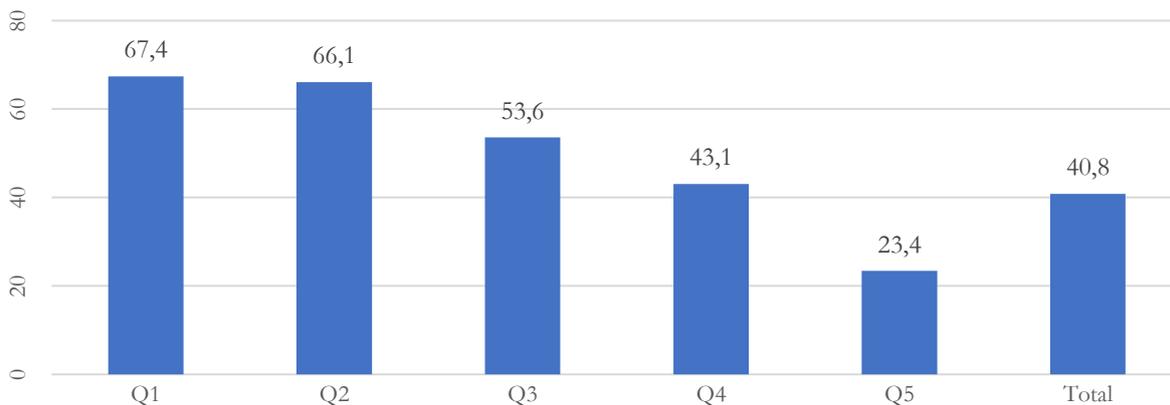
Figure 2 - Distubution of secondary school students per age



Source: self-elaboration based on Ministry of Education

When it comes to higher education, dropout is a major concern. According to data from the Secretary of University Policies (SPU – *Secretaria de Políticas Universitarias*) of 2011, the dropout rate in the public university level was approximately 80%. If we consider the interannual dropout rate calculated as in (Oloriz, Fernandez, Batto, 2020), the mean rate in 48 public institutions is 19%, with a maximum of 36% and a minimum of 11%. Looking at Figure 3, it is easy to tell that the dropout of low-income students between 18 and 30 years old is higher in university level.

Figure 3 - Dropout rates in university level per household income quintile for individuals between 18 and 30 years



Source: self-elaboration based on pool data from Permanent House Survey (EPH) 2003-2014

Even though scholarships (such as *Progresar*) and mentoring programs has been implemented since then to reduce the rate, there are still no policy evaluation empirical studies on the topic (Celada, 2020). This outlines the relevance of our work to test the impacts of the biggest scholarship program in Argentina.

To alleviate the problem the government implemented *Progresar* scholarships, which provide financial support to students between 18 and 24 years of age from the poorest socioeconomic households to complete basic education, and to access and complete higher education. The program includes additional financial incentives for strategic careers and a merit-based component for top performers.

Progresar was established in 2014 as subsidy transfer to foster access and completion at all education levels for students aged 18-24 that belonged to the poorest households. By 2017, the program reached 1 million beneficiaries, but monitoring of eligibility and attendance was carried out poorly.

In 2018, the program was transferred to the Ministry of Education and was modified from a transfer to a scholarship program: *Progresar* Scholarships. The main objectives of this change were to align benefits with education outcomes, to promote more inclusion and merit-based benefits, and to close the gap between education and labor markets by promoting strategic careers in higher education. The program offers four types of scholarships (Figure 4) that focus on i) students completing primary or secondary education; and ii) students accessing and completing higher education.

Figure 4 – Detailed age requirements and institutions for *Progresar* scholarships

	<i>Progresar</i> Basic Education	<i>Progresar</i> Technical professional training	<i>Progresar</i> Tertiary non-university	<i>Progresar</i> Tertiary University
Age	18-24	18-24	18-30	18-30
Education level	Primary and secondary	Technical	Higher	Higher
Institution of enrollment	Adult education schools	Technical training schools	Teacher training institutes, non-university institutes	Public universities

Source: self-elaboration based on Ministry of Education.

The scholarships target vulnerable students which fulfill income and age eligibility requirements. To align incentives and education outcomes, the scholarship was redesigned to include five distinct instruments: i) increased monitoring and evaluation of eligibility requirements: ensuring compliance with conditions of eligibility through administrative data and the certification of regular attendance at the end of the school year. Because of this *cleaning* process, beneficiaries decreased to 606.000, of which 43 percent correspond to youth completing primary and secondary education; ii) differentiated and higher benefits: changes in the amounts of the scholarships by education level and strategic careers; iii) stronger academic requirements: from 2 subjects passed per year to 50 percent of the theoretical study plan passed (3 subjects, on average)⁸; iv) permanence and completion incentives: benefits increasing in trajectory, and eligibility up to 30 years of age for advanced higher education students; v) merit-based component: beneficiaries that pass all the theoretical study plan with a minimum academic average of 8/10 receive twice the scholarship benefits⁹.

Point two (ii-differentiated and higher benefits: changes in the amounts of the scholarships by education level and strategic careers) is an interesting new approach for the scholarship program. With this inclusion, now students receive a higher grant if they are progressing in their studies, and they have a plus in case their career selection is a strategic one. Strategic careers are decided by the Ministry of Education considering the vacancy in the labor market of Argentina (e.g., engineering, nursing, basic sciences, etc.). Regarding the promotion and increase in the number of nurses, the

⁸ As previously mentioned, this requirement was removed in 2021 with the goal of increasing the coverage of the program.

⁹ The award based on merit was only paid in 2019 to beneficiaries of 2018. It is important to outline that the award was not paid to everyone that met the requirements, but to those that applied for the award and met the requirements.

Ministry of Education has decided to create a scholarship program – the National Nursing Formation Program (PRONAFE - *Programa Nacional de Formación de Enfermería*) in 2016. The *Progresar* scholarships absorbed this program but kept the old eligibility requirements for people who were already receiving the scholarship¹⁰. In this study we considered this aspect since the age difference might bias the results. Table 1 explains how grants are distributed based on the year the student is and if the career is considered strategic or not. It is observable that the monthly amount granted to students in dollars is not significantly high with respect to the minimal wage. Value in real terms dropped over the years mainly due to inflation and the lack of an update according to this phenomenon. Therefore, we should consider this when it comes to the analysis of our work. Scholars suggest that adding monetary transfers based on performance or merit do not have an impact on academic returns when an initial grant is already awarded and the requirements are weak (Montalbán, 2019). But the impact of an initial grant that deteriorates its value over the years has not been studied previously and would be of major relevance for developing countries with high macro economical instabilities. Given the lack of administrative data this will not be covered in this paper.

Table 1 - Higher education scholarships grants based on academic advance

Year	Tertiary non-university				Tertiary University			
	non-strategic		strategic		non-strategic		strategic	
	2018 *	2019 **	2018	2019	2018	2019	2018	2019
1st ***	\$ 1.600		\$ 1.800		\$ 1.600		\$ 1.800	
	USD 79	USD 40	USD 89	USD 45	USD 79	USD 40	USD 89	USD 45
2nd	\$ 1.600		\$ 2.000		\$ 1.600		\$ 2.200	
	USD 79	USD 40	USD 99	USD 50	USD 79	USD 40	USD 109	USD 55
3rd	\$ 1.900		\$ 2.600		\$ 1.900		\$ 2.900	
	USD 94	USD 48	USD 129	USD 65	USD 94	USD 48	USD 144	USD 73
4th	\$ 1.900		.		\$ 1.900		\$ 3.800	
	USD 94	USD 48	.	.	USD 94	USD 48	USD 189	USD 95
5th	.		.		\$ 2.300		\$ 4.900	
	USD 114	USD 58	USD 243	USD 123

Table 1 reports the grants awarded to beneficiaries of the program depending on how advanced they are in their career. Grants are different between Tertiary non-university and Tertiary university levels, and strategic and non-strategic careers. The amount is informed in Argentinian pesos (\$) and USD below.

* For 2018 the exchange rate considered is \$20,15 according to the National Bank of Argentina for the 1st of march of 2018

** For 2019 the exchange rate considered is \$39,80 according to the National Bank of Argentina for the 1st of march of 2019

*** Starting students receive 80% of the scholarship at the start of the academic year and the extra 20% if they continue by the end of it.

