The Role of Information in the Demand for Student Loans: Evidence from a Chatbot in Colombia*

Álvaro Carril† Christian Posso‡ Juan Saavedra§

Job Market Paper

December 17, 2024

Latest Version

Abstract

Low-income students are significantly less likely to apply to college than their high-income peers, even when they have the same academic performance. Given the evidence that credit constraints discourage postsecondary attendance, government-provided student loans are widely used to bridge this gap. However, there is evidence of loan assignment inefficiencies: many eligible students who could benefit from student loans do not apply for them, while at the same time an important fraction of loan beneficiaries struggle to repay theirs. One potential explanation for this is that low-income students are less informed about the costs and benefits of higher education and student loans. We document that this is the case in Colombia, where almost 50 % of high school seniors underestimate their eligibility for student loans, and close to 70 % overestimate the costs of these loans. Using a nationwide experiment, we test an information treatment that provides information about student loans using an interactive chatbot. Overall, access to the chatbot alone increases application rates by 10 %, with larger effects for those assigned to more personalized information treatments. Exploiting the random assignment to the chatbot, we estimate that personalized information about student loans increases application rates by 36 %, which translates into a 27 % increase in loan take-up. Importantly, we find that the effect of the intervention is concentrated among students who are misinformed about student loans, and that these heterogeneous effects are consistent with baseline beliefs. For instance, students who underestimate the costs of loan repayment at baseline are less likely to apply for a loan after receiving repayment information. These results suggest that personalized information can play an important role in correcting widespread misperceptions about student loans, and in improving the efficiency of financial aid.

*We thank Zachary Bleemer, Eduard Boehm, Leah Boustan, Christopher Campos, Adam Kapor, David Lee, Nano Ochoa, Sebastián Otero, Jordan Richmond, Karl Schulze, Micaela Sviatschi, Christopher Neilson, Carolyn Tsao, and Damián Vergara for valuable comments and suggestions, as well as participants in the Labor Seminar, Labor Workshop, and the Policy Research Seminars at Princeton University. We thank Felipe Lizarazo and the Research Department of ICFES for their continuous collaboration with this project, and Silvia Granados at ConsiliumBots for excellent research assistance. We aknowledge financial support from the Overdeck Education Innovation Fund at the Princeton School of Public and International Affairs, and from the Graduate Fellowship in Social Data Science at Princeton. All remaining errors are our own.

†Princeton University

‡*Banco de la República de Colombia*

§University of Southern California

Gaps in college attendance rates by family income are large and persistent in many countries, even when controlling for academic achievement. One possible explanation is that low-income students face financial constraints that discourage or prevent them from attending college. While this barrier to entry has been well-documented in the education literature, it is now exacerbated by the rising costs of college observed in recent years. The most recent evidence in fact suggests an increase in the extent to which credit constraints discourage post-secondary attendance (Bailey and Dynarski 2011; Lochner and Monge-Naranjo 2016).

The complexity of the educational decisions that students face and the uncertainty about the college and loan application processes have been shown to deter low-SES students from applying to college (Dynarski et al. 2021). In particular, lack of information about student loans can rationalize the fact that some students fail to apply for college loans even when they are eligible and would benefit from them, and that others may apply for loans that are not optimal for them. This falls in line with a recent literature that emphasizes the importance of information in shaping educational decisions, and documents that students —specially those from low-SES backgrounds— have biased beliefs about the costs and benefits of college.[1] However, in spite of the fact that this lack of information is well-documented and that it predicts the observed behavior under standard models of human capital investment, previous interventions targeting information frictions in educational decisions have had limited success.[2]

In this paper we explore the effects of giving personalized and dynamically adjusted information about student loans to graduating high school seniors. We aim to test the hypothesis that it is very difficult for students to parse complex information about loan eligibility, characteristics and repayment schedules, and how they apply to their own circumstances. We use a low-cost information intervention to correct biased beliefs about college loans and study its effects on students' application and take-up rates of those loans. Several treatment arms with varying degrees of personalization are used to quantify the effects of different types of information frictions. The intervention is implemented in the form of an automated, interactive chatbot that provides information about different loan options, including their characteristics, eligibility criteria, and repayment schedules. The chatbot provides *personalized* information in the sense that it uses the student's own characteristics to filter the options that are relevant to them, and it is *dynamic* in the sense that it adapts to each student's revealed preferences throughout the interaction. Both of these features are important because they allow us to cut down on the complexity of the provided information to a degree that other at-scale interventions have not been able to achieve.

The IcfesBot project is a set of information interventions designed to help students make better-informed decisions about higher education. The project is a collaboration with the Colombian

[1] For example, Bleemer and Zafar (2018) document that students' perceptions of college costs and benefits in the US are substantially biased, with larger biases among low-income students. Similarly, Hastings et al. (2016) use large-scale surveys of Chilean college applicants to show that students from low-SES backgrounds have relatively less accurate expectations about the returns and costs of even their stated top program choice.

[2] (e.g. Bettinger et al. 2012); correcting beliefs does not always lead to changes in behavior (e.g. Dobronyi, Oreopoulos, and Petronijevic 2019)

Institute for the Evaluation of Education (Icfes), the governmental institution that administers standardized exams.[3] In the fall of 2021, we sent a WhatsApp message to a randomly selected, representative subsample of students who had registered to take the college entry exams. This message allowed students to initiate a conversation with a chatbot designed to give information about the student loans provided by Colombian Institute of Credit for Education and Technical Studies in the Exterior (Icetex henceforth, from *Instituto Colombiano de Crédito Educativo y Estudios Técnicos en el Exterior*), the governmental institution that provides the lion's share of student credit in Colombia. All of the provided information was publicly available, but the chatbot made it more accessible and personalized to the students' individual characteristics and stated preferences.