¹⁰ Individuals studying nursing that already receive the Pronafe scholarship prior 2018 did not have an age limit to receive the scholarship with the objective to not interrupt their studies.

2.1. Process of giving the scholarship

The Ministry's decision-making process involves checking the socioeconomic and academic conditions of the candidate. First, the scholarship is not automatically granted by the Ministry, which means that individuals need to apply via the Ministry web page¹¹ to be considered for the grant. In order to receive the benefit, the applicants' household income has to be below the sum of 3 minimal wages, they must be between 18 and 30 years old¹², and Argentinian¹³. Income and personal data of the applicants are not gathered by the Ministry itself, but by the National Social Security Agency (ANSES).

After analyzing the data, a vast majority of the applicants have a registered income equal to zero which can be due for two reasons. In the case the information is simply not available, the National Administration of Social Security marks such cases as flawed applications and rejects these candidates unless they adjust their status, therefore rejecting the application for bureaucratic reasons. So, the first reason is that incomes are generated from informal labor and hence not registered – a phenomenon common to many South American economies. Although such households could potentially earn more than what is found in the records, the database should not be heavily confounded as the vast majority of these applicants are indeed from the poorest classes of society where informality is higher. The second reason is that some individuals are registered as independent from their family income and this might confound our data, since individuals that are not completely independent could be registered with an income equal to zero. For instance, the database for 2020 presents the inconvenience that income is not reported for individuals. Being that the case, all of them present a household income equal to zero. This does not mean that individuals that do not meet the requirement received the scholarship since ANSES runs the control without using the field reported to the Ministry of Education, but it does mean that data is not available for this paper. To be able to work with those individuals as well, the idea in this paper is either using income reported in 2019 or 2021 with the adjustments appropriated.

The input field containing the detail of monthly income is reported to the Ministry by ANSES, but the decision on the socioeconomical requirements is still done by the latter and some students get rejected due to bureaucratic reasons. It might be the case that the respective individual has his/her family income wrongly reported in ANSES's system due to missing data or other administrative

¹¹ <https://www.argentina.gob.ar/educacion/progresar>

¹² This limitation is not applied for cases where the applicant has presented a proof of belonging to minority group, being a registered housekeeper or if the individual already was part of the nursery training scholarships in the previous years.

¹³ Argentinian by birth or in possession of the Argentinian citizenship and living in the country for a minimum of five years.

issues producing an error which automatically denies the benefit. Therefore, there are cases of individuals which satisfy the age requirement, the nationality requirement and the income requirement (based on the informative input field mentioned) but are still rejected. Whether an individual is rejected and the reason for its rejection is documented in another input box via a code which allowed this study to isolate the exclusively administrative issues. This study limits the sample to those rejection codes that indicate problems during the bureaucratic procedure and therefore considers individuals which would have obtained the benefit, if those issues had been resolved.

After deciding whether the applicants are eligible for the scholarship based on their socioeconomic conditions and the age restrictions, the Ministry proceeds to consult the educational institutions regarding the applicant's academic status and the details of their study plans. After receiving the information from more than three thousand institutions¹⁴, the scholarship is granted to those students that have completed a minimum of 50% of their study plan so far.

It must be highlighted that the number of *Progresar* scholarships granted is not limited so that every applicant that fulfills the requirements (based on income and academic performance) obtains the scholarship. This limits our methodology selection since we do not have a natural control and treatment group. Therefore, in the following sections the process of selection of the methodology appropriated for the data is described, together with some comments on the selected one and potential considerations for future research and data gathering.

3. Data and Variables

3.1. Micro database construction

This paper uses a unique link between individual-level administrative data from different sources. Data on the socioeconomic variables comes from the ANSES, data on academic variables comes from the educative institution where the applicant is enrolled and which decides whether the respective student is eligible for the benefit. Information on academic performances of both students that continue studying and those that drop-off or simply did not ask for the benefit again is provided by the National University of the North-East.

The sub-set used in this work consists of beginner students from the National University of the North-East that satisfy the age, nationality and income requirements. This means students between 18 and 24 years, with the Argentinian citizenship and that report a household income lower than the sum of 3 minimal wages. The selection was made based on the recent changes in the design of

¹⁴ 2.627 tertiary non-university institutions, 126 universities and 1.287 faculties.

the program, where the academic requirement does not *apply*¹⁵, and the impossibility to homogenize advanced students.

As explained in the previous section, all the applicants that fulfil the requirement receive the scholarship. Thus, in order to obtain a control and a treatment group to test for the impact of the scholarship, we selected individuals between 18 and 24 years old, that presented an income lower than the limit, that either obtained the benefit or not. The control group did not receive the scholarship based on bureaucratic issues with ANSES at the moment of the application. These individuals do not differ significantly from our treatment group, as it will be analyzed in the next section, since, in principle, bureaucratic inconveniences could be solved in most cases, but the applicants did not resolve the administrative problems and hence, finally did not obtain the scholarship.

The complete microdata set used for assigning the scholarships exceeds 900 thousand individuals with applicants from various educational levels¹⁶. However, complete information is only available for the applicants that request the scholarship every year, thus losing considerable amounts of valuable information of those that drop-off or simply consider themselves not eligible for the scholarship and do not apply the following year. This issue could potentially pose a threat to this work and bias results since students self-select whether they are eligible for the grant both for the academic and the income requirement. In order to address the self-selection inconvenience, additional academic data from the National University of the North-East is gathered. The self-selection bias regarding income is also present in data, individuals with a higher income are poorly represented in the whole database, therefore, the decision was to exclude them and only work with those that are below the income limit.

Altogether the micro dataset consist of applicants starting their degrees over the years 2018-2020 of the National University of the North-East¹⁷. Combining the data from the sources mentioned before and matching the individual identifier of students, we have (i) data on household income; (ii) individual personal data (both stemming from the ANSES), (iii) academic information, such as career and educational institution (from the applicant inscription form), (iv) the study plan for each career and (v) the individual performance over the years (both directly retrieved from the

¹⁵ It is important to highlight again that the academic requirement applies to every applicant of the scholarship but for starting students this requirement is equal to 0, since the 50% of the study plan so far is null.

¹⁶ The data set has information of 908.570 individuals that are distributed in: 410.843 universities students; 284.617 tertiary non-university and 213.110 between technical professional training and basic education.

¹⁷ Information available for 16.243 students accessing university that applied for the program between 2018 and 2020.

universities). Thereupon, students with erroneous or inconsistent data were deleted to preclude the final dataset to be biased due to flawed information.

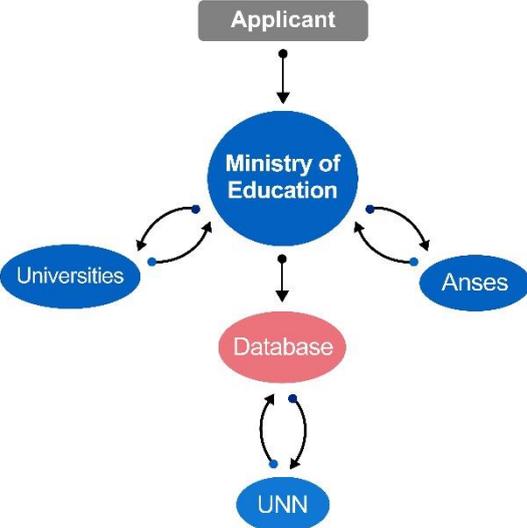
First, the socioeconomic information database from ANSES contains the set of variables: household income, age, nationality, gender, registered residential location, if they suffer a disability and if the student is a registered housekeeper; and determines whether the applicant is eligible. Second, using the information filled by the applicant in their registration form, the Ministry sends the query to the corresponding educational institution to ascertain the student's academic performance. Simultaneously, the institution sends the study plans for all their registered careers alongside with the respective designated number of subjects per year, which is pivotal to determine whether an advanced student satisfies the academic requirements for the scholarship. Lastly, the universities and educational institutions send the individual performance of the previous year and assess whether the student has passed a minimum of 50% of the subjects indicated in the study plan. That's the process done by the Ministry assigning the scholarships.

One extra step we have made in this paper, is the query for applicants starting a career from the National University of the North-East. The University sent us the information for the selection of applicants including the ones that did not applied again for the scholarship and therefore, the Ministry lacks the data of their continuity in the educational system. The data obtained from this particular institution includes the year in which the student started his/her studies, a field explicitly saying if the student continues the studies, the number of subjects approved in total in a specific career and the number of subjects approved in the last year.

This final extra information for the applicants of the National University of the North-East is crucial in our work, since it helps to create a set of objective variables over the years for both the applicants that registered in the next year and those who did not. Sadly, the information regarding student's average grades is not available although future research could exploit it to determine the effect of scholarships.

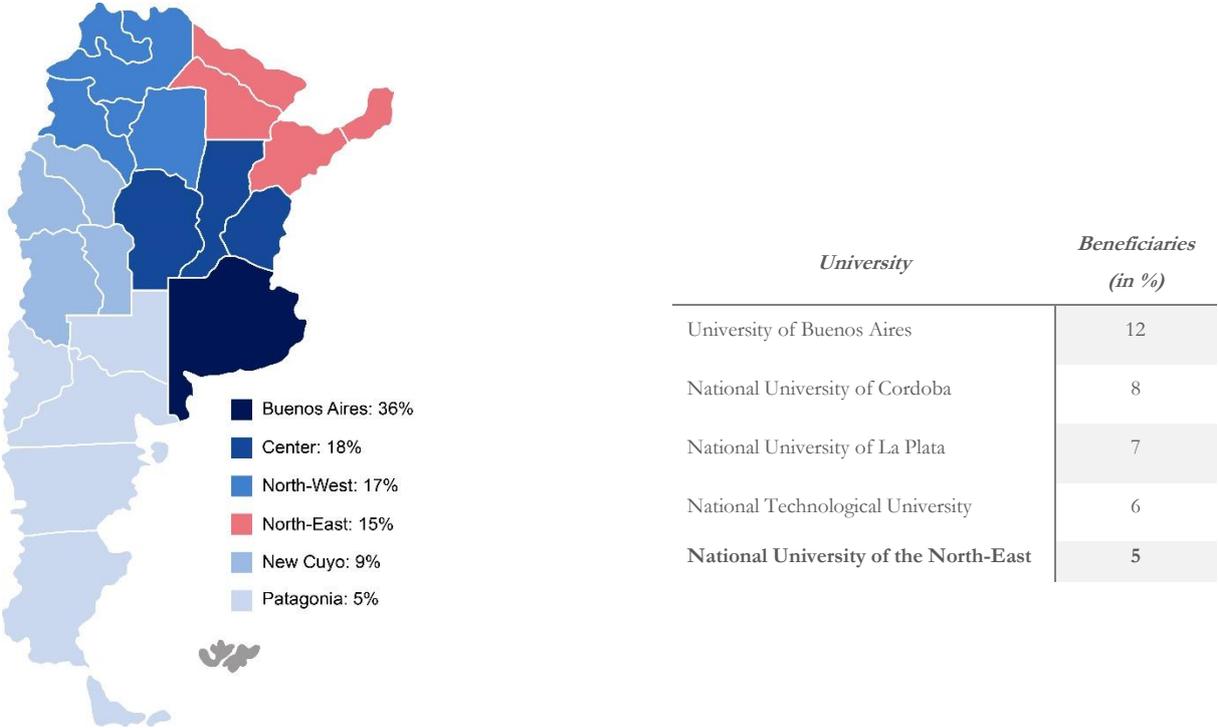
Figure 4 shows how the data from the different organizations is matched in this paper to determine the explicative variables used in the econometric model.

Figure 4 – Map of the connecting databases



A shortcoming of using data on individuals for only one university is that the number of observations is significantly reduced, since data for other universities regarding students that do not apply for the scholarship in the following years is not available. Nevertheless, the dataset still includes more than 3,000 students providing an extensive micro dataset since the representation of the North-East region is high among the *Progresar*'s beneficiaries and the University is 5th in the ranking of institutions with the most scholarships (c.f. Figure 5).

Figure 5 – Distribution of the scholarships per region and top five universities



3.2. Variables and descriptive statistics

The interest variable in the first approach is *cont*, that refer to the continuity in university. It captures students that dropped out of university a year after they applied for the scholarship. Individuals with a value of 0 dropout university, while those with a value of 1 continue with their studies (even though they might not pass any subject the following year).

The set of control variables is decided based on previous literature and the expertise of the Ministry's personnel in charge of the admission process of the scholarships. The summarize of all variables is listed in Table 2.

Table 2 - Variable's descriptions

Variable	Description	Obs	Mean	Std. Dev.	Obs=1
<i>cont</i>	<i>if student continues=1 if not=0</i>	2764	.584	.493	1615
<i>Scholarship</i>	<i>if student received it=1 if not=0</i>	2764	.905	.293	2501
<i>age</i>	<i>individuals age</i>	2764	19.152	1.534	.
<i>female</i>	<i>if student female=1 if not=0</i>	2764	.565	.496	1561
<i>offspring</i>	<i>if student has offspring=1 if not=0</i>	2764	.027	.163	75
<i>applied</i>	<i>applied science career fixed effect</i>	2764	.279	.449	771
<i>social</i>	<i>social science career fixed effect</i>	2764	.408	.492	1128
<i>basic</i>	<i>basic science career fixed effect</i>	2764	.042	.201	116
<i>health</i>	<i>health science career fixed effect</i>	2764	.197	.398	545
<i>humanistic</i>	<i>humanistic science career fixed effect</i>	2764	.074	.262	204
<i>career length</i>	<i>duration in years of the career</i>	2764	5.12	.747	.
<i>Income</i>	<i>monthly reported household income</i>	2764	2675	5084	.
<i>disability</i>	<i>if student has a disability=1 if not=0</i>	2764	.001	.033	3

Table 2 reports variables descriptions and some descriptive statistics. Obs=1 denoted the number of observations that present a value equal to 1.

A first analysis is to test whether our control and treatment groups differ in means for the selection of variables. Table 3 presents the results of the t-test for each variable displayed for both groups (the ones that received the grant and those that did not). Just by looking at the results of the t-test in the outcome variable, means differ between the two groups. Students that receive the scholarship have a mean of 0,60 in *cont* while the ones that did not, exhibit a mean of 0,47. This seems to indicate that individuals that do not receive the benefit have a higher chance to drop out from university.

Table 3 - Two-sample t test with equal variances

Variable	Obs		Mean		dif	St Err	t value	p value
	scholarship=0	scholarship=1	scholarship=0	scholarship=1				
cont	263	2501	0.468	0.597	-.129	.032	-4.05	.000
age	263	2501	19.99	19.06	.928	.098	9.50	.000
female	263	2501	0.563	0.565	-.002	.032	-0,05	.945
offspring	263	2501	0.023	0.028	-.005	.011	-0.45	.650
applied	263	2501	0.178	0.289	-.111	.029	-3.80	.000
social	263	2501	0.498	0.399	.100	.032	3.15	.002
basic	263	2501	0.057	0.041	.017	.013	1.30	.201
health	263	2501	0.187	0.199	-.012	.026	-0.45	.641
humanistic	263	2501	0.080	0.073	.007	.017	0.40	.694
career length	263	2501	5.135	5.118	.017	.049	0.35	.725
income	263	2501	1.841	2.763	-.922	.329	-2.80	.005
disability	263	2501	0.004	0.001	.003	.002	1.40	.160

Table 3 reports variables descriptive statistics and results of the two-sample t-tests with equal variances between control and treatment groups.

Including the gender variable is supported by the increasing participation and graduation of female students in higher education, as shown by Plötz (2017), and Heath and Jayachandran (2016). The ratio between women and men in enrollment and in graduation from higher education has been increasing over the last decades, first in industrialized countries and then in a growing number of developing countries (Heath and Jayachandran, 2016). This increase in the levels of education of women not only led to a convergence in the higher education level by gender but also allowed females to even surpass men and widen the gap in the opposite direction ever since (Plötz, 2017). Females are overrepresented in the enrollment of higher education in 74% of the countries with available data, with the exception of Central and Meridional Asia and Sub-Saharan Africa (UNESCO & IESALC, 2021).

Therefore, the inclusion of the variable is key in our model, since the participation of women in upper education is higher than men in Argentina (Curti, 2004). In the aggregated sample over 56% of the students are female¹⁸, while this number drops to 44% if we consider strategic careers only

¹⁸ Considering advanced students as well, 61% of the North-East applicants are female while 39% are male.