We identified the population of students using administrative data from Icfes, targeting the cohort of students who registered for the Colombian college entry exams corresponding to the fall of 2021. This cohort is comprised of approximately 500,000 students who were in their final year of high school and were about to take the exam that would determine their entry into higher education in the 2022 academic year. We randomized the treatment at the student level, and assigned close to 60,000 students to receive the initial message via WhatsApp.

Overall, we find substantial positive effects of the intervention assignment on loan application and take-up rates. Loan application rates increased by 10.3% for those assigned to receiving the chatbot, and the rate of loan take-up increased by 8.1%, with the largest effects observed among students assigned to more personalized treatment arms. One important advantage of our intervention is that by its nature we are able to precisely observe the information that each student received, so we can identify who are non-compliers. With this information, we exploit the random assignment of the treatment to estimate the casual effects of different types of information. Overall, we find that for students who received their intended information package, their loan take-up rates increased by at least 11 percentage points, which represents a 30% increase relative to the control group.

While the overall treatment effects are positive, we find that there is heterogeneity in the direction of the effect that is consistent with students' baseline beliefs. For example, personalized information about a loan's monthly payment increased loan take-up rates by 26.8% overall. However, the same information *decreased* take-up rates by 70.7% among students who initially underestimate the loan's true monthly payment. Conversely, this information increased take-up rates by 173.1% among students who overestimate the loan's monthly payment. A similar pattern is observed for other types of information treatments, which provides strong evidence that the intervention was highly effective at correcting biased beliefs.

Our findings show that a low-cost information intervention that corrects biased beliefs about student loans can significantly impact educational decisions. We contribute to a growing literature on the importance of information in shaping educational decisions, with a focus on the increasingly important decisions regarding student loans. In ongoing work, we explore the effects of the

[3]Icfes also conducts assessments at other educational levels and plays a key role in educational research and policy development to improve educational outcomes in Colombia.

intervention on college enrollment, choice of institution and major, as well as college performance and completion. In the long term, we hope to examine the effects of the intervention on labor market outcomes and the repayment of student loans.

This paper is organized as follows. Section 1 provides background information on the context of the study. Section 2 describes the design of the intervention, as well as the data and treatment assignment procedures. Section 3 presents the empirical strategy used to estimate the effects of the intervention. Section 4 presents the results of the study, and section 5 concludes.

1 Higher Education in Colombia

Colombia has close to 300 higher education institutions (HEIs) which offer a wide variety of programs that award technical and professional degrees. The former are offered by technological or technical institutes, and typically last between two to three years. Professional degrees are offered by universities, and last between four to five years.

Higher education programs vary widely in terms of quality, selectivity and price. An important distinction is made between HEIs and programs who choose to undergo the process to obtain a "high-quality accreditation", which is granted by a council comprised of members from the academic and scientific communities.[4] This accreditation is correlated with higher college exit exam scores and higher graduates' earnings (Camacho, Messina, and Uribe Barrera 2017), and is generally perceived as a signal of high-quality education provision. In 2014, only 12 % of HEIs had this accreditation, accounting for 9 % of the programs offered. Around 75 % of HEIs that have this accreditation are universities, and close to 60 % of these universities are private (OECD 2016). In terms of enrollment, approximately 33 % of students attend any high-quality HEIs, and 20 % attend high-quality universities. Those who attend these institutions have significantly higher standardized entry and exit exam scores, and are more likely to graduate on time (Londoño-Vélez, Rodriguez, et al. 2023).

Financial resources are a major barrier to higher education in Colombia, specially to access the most elite institutions (Riehl, Saavedra, and Urquiola 2016). High-quality schools are around twice as expensive as non-accredited schools on average, and private HEIs are eight times more expensive than public ones (OECD 2016). Private schools offer very little financial aid, and only a handful of public, high-quality institutions can offer relatively low fees that are subsidized by the government. Therefore, students from high-income backgrounds can afford to attend high-quality private institutions, while students from low-income backgrounds are limited to public institutions. Because of this, access to high-quality public HEIs is highly competitive, and most students from low-income backgrounds end up attending lower quality institutions, if they enroll into higher education at all (Ferreyra et al. 2017). This results in an allocation into higher education that is heavily influenced by financial constraints, producing a misallocation of talent

[4]Specific programs can apply to obtain high-quality accreditation, but if an institution applies as a whole, every program offered by it gets the accreditation automatically. The accreditation must be renewed periodically.

that perpetuates inequality.

In order to alleviate the financial barriers to higher education, the Colombian government provides support to disadvantaged students through its national student loan agency, Icetex. Incorporated in 1953 as a student loan program, in the last 20 years Icetex manages student loans and grant aid for tertiary education on behalf of public and private organizations. According to Icetex, from 2003 to 2023 it has added more than 620,000 students to tertiary education. Moreover, the proportion of beneficiary students from disadvantaged backgrounds has increased from 78 % in 2003-2014 to 90 % in 2015-2023 (World Bank 2024). In terms of total enrollment, Icetex loan beneficiaries accounted for 20 % of all undergraduates in Colombia in 2011, the largest share of any student loan program in Latin America at the time (World Bank 2012).