(that include STEM). This is consistent with the 2021 UNESCO report (UNESCO & IESALC, 2021) that outlines the increasing opportunities in the labor market related with those degrees. Obviously, females are therefore over-represented in the non-strategic career in the university analyzed in this paper¹⁹.

Considering this, the inclusion of the gender variable might bias our estimates considering that there is a disproportion between males and females in strategic and non-strategic careers respectively. Therefore, one must include a dummy variable to control for the isolated effect of those careers in the outcome variable. Since using each career²⁰ would reduce the degrees of freedom to a bare minimum, careers are grouped into 5 different groups -see Table 2- with similar study plans and students' characteristics. Also, we include the length of the career to control for any negative effect on the outcome variable of longer careers. By looking at the results of Table 3, the fact that our sample is balanced between the control and treatment group suggests that there is no bias towards a gender difference, since means do not differ statistically significantly.

Students' age is recorded at the day the inscription to the benefit is closed following the Ministry's admission procedure (31st of march). The t-test shown in Table 3 suggests that there is a statistically significant difference between the two groups. Therefore, the inclusion of age and age square is necessary to control for its direct and marginal effect. Furthermore, a set of variables referring to minority groups²¹ is also considered in the analysis. Considering the small size of each group it is not possible to isolate the effect of each of them, so that they had to be subsumed in our sample. Nevertheless, minority groups are not expected to have a significant impact on dropouts mainly due to a lack of observations.

As mentioned before, the income variable that refers to the monthly household income registered in ANSES, presents some methodological difficulties. According to data an outstanding portion registered an income equal to zero²², which is obviously questionable. An income of zero can come about for a couple of reasons. First, informality is a significant phenomenon all over South America and Argentina is no exception. As of 2020, the level of informality amounts to almost 36% (Ministry of Productive Development, 2020). Second, as mentioned earlier, all data that stems from the National Social Security Agency (ANSES) that is before the Ministry of Education was in

¹⁹ The percentage of women in the non-strategic career ascends to 64.

²⁰ In the sample for students starting a degree in the National University of the North-East there are 57 different study plans/careers.

²¹ These variables as shown in Table 2 refer to the specific situation of the individual, whether he/she has offspring or if they have a disability.

²² A 68% of the whole sample presents an income of zero in the National Social Security Agency (ANSES) database. If we consider the National University of the North-East the percentage decrease to 65.

charge is neither precise nor accurate. During 2014 and 2017 individuals could simply submit a statement with which they confirm that they are independent of their parents' income so that their own income was used to assess their eligibility to the grant. Considering this, individuals that presented the statement of independence during those years might present an income equal to zero even though this might not have been the case. Considering that the sample only includes students applying for scholarships from 2018 to 2020 and requires them additionally to be below the age of 24, only few of those students will still receive the grant and hence their share in the sample should be negligible and consequently not confound the data. This affects the whole sample, both the treatment and control group equally, so no bias might come to our selection from this field considering that the sub-set used only includes students with a reported income lower than the limit for both groups. Still, data on income is neither abundant nor robust so that it must be interpreted with caution.

The idea of the second approach is to measure the performance of the individuals after they have received the grant. Since the applicants are distributed over different careers with different study plans, we used the number of passed subjects in the subsequent year, over the number of subjects corresponding to the first year in their respective study plan to homogenize the variable (c.f. Eq [1]). For example, consider an individual that applied in 2019 to study mechanical engineer. In the first year of his/her career, the total number of subjects is 8. If he/she passed 5 of them in 2019, the outcome variable takes the value of $\frac{5}{8} = 0.625$. Although this method homogenizes the performance variable across different courses of study in that it takes a value between zero and one for every student, differences in student characteristics across careers must still be considered. This is why the previously mentioned career fixed effects are included in this study.

$$Performance_i = \frac{\text{subjects passed in 2019}_i}{\text{total subjects of the first year study plan}_i} \quad [1]$$

4. Methodology and empirical model

In this section the methodology selected and the procedure of selection of such methodology is going to be described. Based on previous literature on how to analyze the impact of scholarships on students' performance outcomes the most used methodologies are a Regression Discontinuity Design (RDD), Panel Data, OLS or Logit for the cases where the outcome variable is dichotomic.

The first attempt, trying to follow the steps of Montalbán (2019) was to use an RDD to see the impact of the scholarships on the performance of students the following year, with the objective

of getting a short-term effect of the grant. Two major problems did not allow the use of such methodology. First, the lack of continuous variables both for the running and outcome variables. Since household income as mentioned in the previous section should be used with caution, presenting in most cases an income equal to 0, the natural running variable to get the threshold group needed in an RDD was not available. Of course, this is done to use the assumption that individuals around the threshold are similar to each other, but the number of observations around the cutoff in income to receive the scholarships is not enough in our dataset. An alternative was to use as running variable the percentage of the study plan passed so far (that is the academic requirement that decides whether an advanced student receives or not the grant). But the inconvenient that study plans, and students of different careers, are not homogenous makes the assumption of similarity to not hold. Also, the variable was not continuous by nature, since individuals can pass only integers number of subjects. Secondly, the outcome variable faced the same fate as the suggested running and it was no possible to homogenize students with different progress in their degrees.

Considering this and the most recent change in the design of the program, the focus was on individuals that were starting a degree and not on those advanced with the academic requirement.

Then the possibility of using a Panel Data was discarded since information regarding the income variable in 2020 is equal to zero for every individual.

Finally, the idea of testing the impact of the scholarship on whether first year students continue in university using a Binomial Logistic Regression was the option selected. Firstly, due to the recent change in the program removing the academic requirement, this scheme of scholarships -the ones that initial students have- is more relevant and is going to be the same across the sample²³. Secondly, information is available and testing the dropout of a needed-based grant for low-income students, where continuity is challenging enough, is of significant relevance.

In a second approach to test for the effects of the scholarship on the performance of students we will run an OLS regression. The interest variable will be performance (c.f. Eq [1]) and the set of control variables is the same as in the logical regression.

4.1. Empirical Model Logit

Considering that the outcome variable *cont* is not a continuous random variable the model logit uses the dummy variable to analyze the impact of the scholarship, together with the selected control

²³ Recall that the only academic requirement for a student starting a career is to be enrolled in a higher educational institution.

variables. The model consists of the relationship between a categorical outcome that takes the value of 1 and 0 and categorical, continuous or binary variables. The linear regression of such outcome violates the assumptions of normality in the error term and heteroskedasticity. Using a linear estimation to calculate the probability of dropout could estimate even negative probabilities. Therefore, a Logit or Probit approach is more appropriated. The selection of logit is based on the lack of requirement of a normal distribution (Pradhan and Lee, 2010). Although estimations do not differ considerable in our case.

The methodology is based on the logistic function P_i , determined as:

$$P_i = \frac{\exp(Y)}{(1 + \exp(Y))}$$

where P denotes the probability of dropout related to a certain student, and Y is defined as

$$Y_i = \beta_0 + \beta_1 \text{scholarship}_i + \beta_j X'_i + \varepsilon_i$$

Here scholarship_i takes the value of 1 if student i received the scholarship and 0 if not. X'_i corresponds to the vector of control variables for each student and ε_i is the error term.

As it has been described, the interpretation of the coefficient is not as linear as in OLS, therefore the results will be detailed in the following section together with the outcomes of the regression.

4.2. Empirical Model OLS

The model estimates the impact of the scholarship on the performance of students after a year. The idea is to see if the scholarship has an impact on the percentage of the first-year study plan that a student archives. To do so, the following baseline regression is done,

$$Y_i = \beta_0 + \beta_1 \text{scholarship}_i + \beta_j X'_i + \varepsilon_i$$

where Y_i is the performance as described in equation 3. scholarship_i is the dummy variable if student i received the scholarship. X'_i is the vector of control variables for each student and ε_i the error term.

5. Results

In this section the results for both regression are described and analyzed using an economical, educational and statistical approach.. First, we will focus on results obtained in the logistic regressions. Then, we are going to check if the results obtained from the first model are consistent

with the OLS regression. Finally, the robustness checks are described for both models, together with some preliminary conclusions on the results.