The loan program implemented by Icetex is not without its problems. It boasts one of the highest interest rates and shortest repayment terms of the world (Mackenzie 2022), and lacks income-contingent repayment plans, so students must repay their loans even if they do not graduate or do not find a job after graduation. Lozano-Rojas (2018) estimate that in 2011-2012, 40 % of loan beneficiaries do not graduate, and approximately half of them will fail to repay their loans. In response to these circumstances, the country has seen multiple student protests in the last decade, including the national university strike in 2018. Qualitative research has shown that students are often uninformed about the terms of their loans, and that this lack of information can lead to poor decisions about higher education (Mackenzie 2022; Serrano Mora et al. 2023). These studies find that a significant proportion of beneficiaries are incorrect about the terms of their loans, and report that many students regret taking them.

Application and Admission Process. Admission to higher education in Colombia is largely based on a national, standardized entry exam (S11 henceforth, from *Saber 11*) administered by Icfes. S11 is generally similar to the SAT in the United States, but has a wider coverage — more than 90 % of high school seniors take it — and is arguably of bigger importance in the process, with 80 % of HEIs using its score as an admission criterion. Admission requirements and the weights assigned to them vary across institutions, but many schools rely exclusively on S11 scores to determine admission, and almost all of them require having taken the exam. The S11 exam is offered twice a year, in order to accommodate the two different high school calendar regimes that are used by high schools in Colombia (A and B calendars). The A calendar is by far the most commonly used, typically concentrating at least 95 % of the students that take the exam in any given year.[5] Our intervention targeted students who registered for the S11 exam in the A calendar of year 2021, which determined access into higher education for the entering class of 2022. In this cohort, 566,478 students registered for the exam in the A calendar (97.2 % of the total registered). Almost all students who register for the exam end up taking it, with a no-show rate of less than 3 %.

[5]The B calendar is used by a small number of schools, with its main advantage being that it coincides with the northern hemisphere's school year.

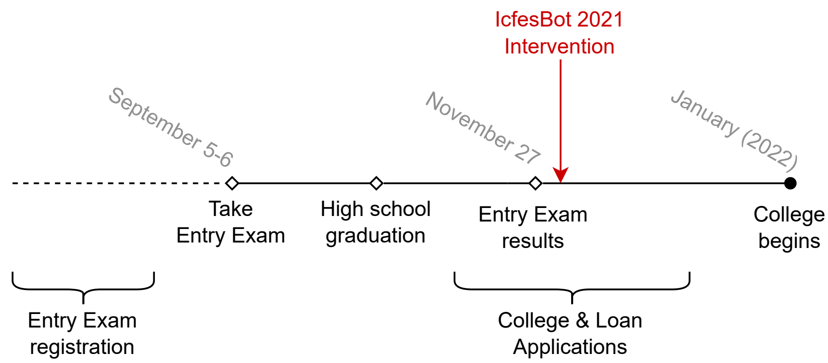

Figure 1: Timeline of the college application process in Colombia (2021)

We focus on the enrollment process for the cohort of students who registered for the exam in the A calendar of 2021, which are the ones targeted by our intervention. The timeline of this process is summarized in Figure 1. It begins with registering for the S11 exam, which must occur before students graduate from their final year of high school. Students take the exam at the beginning of September, and receive their scores by the end of November, after graduating from high school. Many HEIs begin taking applications before the entry exam scores are released, and some students might start their application processes (e.g. submitting other required documentation) before they receive their scores. However, almost all institutions require the S11 scores to be submitted as part of the application—most of them use the scores as an important criterion for admission, and virtually all of them require having taken the exam and meet a minimum score. Similar to the US, the process of applying to higher education in Colombia is somewhat decentralized, with each institution having its own application process and admission criteria.

Application to government-backed college loans is a separate process that runs in parallel to college applications. The specific timelines vary by loan product, but in general students can apply for them after they have received their S11 scores in late November, as most loans require a minimum score to be eligible. In terms of timeline, there are two main types of loans: those that are awarded before the students are admitted to a program, and those that are awarded shortly after, typically because they are contingent on the student being admitted to a program. The former are typically awarded in December and the latter in February or March of 2022, with most programs starting in January 2022. In the following section we provide more details about student loans.

Student Loans in Colombia. The main provider of student loans in the country is the Colombian Institute of Credit for Education and Technical Studies in the Exterior (Icetex henceforth, from *Instituto Colombiano de Crédito Educativo y Estudios Técnicos en el Exterior*), a public institution whose mission is to "promote access to higher education through the provision of financial support to students who are financially disadvantaged and who have good academic performance". Icetex depends on the Ministry of Education, providing financial support using its own resources as well as through administering external educational funds.[6] In 2021 alone, it disbursed 376 million USD in financial support for nearly 400,000 beneficiaries of student loans, including 68 million USD for 46,184 new beneficiaries.[7] In agreement with Icetex' mission, over 90 % of these new beneficiaries are students from the lowest half of the socioeconomic status (SES) distribution.[8]

Icetex finances higher education tuition costs through several *credit lines*: uniquely named loan products each with its own characteristics (i.e. contract terms) and eligibility requirements. Characteristics include the interest rate and the overall repayment term, among others. Eligibility requirements include two measures of the student's SES and the student's academic performance, as measured by their S11 score; these are discussed in more detail below. In the fall semester of 2021, Icetex offered 35 distinct credit lines for undergraduate studies in Colombia, which are listed along their characteristics and eligibility requirements in Table 1. All credit lines finance the full cost of tuition of a program, but some percentage of the debt is to be repaid during the study period, and the overall repayment term is limited to a multiple of the program's duration. There is no income-contingent repayment plan, and the real interest rate is fixed for the duration of the loan.

One of the most important characteristics of a loan product is its interest rate, which determines the cost of borrowing. The annual rates associated to the different Icetex credit lines are summarized in column 2 of Table 1. All interest rates are in real terms, and are fixed for the duration of the loan. Many credit lines have a dual interest rate, with the lower rate being accessible to those who meet the minimum S11 score requirement. Values range from 0 % to 10 %, with the most common rate being 9 %, which is considerably higher than many other countries that employ government-backed student loans. For instance, federal student loans in the United States had a nominal interest rate of 4.3 % for direct loans first disbursed until November 2024.[9] Peru, a Latin American country with a similar per-capita GDP to Colombia, has a nominal interest rate of 3.55 % for government-backed student loans. Chile, which ranks somewhat higher in terms of per-capita GDP, has a real interest rate of 2 % for government-backed student loans.[10] The sheer number of credit lines and the variety of interest rates and other characteristics make the choice of a credit line a complex decision for students, which also contrasts with the situation in other countries like the ones mentioned above, where there are fewer options and the terms are more standardized.

[6]Icetex finances several of its student loans using "administered" funds (i.e. external funds, from *fondos en administración*). These are resources provided by public or private entities, and usually target a specific population. For example, there are student loans financed by NGOs that specifically support students with disabilities.

[7]In 2021 Icetex disbursed 1,421,044,133,686 COP for 230,191 of its current beneficiaries, and 315,938,465,597 COP for 46,184 new beneficiaries. The exchange rate used is 4,580.5 COP/USD, and the data was obtained from `https://web.icetex.gov.co/en/el-icetex/informacion-institucional/estadisticas-oficiales-icetex` (April 5, 2023).

[8]Further details about SES indicators are given in section 1.

[9]See `https://studentaid.gov/understand-aid/types/loans/interest-rates`.

[10]See `https://www.chileatiende.gob.cl/fichas/9583-credito-con-garantia-estatal-cae`.

Table 1: Credit lines for undergraduate programs offered by ICETEX (2021)

Name	Line characteristics				Line pre-requisites		
	Interest rate	% While studying	Grace period (months)	Payment deadline (× program length)	Max. *estrato*	Max. S4 group	Min. S11 score
Tú Eliges - 0%	0	0	12	2.0	3	C7	300
Más Colombiano Que Nunca - 10%	0	10	12	2.0	3	C7	210
Tú Eliges - 25%	0 / +9[*]	25	12	2.0	3	All	270
Fondo Garantías Covid (a. económica) - 0%	0	0	12	2.0	3	C7	290
Fondo Garantías Covid (a. económica) - 10%	0	10	12	2.0	3	C7	290
Fondo Garantías Covid (a. económica) - 25%	0 / +9[*]	25	12	2.0	3	C7	290
Fondo Garantías Covid (a. salud) - 0%	0	0	12	2.0	3	C7	290
Fondo Garantías Covid (a. salud) - 10%	0	10	12	2.0	3	C7	290
Fondo Garantías Covid (a. salud) - 25%	0 / +9[*]	25	12	2.0	3	C7	290
Tú Eliges - 30%	+9	30	6	1.5	All	All	260
Tú Eliges - 40%	+8	40	0	1.0	All	All	240
Tú Eliges - 60%	+7	60	0	1.0	All	All	240
Tú Eliges - 100%	+7	100	0	0.0	All	All	240
CERES	0 / +9[*]	25	12	2.0	3	All	260
Reservistas de Honor	+9	0	6	2.0	All	All	240
Alianzas	0 / +9[*]	0	12	2.0	3	All	210
Oficiales	+9	0	6	2.0	All	All	240
Suboficiales	+9	0	6	2.0	All	All	240
Comunidades con Protección Constitucional	0	0	12	2.0	All	All	210
Funcionarios MEN - 25%	+4 / +10[†]	25	12	2.0	3	All	270
Funcionarios MEN - 30%	+4 / +9[†]	30	6	1.5	All	All	260
Funcionarios MEN - 40%	+4 / +9[†]	40	0	1.0	All	All	240
Funcionarios MEN - 60%	+4 / +9[†]	60	0	1.0	All	All	240
Funcionarios MEN - 100%	+4 / +9[†]	100	0	0.0	All	All	240
Servidores Públicos - 30%	+6 / +9[†]	30	6	1.5	All	All	260
Servidores Públicos - 40%	+6 / +9[†]	40	0	1.0	All	All	240
Servidores Públicos - 60%	+6 / +9[†]	60	0	1.0	All	All	240
Servidores Públicos - 100%	+6 / +9[†]	100	0	0.0	All	All	240

(continued . . .)

Table 1: Credit lines for undergraduate programs offered by ICETEX (2021) *(continued)*

Name	Interest rate	% While studying	Grace period (months)	Payment deadline (× program length)	Max. *estrato*	Max. S4 group	Min. S11 score
Francisco José de Caldas	+8	30	6	1.5	All	All	240
Volvamos a Clases	+7	30	6	1.5	All	All	0
Ser Pilo Paga	+4	0	12	2.0	All	All	0
Buenaventura - 0%	0 / +7	0	12	2.0	3	All	210
Buenaventura - 10%	0 / +7	10	12	2.0	3	All	210
Buenaventura - 25%	0 / +7	25	12	2.0	3	All	210
Apoyo IES	+9	25	12	2.0	All	All	240

Other important loan characteristics refer to the repayment terms: the percentage of the debt that is to be repaid during the study period, the grace period after graduation, and the overall repayment term, summarized in columns 3 through 5 of Table 1. Credit lines finance the full cost of tuition of a program, but some percentage of the debt is to be repaid during the study period (column 3). The grace period after graduation (column 4) is the time during which the student does not have to make any payments, and the overall repayment term (column 5) is the maximum time allowed after graduation to repay the entirety of the debt, in terms of the program's duration. For example, a credit line with a 2.0 repayment term that is used to finance a 4-year program (which is the most common duration for undergraduate programs in Colombia) has to be repaid within 8 years after graduation. In general, repayment terms are less favorable than that of other countries. For instance, federal student loans in the United States have a grace period of 6 months after graduation, and a repayment term of 10 years, with the possibility of extending it to 25 years through income-driven repayment plans.[11] Chile's state-guaranteed student loans have a grace period of 18 months after graduation, and a repayment term of 15-20 years (World Bank 2011).