5.1. Logistic model results

The first model is defined around the *cont* (continuity in the studies) outcome from the academic year 2019 for individuals that applied for the scholarship in the beginning of 2019. This means that for students applying for the scholarship in March of 2019 (the start of the academic year), we observe if they dropped out university by March of 2020 (the end of that academic year and beginning of the next one). This guarantees that there is no endogeneity between the variables since all the controls together with the explicative variable are set in 2019 and this cannot be affected by something occurring in the future.

As presented in Table 4, the results show a positive and significant -even at a 1% level- effect of the scholarship on the outcome. If we take the interpretation of the logit estimators, it can be said that odds of continuity in university after the first year are 43% higher for students that received the grant. For the set of control variables, we see a significant and positive effect of female on the outcome at a 5% level. Being a female increases by 20% the odds of the outcome variable with respect to males. These results are consistent with the premise of this work and previous literature.

When it comes to individuals that have an offspring, we can see the negative and significant effect on continuity with their studies with a 5% level as well. The inclusion of this variable, even though as a sub-group only represents 3% of the sample, is significant enough in this study to have an impact on the dropout of students. This is also consistent with previous literature and the expected sign. The minority group variable it would be expected to present a similar estimator, but due to the lack of observations and a low representation in the sample -only a 0.24 percentage- is not sufficient to have an impact and significance in the outcome variable.

For age and age square it can be interpreted that they are not significant at a 5% level, but since in the comparison between the control and treatment groups the result was that there were some statistical differences between groups, it is important to leave them as control and to not bias our estimators. Besides this, the estimators are logical since the effect is negative for every year the student gets older but a decreasing rate as in shown in the positive effect of age squared.

Table 4 - Logistic regression with Odds Ratios

cont	Odds Ratio	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
scholarship	1.434 (.36)	.193 (.134)	2.68	.007	1.102 (.097)	1.865 (.623)	***
age	.502 (-.69)	.316 (.629)	-1.10	.273	.146 (-1.923)	1.723 (.544)	
age^2	1.014 (.014)	.016 (.016)	0.90	.369	.984 (-.016)	1.045 (.044)	
female	1.203 (.185)	.099 (.083)	2.24	.025	1.023 (.023)	1.415 (.347)	**
offspring	.521 (-.652)	.134 (.257)	-2.54	.011	.315 (-1.156)	.862 (-.148)	**
income	1 (0)	0 (0)	0.45	.651	1 (0)	1 (0)	
career length	1.082 (.079)	.061 (.056)	1.40	.161	.969 (-.031)	1.208 (.189)	
applied	2.025 (.705)	.341 (.168)	4.19	0	1.456 (.376)	2.816 (-1.035)	***
social	1.366 (.312)	.217 (.159)	1.96	.05	1 (0)	1.865 (.623)	*
basic	1.075 (.072)	.257 (.239)	0.30	.763	.673 (-.396)	1.716 (.54)	
health	1.096 (.091)	.191 (.174)	0.53	.599	.779 (-.25)	1.541 (.432)	
Constant	1392.806 (-7.239)	8882.998 (-6.378)	1.14	.256	.005 (-5.261)	3,74E+11 (19.739)	
Mean dependent var		0.584					0.493
Pseudo r-squared		0.025					2764
Chi-square		93.124					0.000
Akaike crit. (AIC)		3681.700					3752.793

Table 4 reports the results obtained from the logistic regression with the respective Odds Ratios. Coefficient in brackets correspond to the logit estimation. The logical estimation used robust standard error to control for heteroskedasticity. Endogeneity was already discarded as a potential threat to our estimators.

*** p<.01, ** p<.05, * p<.1

Income and career length are not significant at a 5% level of confidence, but income presents the expected sign. It can be interpreted that the effect is positive, meaning that individuals with a higher income -between those below the limit- have a higher probability of continuity in their studies, which is consistent with Young (2018). The assumption is that if the income variable was more complete, this variable should be significant with the same sign as in this model. The career length is included in the model as a control for different duration and persistence of students in their degrees, though this effect should not be significant at the beginning of the career.

Finally, fixed effects of careers show that the odds of individuals studying an applied science (such as engineer, architecture, etc.) to continue in their studies increase by 102% with respect to those studying humanistic (reference dummy). This estimator is significant even at a 1% level of confidence. A similar result is found for social sciences, where the odds increase by 36% and it is significant at a 5% level. For the other career fixed effects, it is not found a significance in estimators, but both are included in the model as controls.

5.2. OLS results

To support the results obtained so far, an OLS estimation of the performance variable in the academic year where he/she received the scholarship is shown in Table 5 and 6. The variable as described before consists of the number of subjects passed in 2019 for those students that receive the scholarship in the beginning of the scholar year (c.f. Eq [1]).

Table 5 - OLS regression excluding dropouts

performance	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
scholarship	.048	.029	1.69	.091	-.008	.104	*
age	-.279	.139	-2.01	.045	-.551	-.007	**
age^2	.007	.003	1.90	.058	0	.013	*
female	.061	.019	3.20	.001	.024	.098	***
offspring	-.016	.071	-0.22	.824	-.154	.123	
income	0	0	-0.71	.476	0	0	
career length	.146	.016	9.13	0	.115	.178	***
applied	.298	.020	14.97	0	.259	.337	***
social	.306	.019	15.96	0	.268	.343	***
basic	.199	.033	6.01	0	.134	.264	***
health	.353	.029	12.02	0	.296	.411	***
Constant	2.15	1.382	1.56	.12	-.561	4.861	
Mean dependent var		0.359		SD dependent var			0.401
R-squared		0.148		Number of obs			1692
F-test		97.724		Prob > F			0.000
Akaike crit. (AIC)		1464.265		Bayesian crit. (BIC)			1529.469

Table 5 reports the results obtained from the OLS regression without dropouts. 43% of the observations have an outcome equal to 0. The estimation used robust standard error to control for heteroskedasticity. Endogeneity was already discarded as a potential threat to our estimators.

*** p<.01, ** p<.05, * p<.1

Table 6 - OLS regression including dropouts

performance	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
scholarship	.036	.018	1.98	.047	0	.071	**
age	-.262	.085	-3.07	.002	-.429	-.094	***
age^2	.006	.002	2.84	.005	.002	.01	***
female	.051	.013	3.86	0	.025	.077	***
offspring	-.061	.032	-1.89	.059	-.124	.002	*
income	0	0	-0.40	.693	0	0	
career length	.086	.01	8.38	0	.066	.106	***
applied	.214	.014	15.67	0	.187	.241	***
social	.182	.011	15.92	0	.16	.205	***
basic	.115	.019	6.08	0	.078	.153	***
health	.176	.02	8.61	0	.136	.216	***
Constant	2.34	.858	2.73	.006	.658	4.023	***
Mean dependent var		0.211		SD dependent var			0.355
R-squared		0.088		Number of obs			2764
F-test		77.797		Prob > F			0.000
Akaike crit. (AIC)		1961.071		Bayesian crit. (BIC)			2032.633

Table 6 reports the results obtained from the OLS regression including dropouts. 66% of the observations have an outcome equal to 0. The estimation used robust standard error to control for heteroskedasticity. Endogeneity was already discarded as a potential threat to our estimators.

*** p<.01, ** p<.05, * p<.1

The results obtained are the same in terms of significance and sign if individuals that dropout university are included or not (Table 6 and 5 respectively). Both regressions were run as a robustness check in the case some bias might occur for including several students that dropout, giving an outcome equal to zero. Including the whole sample has a 66% of individuals with an outcome equal to 0, while the percentage drops to 43% if dropouts are excluded. In both regressions the results show a significant and positive effect of receiving the scholarship on the performance the following year. In the case where students that abandoned their studies in the university are not included in the regression, the impact of the scholarship is positive and significant at a 10% level of confidence, with an interpretation that individuals that receive the scholarship outperform those that did not by 5% the next year. This means that they do a 5% more of their study plan in their first year at university. For the regression without the sample limitation the effect drops to 3.5 percent but is still positive and significant at a 5 percent level of confidence. By looking at the R² of the regressions we can say that the one excluding the individuals that dropped out has an explicative power of 15% while the one including is 9%. This makes sense considering that the observations added in the latter, are all equal to 0 in the outcome variable, bringing less variability to the sample. But for both cases we can see that there might be some omitted variable bias due to the low R² obtained. The same analysis should be done in future research if more data becomes available for the same sample.