Loan Eligibility Requirements. Icetex student loans have both merit- and need-based eligibility criteria that students must meet in order to be eligible for a particular loan product. The two most important criteria used to determine loan eligibility are (i) the student's academic performance, and (ii) the SES of the student's household. Academic performance is measured by the student's total S11 score, and almost all credit lines require students to have a minimum score.[12] As indicated in the last column of Table 1, the exact score requirement varies widely by credit line, ranging from the 27th to the 84th percentile of the 2021 distribution of entry exam scores (i.e. 210 to 300 points). For the "*Tú Eliges*" credit lines, which are by far the most popular, the minimum scores range from the 49th to the 69th percentile of the distribution of scores (240 to 270 points). Dual score requirements interact with the interest rate, with the lower score requirement being tied to a higher interest rate, and *vice versa*. Figure 2 shows the distribution of S11 scores for the 2021 fall semester, along with all the cutoffs for the 2021 credit lines. This figure highlights that academic performance requirements are very relevant for a large proportion of students, as a significant mass have a score that is around the cutoffs.

Socioeconomic eligibility criteria are tied to the student's household, and are determined by two indicators: its *Sisben* group, and its *estrato*. The first one depends on the System for the Identification of Beneficiaries for Social Programs (Sisben henceforth, from *Sistema de Identificación de Potenciales Beneficiarios de Programas Sociales*), which uses a voluntary survey to collect information on the living conditions of a household, and then categorizes it into one of 51 groups (DNP 2016).[13] As of March 2020, 39.4 million people (78 % of the population) are covered by Sisben's vulnerability

[11]See `https://studentaid.gov/manage-loans/repayment/plans/standard`.

[12]For the rare credit lines tht do not require a minimum S11 score, these have some other demographic requirement, such as being a member of a specific indigenous group.

[13]Strictly speaking, Sisben assigns each household to one of four groups: group A ("extreme poverty"), group B ("moderate poverty"), group C ("vulnerable"), and group D ("non-vulnerable, non-poor"). However, these groups are further divided into subgroups, with group A having 5 subgroups, group B having 7, group C having 18, and group D

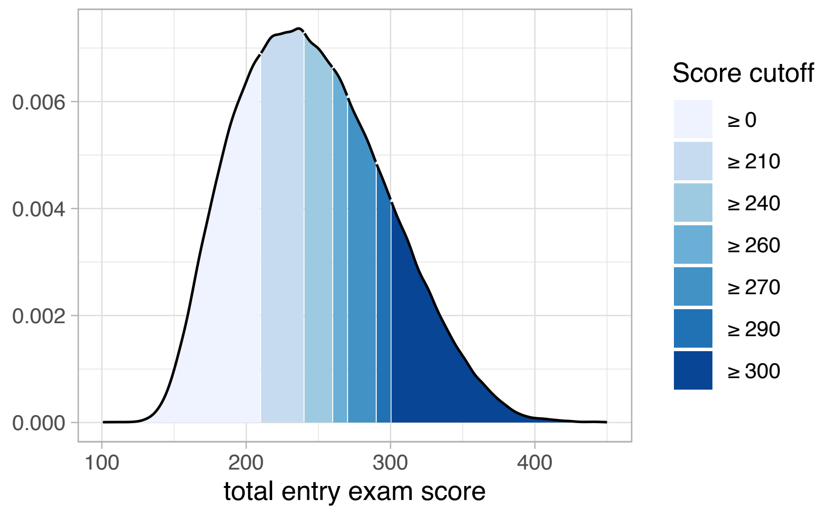

Figure 2: Entry exam score distribution and Icetex credit lines cutoffs (2021)

assessment (Wiseman 2021), with the survey results being used to determine eligibility for a variety of social programs. Insofar as Icetex student loans are concerned, approximately two thirds of the available credit lines do not have a Sisben requirement, and the remaining third require students to be in group C7 or poorer (see Table 1). For the population of students who registered for the S11 exam in the fall semester of 2021, 47 % of them meet this criterion, with 7 % of them not meeting it, and the remaining 46 % not having information on their Sisben group.[14]

The second indicator of socioeconomic status utilized to determine loan eligibility is the household's *estrato* (stratum henceforth), which is assigned at the level of the building in which the household resides. Strata are determined by the municipality in which the building is located, and influence the price of public services, such as gas, water and electricity. Strata range from 1 to 6, with 1 being the lowest (poorest) and 6 the highest (richest). Broadly speaking, strata 1 through 3 will pay subsidized prices, stratum 4 will pay market prices, and strata 5 and 6 will pay a premium on utility prices. All credit lines are available to students from strata 1 through 3, and roughly half of them are available to students with strata 4 through 6 (see Table 1). For the students who registered for the S11 exam in the fall semester of 2021, 50.4 % of them are in strata 1 through 3, and less than 1 % of them indicate being in strata 4 through 6, with the remaining 49.3 % not having information on their strata.[15]

having 21, adding up to a total of 51 subgroups. Subgroups have a numerical index where lower is poorer (e.g. B1 is poorer than B2). However, in the paper we don't make a distinction between groups and subgroups, and simply say that a person or household is in (Sisben) group C7, for example.

[14]This lack of representation of the richer groups is due to the fact that the Sisben survey is voluntary, and richer households are arguably less likely to participate, because there are no benefits for them.

[15]While the percentage of students in the sample that do not have strata information is similar to the percentage that don't have Sisben information, the reasons for this are different. While the Sisben survey is voluntary, municipalities are required to assign a stratum to each building, and the information is publicly available. The lack of strata information is likely due to the fact that students did not provide it when registering for the exam, arguably because they did not

Meeting the minimum academic and socioeconomic requirements is necessary but not sufficient to obtain a loan from Icetex. First, most credit lines typically have additional requirements that are tied to the program of study or to the student's background. Examples of these include applying to a program offered by a particular institution or set of institutions (e.g. outside of Bogotá), or being a member of an indigenous group. Second, each credit line has limited funds, and the number of loans that can be disbursed is depends on the budget allocated to the credit line, the number of applicants, and the tuition of the programs tied to these applications. Finally, students must complete the application process, which includes submitting the required documentation, and signing the loan contract.

2 The IcfesBot Intervention

IcfesBot is a chatbot developed by researchers in collaboration with Icfes, with the overarching goal of understanding how access to information affects students' decisions about higher education. In the wake of the crisis of financing public education that culminated in the 2018 student protests in Colombia, there was increased interested in improving access to higher education, and recent research had underscored the role of information as a cost-effective tool to achieve this goal (see for example Bettinger et al. 2012; Oreopoulos and Dunn 2013; Hastings et al. 2016). The IcfesBot project had its first iteration in 2017, where it was used to provide static "information pop-ups". However, in years 2018 and 2019 a truly interactive chatbot was developed, which was used to provide personalized information about higher education institutions and programs to students at the time of receiving their S11 scores. Because of the Covid pandemic, there was no intervention in year 2020.

In this paper we focus on the 2021 iteration of the IcfesBot, which implemented two important improvements that are central to this study. First, previous iterations of the IcfesBot were designed to only provide information about higher education institutions and programs, without information about how to finance them. In 2021, the chatbot was redesigned to also include information about Icetex student loans, with emphasis on loan *eligibility* and *repayment*. Eligibility information encompasses the requirements for the different credit lines and the student's eligibility for each of them, and repayment information includes the terms of the loan, and the monthly payment that the student would have to make after graduation.[16] Second, the 2021 iteration added a WhatsApp interaction channel to the chatbot, which was previously only available through a web-based interface. This change was made to increase the reach of the chatbot, as the web-based channel had low ($< 5\,\%$) take-up rates in previous iterations, which made it difficult to evaluate the impact of the intervention.

The 2021 version of the IcfesBot used a mix of rule-based and retrieval-based models to generate responses to students' input. In other words, it used a combination of pre-defined responses and

know it.

[16]These two dimensions of information were chosen after a series of focus groups with high school students, as well as based on feedback gathered from Icfes and Icetex.

responses queried from different application programming interfaces (APIs), as opposed to the use of generative artificial intelligence to parse and generate responses which became more common starting in year 2022.[17] Generative AI was a less developed technology at the time, and so this design choice was made to ensure that the chatbot would provide accurate information, as well as to be able to control the information that was being provided to students assigned to different treatment arms. The "conversation" or interaction with the chatbot is structured as a series of questions and answers, and pre-programmed rules determine the flow of the conversation.

The 2021 iteration of the IcfesBot intervention was designed to test the impact of providing information about student loans on students' decisions about higher education. To do this, the chatbot was programmed to provide different information to students depending on the treatment arm to which they are randomly assigned. There are four treatment arms in the intervention.

1. **Career Explorer.** Students in this arm receive information about higher education institutions and programs, but no information about student loans. Based on the student's response to the baseline survey, the chatbot offers the student a list of programs that are related to their interests, and provides information about the institutions that offer these programs.
2. **Loan Availability.** Students in this arm receive non-personalized information about the availability of student loans. The chatbot informs the student that Icetex has a variety of loans available, and that they can apply for them if they meet the eligibility requirements. A URL that points to the official Icetex page that contains a directory of undergraduate loans is provided.
3. **Personalized Loan Eligibility.** Students in this arm receive personalized information about their eligibility for Icetex student loans. The chatbot informs the student about the credit lines for which they are eligible, based on their S11 score and SES indicators. Students can ask the chatbot for more information about any of the loans fo which they meet the requirements, and the chatbot provides information about the terms of the loans, including the interest rate, the percentage of the debt that is to be repaid during the study period, the grace period after graduation, and the overall repayment term.
4. **Personalized Loan Repayment.** Students in this arm receive personalized information about the repayment terms of Icetex student loans. The chatbot informs the student about the terms of the loans for which they are eligible, including the monthly payment that the student would have to make after graduation. The chatbot also provides information about the grace period after graduation, and the overall repayment term.

Throughout the rest of the paper, we refer to these treatment arms by these names or by their corresponding numbers (1 through 4). Treatment 0 denotes the control group, which are students not assigned to receiving the chatbot.

The information that is both collected and provided by the chatbot is dependant on the student's pre-assigned treatment arm. A simplified schematic of the interaction tree is shown in Figure 3.

[17]See Adamopoulou and Moussiades (2020) for more rigorous definitions of these concepts.