For the rest of the control variables the results are with the same sign as in the logit regression. Females outperform man by 6% with a 1% level of significance. Age has a negative and significant impact on performance of 28% for every extra year, but age square is positive and significant, meaning that this effect is decreasing with aging until the 24-years-old limit. Having a family is still negative but statistically insignificant in this regression. The variable has a small representation on the sample, which leads to these results, but the signs are negative as expected. If we focus on having a family, the variable is significant if the students that dropout are included in the sample. This is consistent and logical based on the results of the logit model, where the probability of dropout is higher for individuals with offspring.

For the career fixed effects and length, the results are like in the logit regression, with positive estimators. Now all the fixed effects are statistically significant at a 1% level, meaning that all the other sciences outperform humanistic. Applied sciences by 30 percent, social by 31, basic 20 and health 35. Finally, income is statistically insignificant and with the negative sign. This result might come from some bias due to the high representation of individuals with an income equal to zero. Therefore, we keep the variable in the model to control but the interpretation should be taken with caution.

5.3. Robustness checks

In this subsection we analyze the robustness of the results obtained so far by testing the prediction power of the logical model. Also, the same empirical strategy is applied for data of 2018 and 2020 with some considerations explained later. All the regressions are run with a robust indicator to control for heterogeneity, and it was already exposed in this paper, endogeneity is not a concern since future outcomes do not affect past variables, supporting the robustness of the standard errors.

First, to see the power of this model on predicting the outcome variable, a comparison between the estimated outcome and the data outcome is done. Considering the data driven outcomes are dichotomic while the estimated ones are probabilistic, in order to compare we need to transform the estimated values. All the estimators for each individual that surpasses a value of 0.5 were considered to predict a continuity in the studies, while the ones below the level were considered dropouts (c.f. Eq. [2]).

$$\hat{Y}_i = \begin{cases} 1 & \text{if predicted value} \geq 0.5 \\ 0 & \text{if predicted value} < 0.5 \end{cases} \quad [2]$$

The results in Table 7 show that this model has a power of prediction of 60%, with a tendency to assume that individuals continue in their studies while they dropped university. Therefore, the

estimators suffer from some bias probably coming from the omission of relevant variables (both observable and non-observable) and few observations. In order to obtain more robust results, this paper encourage institutions to demand more data and do research to test the impact of policies. This will not only provide significant conclusions for research in general, but specially it can provide information to policymakers.

Table 7 - Predictive power of the logit model

$cont - \widehat{cont}$	Freq.	Percent	Cum.
-1	892	32.27	32.27
0	1665	60.24	92.51
1	207	7.49	100.00

Table 7 reports the difference calculated between the real data outcome and the predicted by the logit model. Values equal to 0 are the correct predictions of the model. -1 and 1 means that the model predicts that the student continues (drops out) when the real outcome is that he/she dropped out (continue).

Another approach done in this paper to see the robustness of our estimators, is testing if the results hold for different years that are available. It is important that these tests are not part of the central study in this work because data presents some inconvenience and therefore it would not be as precise as they are for 2019.

The same model and logic are applied for the transition of students that were starting a career in the beginning of the academic year of 2020 (March) and the National University of the North-East gave the outcomes for March of 2021. One central issue that it is faced using data of 2020 is, as mentioned before, that income informative data is not available, and the individuals present an income equal to zero. In order to apply the same methodology as before it is needed to split the sample into a control (that satisfies the conditions besides the bureaucratic problems) and a treatment group. Since income was central together with age to determine the control group, the available income of individuals in 2019, with the same limit as in 2019 is used in this model²⁴. It might not be the best approach since the assumption that income does not change over a year could be considered strong. Although, significant changes in income of low-income families from one year to the next one, do not have a high probability. It is important to highlight that the approach of using income of 2021 -that is available- was evaluated, but since Argentina is a country with a high annual inflation -36.1% from 2020 to 2021²⁵-, adjusting income over the years according

²⁴ To clarify, the income reported in 2019 was used for those students that also applied in 2020 with the same limit of the sum of 3 minimal wages (considering the minimal wage used in 2019).

²⁵ Data obtained from the National Institute of Statistics (INDEC).
<https://www.indec.gob.ar/indec/web/Nivel4-Tema-3-5-31>

to inflation while minimal wage did not follow the same fate²⁶, was not accurate. Therefore, using the same information of 2019 was considered appropriated.

Using information of 2019 for applicants of 2020 limits the number of observations for the model, lowering the power of our estimators. For the OLS regression the decision was to include the dropouts since observations drops drastically if not.

By looking at Table 7 and 8 it is not noticeable drastic changes from one year to the next one. Scholarship is still positive for both continuity and performance, though not significant for latter, which strengthens the results obtained in this paper. For the rest of the control variables, we observe the same positive effect of females on both continuity and performance with a statistical significance of 5%. Similarly, the results resemble in sign and magnitude for the other controls, such as offspring, career fixed effects and career length. Some changes occur though with the age variable but considering the statistical insignificance in both years studied, the interpretation will not be considered.

It can be said that the scholarship seems to show a positive and significant effect on the continuity of students in their studies. Also, the control for female shows some high impact both in continuity and performance, leading to the interpretation that women are outperforming men in the first year of university. And careers from applied and social sciences outperform humanistic sciences. Finally, students that already have a family between the ages of 18 and 24 are more likely to dropout university in their first year. These results are insignificant for the other estimations done in this paper, but this is mainly due to the lack of observations related to the variable and not on its effect on the outcomes.

²⁶ Minimal wage increased by 28% from 2020 to 2021 according to Res 4/2020 of the Ministry of Labor, Employment and Social Security.

Table 8 - Logistic regression with Odds Ratios for applicants in 2020

cont	Odds Ratio	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
scholarship	1.247 (.22)	.167 (.134)	1.65	.099	.959 (-.042)	1.62 (.482)	*
age	3.189 (1.16)	3.702 (1.161)	1.00	.318	.328 (-1.115)	31.026 (3.435)	
age^2	.97 (-.03)	.027 (.028)	-1.09	.277	.919 (-.084)	1.024 (.024)	
female	1.508 (.41)	.211 (.14)	2.93	.003	1.145 (.136)	1.984 (.685)	***
offspring	.729 (-.317)	.361 (.495)	-0.64	.522	.276 (-1.286)	1.922 (.653)	
career length	1.138 (.129)	.106 (.093)	1.39	.165	.948 (-.053)	1.365 (.311)	
applied	3.609 (1.283)	1.009 (.28)	4.59	0	2.086 (.735)	6.243 (1.831)	***
social	1.784 (.579)	.473 (.265)	2.19	.029	1.061 (.06)	2.999 (1.098)	**
basic	2.085 (.735)	.726 (.348)	2.11	.035	1.054 (.053)	4.125 (1.417)	**
health	.667 (-.404)	.203 (.304)	-1.33	.183	.368 (-1)	1.211 (.191)	
Constant	0 (-12.305)	0 (12.122)	-1.02	.31	0 (-36.063)	94207.198 (11.453)	
Mean dependent var		0.591		SD dependent var			0.492
Pseudo r-squared		0.060		Number of obs			1039
Chi-square		74.822		Prob > chi2			0.000
Akaike crit. (AIC)		1342.767		Bayesian crit. (BIC)			1397.173

Table 8 reports the results obtained from the logistic regression with the respective Odds Ratios for applicants in 2020. Coefficient in brackets correspond to the logit estimation. Income variable is excluded since it was used the value of 2019 due to the lack of data in 2020. The logical estimation used robust standard error to control for heteroskedasticity. Endogeneity was already discarded as a potential threat to our estimators.