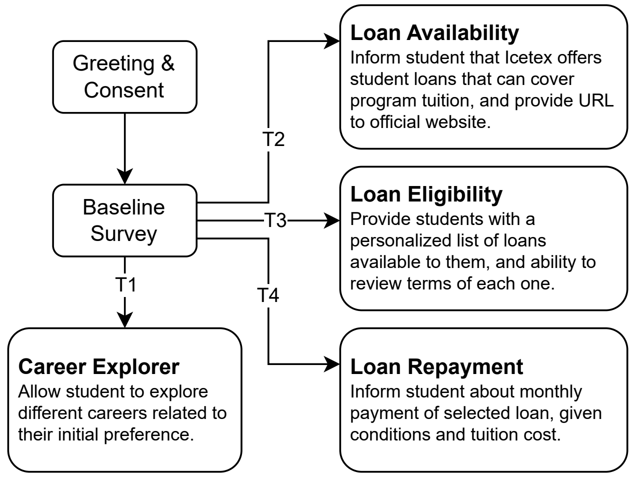

Figure 3: Simplified flow of the IcfesBot 2021 intervention

All students contacted by the chatbot receive a greeting message that contains their first name, and then are asked to provide their consent to continue the conversation and participate in the study. A baseline survey follows, where we elicit information about the student's plans for higher education. After the baseline survey, the chatbot provides the student with the information that corresponds to their treatment arm.

Treatment Assignment. We assign treatment status at the student level, following a three-step process summarized in Figure 4. First, we define an estimation sample of 103,831 students by excluding those who are assigned to the web-based treatment, and those whose phone numbers are invalid.[18] Second, we randomly assign students in this sample to a treatment dummy indicating whether they are contacted by the chatbot via WhatsApp. Our budget and expected response rates determined the number of students assigned to the treatment group, which is 60,139. Therefore, the remaining 48,234 students are assigned to a pure control group with no access to the chatbot through either channel. Third, the 60,139 students in the coarse treatment group are randomly assigned to one of the four treatment arms in fixed proportions: $16.\overline{6}\%$ ($N = 9{,}986$) to treatment 1, $16.\overline{6}\%$ ($N = 10{,}078$) to treatment 2, $33.\overline{3}\%$ ($N = 19{,}910$) to treatment 3, and $33.\overline{3}\%$ ($N = 20{,}165$) to treatment 4.[19]

We carry out a stratified treatment assignment in an effort to preserve covariate balance across treatment arms. We stratify the sample by student gender and whether they presented the S11

[18]In the fall of 2021, the IcfesBot was deployed through two different channels: WhatsApp and a web-based chatbot. Icfes required that 450,000 students were able to interact the chatbot via the web-based channel, leaving the remaining 116,478 students for the WhatsApp experiment that we focus on in this paper; assignment to either group was random. Out of the 116,478 students not assigned to the web-based treatment, we exclude those whose phone numbers were missing, repeated, or otherwise invalid, leaving us with a final sample for the experiment consisting of 108,373 students.

[19]The proportions of each one are determined by institutional constraints imposed by Icfes. The final number of students in each treatment arm is slightly different from the target number due to the presence of *misifts* across the strata implied by the stratified randomization (see Carril 2017).

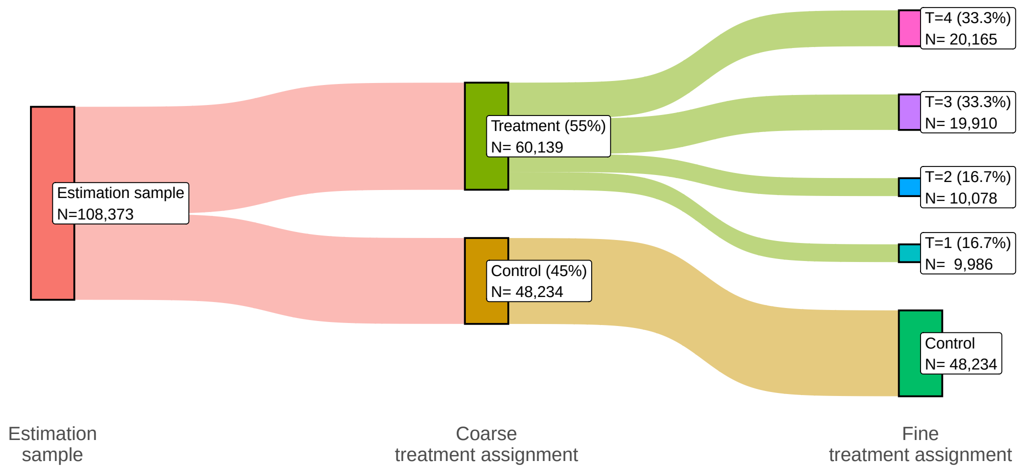

Figure 4: Diagram of the treatment assignment process.

exam, as well as by a measure of their household's SES group and its coarse location at the administrative level[20], and by the student's school administration type and whether it is located in an urban or rural area. This defines 872 strata, and random assignment into the treated group is carried out within each strata in a proportion that is consistent with the total target number of students that are assigned to the WhatsApp intervention. The same process is carried out for the assignment of students in the treatment group to the four treatment arms, this time assigning a fixed proportion of students in each strata to the corresponding treatment arm, as described above.

Treatment assignment was carried out at the student level, in order to maximize statistical power. The main risk of this approach is that there may be spillover effects between students in the treatment and control groups—in particular, we worry that students in the control group might receive the treatment. Similar information interventions usually circumvent this issue by assigning treatment at a coarser level (e.g. Dynarski et al. 2021, assign treatment at the school level).

There are two reasons to expect no major spillover effects in our intervention. First, the intervention occurred after students had graduated from high school, which creates a practical barrier to learning about and receiving the treatment. Second, it is made clear that the information provided by the chatbot is publicly available, and any personalized information provided by the chatbot is based on the recipient's own data. This reduces the incentive for students in the control group to seek out talking to the chatbot. After the intervention, we verify that less than 1 % of the students in the treatment group interacted with the chatbot more than once, and that 98 % of interactions occur in the first 30 minutes after the chatbot's first message. This suggests that it is unlikely that students in the control group would have sought out the chatbot on their own.