*** p<.01, ** p<.05, * p<.1

Table 9 - OLS regression including dropouts for the 2020 applicants

performance	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
scholarship	.02	.027	0.73	.468	-.034	.073	
age	-.375	.344	-1.09	.276	-1.05	.3	
age^2	.009	.008	1.11	.266	-.007	.026	
female	.086	.025	3.47	.001	.037	.135	***
offspring	-.099	.082	-1.21	.228	-.26	.062	
career length	.052	.02	2.53	.011	.012	.092	**
applied	-.096	.048	-2.02	.043	-.19	-.003	***
social	.051	.052	0.98	.327	-.051	.154	
basic	-.162	.05	-3.27	.001	-.26	-.065	***
health	-.03	.055	-0.55	.585	-.138	.078	
Constant	3.682	3.429	1.07	.283	-3.047	10.41	
Mean dependent var		0.250		SD dependent var			0.433
R-squared		0.063		Number of obs			1039
F-test		7.644		Prob > F			0.000
Akaike crit. (AIC)		1160.184		Bayesian crit. (BIC)			1214.591

Table 9 reports the results obtained from the OLS regression including dropouts for applicants in 2020. The estimation used robust standard error to control for heteroskedasticity. Endogeneity was already discarded as a potential threat to our estimators.

*** p<.01, ** p<.05, * p<.1

The same model and logic are applied for the transition of students that were starting a career in the beginning of the academic year of 2018 (March) and the National University of the North-East gave the outcomes for March of 2019. It is important to highlight that the only difference with respect to the central test in this paper is the change in minimal wage²⁷. Considering the methodology used in this paper to divide between our control and treatment group, we limited the analysis for individuals that met the income and age requirements and only got rejected due to bureaucratic issues. This means the sample was split using the income limit for 2018 as well in order to see the differences between the two groups.

The t-tests results suggests as in the 2019 analysis, that there is a difference in the continuity in university (*cont*) between the control and treatment group. The first ones present a mean of 0.32 while the latter a mean of 0.43, and the difference being statistically significant. By looking at the results exposed in Table 10, we can confirm the same trend found so far in the logit model. The scholarship has a positive and significant impact on the continuity. Females still present a positive coefficient, but it is not significant. Age is still statistically insignificant as income and career length. While offspring still has a negative effect as expected, though is not significant due to the lack of

²⁷ Minimal salary considered for march of 2018 is \$9.500 according to Ministry of Labor, Employment and Social Security. The requirement for the scholarship is \$28.500 as the sum of 3 minimal wages.

observations in the sample. The results for applied science are consistent with the results exhibited before for the previous years and is the only significant variable.

Table 10 - Logistic regression with Odds Ratios for applicants in 2018

cont	Odds Ratio	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
scholarship	1.407 (.341)	.17 (.121)	2.82	.005	1.109 (.104)	1.783 (.579)	***
age	1.203 (.185)	.646 (.537)	0.34	.731	.42 (-.868)	3.446 (1.237)	
age^2	.993 (-.007)	.013 (.013)	-0.54	.586	.968 (-.033)	1.019 (.019)	
female	1.108 (.103)	.086 (.078)	1.32	.188	.951 (-.05)	1.291 (.255)	
offspring	.627 (-.467)	.178 (.284)	-1.64	.1	.36 (-1.023)	1.094 (.09)	
income	1 (0)	0 (0)	1.25	.211	1 (0)	1 (0)	
career length	1.023 (.023)	.052 (.05)	0.45	.649	.927 (-.076)	1.129 (.122)	
applied	1.434 (.361)	.208 (.145)	2.49	.013	1.08 (.077)	1.905 (.645)	**
social	.94 (-.061)	.128 (.136)	-0.45	.653	.72 (-.329)	1.229 (.206)	
basic	1.167 (.155)	.265 (.227)	0.68	.495	.749 (-.29)	1.82 (.599)	
health	.88 (-.128)	.133 (.152)	-0.85	.398	.653 (-.426)	1.184 (.169)	
Constant	.17 (-1.773)	.927 (5.458)	-0.32	.745	0 (-12.471)	7522.024 (8.926)	
Mean dependent var		0.417		SD dependent var			0.493
Pseudo r-squared		0.018		Number of obs			3158
Chi-square		74.210		Prob > chi2			0.000
Akaike crit. (AIC)		4238.036		Bayesian crit. (BIC)			4310.729

Table 10 reports the results obtained from the logistic regression with the respective Odds Ratios for applicants in 2018. Coefficient in brackets correspond to the logit estimation. The logical estimation used robust standard error to control for heteroskedasticity. Endogeneity was already discarded as a potential threat to our estimators.

*** p<.01, ** p<.05, * p<.1

Table 11 presents the results for the OLS regression on the performance of the same groups of students after a year. The results show that the scholarship is not statistically significant like in the analysis of 2020 applicants.

Table 11 - OLS regression without dropouts for the 2018 applicants

performance	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
scholarship	.016	.029	0.54	.588	-.042	.074	
age	-.391	.134	-2.91	.004	-.654	-.128	***
age^2	.009	.003	2.69	.007	.002	.016	***
female	.001	.02	0.05	.961	-.039	.041	
offspring	.081	.088	0.91	.361	-.092	.253	
income	0	0	-1.21	.226	0	0	
career length	.014	.013	1.07	.284	-.012	.04	
applied	-.177	.034	-5.24	0	-.243	-.111	***
social	.1	.035	2.86	.004	.032	.169	***
basic	-.072	.052	-1.40	.162	-.174	.029	
health	-.077	.041	-1.87	.062	-.159	.004	*
Constant	4.517	1.326	3.40	.001	1.915	7.119	***
Mean dependent var		0.410		SD dependent var			0.393
R-squared		0.110		Number of obs			1454
F-test		19.075		Prob > F			0.000
Akaike crit. (AIC)		1262.275		Bayesian crit. (BIC)			1325.660

Table 11 reports the results obtained from the OLS regression without dropouts for applicants in 2018. The estimation used robust standard error to control for heteroskedasticity. Endogeneity was already discarded as a potential threat to our estimators.

*** p<.01, ** p<.05, * p<.1

As a conclusion of the results of the models, it can be said that the scholarship has a positive impact on the continuity of students and the performance in the following year, though this last result seems not to be consistent. This means that receiving the scholarship allows students to focus on higher education, but the differences between individuals receiving and not receiving the benefit are much higher in the continuity scheme. Therefore, the scholarship is more significant to retain students in higher education than increasing performance. As previous literature suggests, it was also found that females are more consistent and have a higher representation in upper education. Also, they outperform man in the first year of studies.

The fact that a program that started as a conditional transfer to facilitate access of low-income individuals to higher education, shows results consistent with that is not a surprise, but helps to understand and justify why the Ministry is making a significant expense in this program. Further comments on policy implications and recommendations are developed in the following section, together with the final conclusions of this work.

6. Conclusions and policy implications

Through this paper it has been analyzed a conditional transfer program that muted into a scholarship in 2018 using a merit-based scheme. The unique database obtained directly from the administration of the program is a good first opportunity to a policy analysis with an empirical approach to conclude and recommend policy implication for further steps. Database as mentioned has some flaws that do not allow this paper to bring completely robust and conclusive results, though the objectives reached are according to the expectations.

It is important to highlight that the definition of continuity or dropout is not unanimous. Dropout can be defined as leaving the formal system of higher education (Tinto, 2002) from an institutional perspective. But what might be a problem for the educational institution (in our case the National University of the North-East) could be just a transfer between institutions (Fanelli, 2002). Although, what is considered failure from an educational point of view, cannot necessarily mean the same from the individual's position. For the student, dropping out of a university might be a step in his/her own objectives, an adjustment to another career, new study plan, different institution or just simply a change in their personal life not necessarily associated with failure (Parrino, 2014). In this paper we analyzed the scholarship program from an efficiency perspective of the policy, with a special focus on the continuity rate in the National University of the North-East, and the results must be interpreted according to that.

We found that the scholarships present a positive effect in the continuity of students in their degrees for a subset of individuals of the National University of the North-East in the year 2019. The result holds for the same analysis in 2018 and 2020 with positive and significant impact. This result provides an inside of one of the main objectives of the scholarship that is to bring opportunities for low-income individuals to be part of the upper education system (even though is public and free to study in Argentina, socioeconomical barriers still exist) and even more substantial, to keep them in university. Considering data obtained from the National Institute of Statistics (INDEC) in 2014, 18.5% of young people between 15 and 24 years old were unemployed or even economically inactive, without studying. The results obtained in this paper brings important evidence on the effect on keeping low-income students in university, but the results based on performance are not that considerable.

Individuals that do receive the benefit only outperform those that do not by a small percentage, therefore it can be said that the scholarships are good for retaining and even bringing students to university but lacks power to increase the performance of students. This result is not significant in

the other years analyzed. As it has been seen in the literature discussion, the debate is still open whether scholarships or monetary transfers increase the performance of students without other policies supporting. Low-income students face other barriers to access higher education that are not only monetary. There is an educational barrier, a socioeconomical and even a sociodemographic one in some places in Argentina. The system of National Universities could be considered both broad since there is at least one university in each province. And it could be considered insufficient since there are some provinces of considerable size, both in distances to get to school and population, with a low representation. Buenos Aires groups the most academic units with 244 while provinces like San Juan or Santa Cruz only have 6. The COVID-19 pandemic forced academical institutions to install the capacities to bring online classes. This is an outstanding opportunity to delete those sociodemographic barriers in access to higher education.

Therefore, the objectives to lower inequality in education must be an integral policy structure in order to attack all the problems together. This seems to be quite ambitious, but the good communication and coordination between the National and the Provincial authorities has room for improvement.

Also, the objectives of the scholarships should be defined to focus different populations that can be found within the sample. Merit-based grants increase the gap between the poorest and the highest quartiles of earners (Young, 2018), then a different aim for the scholarships should be done for students that belong to the poorest segments of Argentinian population, since income is not the only issue that students face when they start a higher education degree. A suggestion might be to divide the structure of the program between an access to higher education universal assistance for students, and a merit or performance-based program for students of strategic careers. Trying to solve the access to higher education, the desertion rate and the lack of professionals in strategic areas with just one scholarship program doesn't seem to be the solution considering the results obtained in this paper and the mentioned problems of the Argentinian higher education.

For future research it is a must to highlight the relevance of data availability. One of the main issues when it comes to policy decisions is the lack of empirical evidence. A focus on data gathering and accuracy is a key aspect in policy evaluation, and considering the coverage of a program like *Progresar*, it is a great opportunity to develop this aspect. In that sense, this paper is a first approach to a unique dataset that could be extremely exploited with extra data. We encourage future research to analyze this program using the grade scores of students, the number of subjects passed over the enrolled ones and if it is available, a precise income data. The personnel designed to administrate

the scholarships is facing in that direction since every year new variables are included both quantitative and qualitative.

7. Bibliography

- Angrist, J., Lang, D. and Oreopoulos, P. (2009) *Incentives and Services for College Achievement: Evidence from a Randomized Trial*. American Economic Journal: Applied Economics, Vol. 1, No. 1 (January 2009), pp. 136-163
- Angrist, J., Lang, D., Oreopoulos, P. and Williams, T. (2014) *When Opportunity Knocks, Who Answers? New Evidence on College Achievement Awards*. The Journal of Human Resources, Summer 2014, Vol. 49, No. 3, pp. 572-610
- Barro, R. (1997) *Determinants of Economic Growth: A Cross-country Empirical Study*. Lionel Robbins Lectures, MIT Press, Cambridge MA
- Barro, R. and Lee, J-w. (1996) *International measures of schooling years and schooling quality*. American Economic Review 86, Papers and Proceedings, 218-223
- Barro, R. and Lee, J-w. (2000) *International data on educational attainment, updates and implications*. NBER Working Paper 7911
- Celada, V. L. (2020) *Acerca de las causas de deserción universitaria en Argentina a principios del siglo XXI, de las políticas implementadas y nuevas propuestas de retención de población estudiantil*. Revista Científica de UCES
- Cohen, D. and Soto, M. (2007) *Growth and human capital: Good data, good results*. Journal of Economic Growth I 2, 51-76
- Curti, C. (2004) *Los géneros en la Educación Superior Argentina. Análisis Estadístico*. IESALC (Instituto Internacional para la Educación Superior en América Latina y el Caribe). UNESCO
- De Gregorio, J., and Lee, J. (2002) *Education and income inequality: new evidence from cross-country data*. Review of Income and Wealth. Series 48, Number 3
- Fanelli, A. M. (2002), *Indicadores y estrategias en relación con el abandono universitario*. La Agenda Universitaria. Colección Educación Superior. Universidad de Palermo. Buenos Aires

- Fanelli, A. M. (2014), *Rendimiento académico y abandono universitario: Modelos, resultados y alcances de la producción académica en la Argentina*. Revista Argentina de Educación Superior
- Fanelli, A. M. (2017), *Políticas públicas ante la masificación de la educación universitaria: el reto de elevar la graduación, garantizando la inclusión y la calidad*. La Agenda Universitaria III. Propuestas de políticas y acciones. Universidad de Palermo. Buenos Aires
- Freeman, C. E. (2004) *Trends in educational Equity of Girls & Women:2004*. Education Statistics Quarterly, Volume 6, Issue 4, 2004. National Center for Educational Statistics
- Gloom, G. and Ravikumar, B. (2001) *Human capital accumulation and endogenous public expenditures*. Canadian Journal of Economics
- Gneezy, U., and Rustichini, A. (2000) *Pay enough or don't pay at all*. The Quarterly Journal of Economics, 115(3): 791–810
- Heath, R. and Jayachandran, S. (2016) *The causes and consequences of increased female education and labor force participation in developing countries*. National Bureau of Economic Research. Mimeo
- Jacob, B. (2002) *Where the boys aren't: non-cognitive skills, returns to school and the gender gap in higher education*. Economics of Education Review, 21, 589 – 598. National Bureau of Economic Research
- Jensen, E. (1983) *Financial Aid and Educational Outcomes: A St. John, Kirshstein, & Noetf/Financial Aid 403 Review*. College and University 58: 287-303
- Krueger, A. and Lindahl, M. (2001) *Education for growth: Why and for whom? Journal of the convergence hypothesis*. New York University. Mimeo
- Lucas, Jr., R. E. (1988). *On the Mechanics of Economic Development*. Journal of Monetary Economics. University of Chicago, Chicago
- Ministry of Productive Development (2020) *Technical Report*. Buenos Aires, Argentina.
- Montalbán, J. (2019) *Countering moral hazard in higher education: The role of performance incentives in need-based grants*. Stockholm University. Mimeo
- Oloriz, M. G., Fernandez, J. M. and Batto, M. (2020) *Relación entre la planta docente y el abandono interanual en las universidades Argentinas*. Políticas nacionales y gestión institucional para reducir el abandono. 2019: Congreso CLABES IX, Bogotá D.C. Colombia

- Palermo, A. I. (2008) *El acceso de las mujeres a la educación universitaria*. Revista Argentina de Sociología, vol. 4, núm. 7, noviembre-diciembre, 2006, Buenos Aires, Argentina. Consejo de Profesionales en Sociología
- Parrino, M. C. (2014), *Factores intervinientes en el Fenómeno de la Deserción Universitaria*. Revista Argentina de Educación Superior. UNTREF – UNSAM
- Patel R., Richburg-Hayes L., de la Campa E., and Rudd T. (2013) *Performance-Based Scholarships: What Have We Learned?* MDRC Policy brief. New York
- Plötz, J. (2017) *Gender Gap Developments in Tertiary Education - A cross-country time-series analysis on European level*. Universidade Católica Portuguesa
- Pradhan, B. and Lee, S. (2010), *Regional landslide susceptibility analysis using back-propagation neural network model at Cameron Highland, Malaysia*. Landslides. 7(1):13–30
- Richburg-Hayes, L. et.al. (2009) *Rewarding Persistence: Effects of a Performance-Based Scholarship Program for Low-Income Parents*. MDRC Opening Doors. New York
- Sylwester, K. (2002) *Can education expenditures reduce income inequality?* Economics of Education Review 21, pp. 43-52
- Tinto, V. (1995), *Educational Communities and Student Success in the First Year of University*. In *Transition from Secondary School to University*. Monash University Conference Proceedings. Melbourne
- Tinto, V. (2002), *Promoting Student Retention: Lessons Learned from the United States*. 11th Annual Conference of the European Access Network. Prato, Italy
- UNESCO and IESALC (2021) *Women in higher education: has the female advantage put an end to gender inequalities?* UNESCO and International Institute for Higher Education in Latin America and the Caribbean (IESALC)
- World Bank (2019) *Improving Inclusion in Secondary and Higher Education*. Program Information Documents
- Young, K. (2018) *Threatening the American Dream: How State-Funded Merit Scholarships Impact Intergenerational Economic Mobility*. Harvard University