Indirect spillover effects such as students in the control group learning from their treated peers or being incentivized to look for information on their own are harder to rule out. However,

[20] Colombia is subdivided into 33 distinct administrative areas called *departamentos*.

Table 2: Balance Table: Mean Characteristics for Students by Treatment Status

	Control (N=48234)		Treated (N=60139)			
	Mean	Std. Dev.	Mean	Std. Dev.	Diff. in Means	p
Individual						
Share Female	0.54	0.50	0.54	0.50	0.00	0.88
Share Minority Ethnic Group	0.05	0.22	0.05	0.22	0.00	0.69
Share Foreign	0.01	0.10	0.01	0.10	0.00	0.96
Share Has Job	0.38	0.48	0.37	0.48	0.00	0.40
Share Takes Entry Exam*	0.96	0.20	0.96	0.20	0.00	0.19
Entry Exam Percentile*	51.35	28.71	51.74	28.69	0.40	0.03
Household						
SES index	2.27	1.27	2.28	1.28	0.00	0.59
Mother's education (years)	10.11	4.48	10.12	4.48	0.01	0.78
High School						
Share Public	0.78	0.42	0.77	0.42	0.00	0.71
Share Rural	0.14	0.35	0.14	0.35	0.00	0.29
Share Vocational	0.29	0.34	0.29	0.34	0.00	0.77

* Values were realized after randomization.

these would in any case attenuate estimated effects towards zero, which would make our results conservative.

Data. Our target population is high school seniors who are considering enrolling into higher education in Colombia. In particular, we focus on students who registered to take the S11 exam in the fall semester of 2021, which determined entry into higher education for the entering class of 2022-1. We identify and characterize these students using longitudinal, student-level administrative data that contain the population that registered to take the entry exam, which for the 2021-2 process consists of 566,478 students. Administrative data of the target population come from Icfes, and encompass a rich set of demographic, school, and household characteristics. Demographic information includes gender, ethnicity and nationality of the student, and household information includes location at the municipal level, and the two measures of socioeconomic status discussed in detail in the context of loan eligibility requirements (section 1). School data include its area type (i.e. urban or rural), location at the municipal level, as well as school type (i.e. private or public). After the intervention, we supplement these data with information on the students' performance in the S11 exam as well as information on loan application and approval, which exist for 540,755 students (96%).[21] The data on loan application and approval come from Icetex, the public institution that provides student loans in Colombia (see section 1 for details).

3 Empirical Strategy

Intent to Treat. We are firstly interested in measuring the effect of *assignment* to the Icfesbot intervention on the rate of loan applications and take-up for undergraduate loans provided by Icetex, which corresponds to the effect of the Intent to Treat (ITT). The ITT is policy-relevant because it represents the causal effect of being targeted by the intervention, regardless of whether the recipient actually decides to participate (i.e. engage with the chatbot). This measure allows us to compute the cost-effectiveness of the intervention, because it considers the effect on every participant on which money was spent to provide the treatment.[22]

We obtain the "coarse" ITT effect of the intervention by comparing the outcomes of students assigned to *any* treatment to the outcomes of those assigned to the control group. That is,

$$\mathbb{E}[Y_i \mid \mathbb{1}(Z_i > 0), \boldsymbol{X}_i] - \mathbb{E}[Y_i \mid Z_i = 0, \boldsymbol{X}_i], \tag{1}$$

where Y_i is an outcome of interest for student i. Here $Z_i \in \{0, 1, 2, 3, 4\}$ is a categorical variable indicating the treatment arm to which student i is assigned, so $\mathbb{1}(Z_i > 0)$ is a binary indicator of assignment to receiving the chatbot. Finally, in most models we include $\boldsymbol{X}_i$, a vector of the covariates listed in Table 2. Empirically, we measure this difference by estimating by ordinary least squares (OLS) the equation

$$Y_i = \beta_0 + \beta_1 \mathbb{1}(Z_i > 0) + \beta_2 \boldsymbol{X}_i + \epsilon_i. \tag{2}$$

The parameter of interest is β_1, which measures the average causal effect on the outcome Y of being randomized into any group that receives the chatbot (i.e. $Z > 0$).

Estimating the effect of being assigned to receive the chatbot is an aggregate measure of the effect of the intervention, as it does not distinguish between the kinds of information provided by it. Heterogeneous effects of different types of information shed light on the mechanisms through which the intervention operates, and can inform the design of future policies. Therefore, we are also interested in estimating the effect of being assigned to treatment arm $j \in \{1, 2, 3, 4\}$ relative to the control group. That is,

$$\mathbb{E}[Y_i \mid \mathbb{1}(Z_i = j), \boldsymbol{X}_i] - \mathbb{E}[Y_i \mid Z_i = 0, \boldsymbol{X}_i], \tag{3}$$

[21] Entry exam scores data were only available after the intervention. Icfes has a strict policy of not disclosing exam scores before the official release date, which implied that we had to define the sample and assign treatments without knowing the students' scores.

[22] However, by the nature of our intervention the cost of providing the treatment actually increases with the level of engagement by the participant, as the cost increases with each message sent and received by the chatbot.

where $\mathbb{1}(Z_i = j)$ is an indicator variable for student i being assigned to treatment arm j. The regression analog of this difference is

$$Y_i = \beta_0 + \sum_{j=1}^{4} \beta_{1j} \cdot \mathbb{1}(Z_i = j) + \boldsymbol{X}_i \boldsymbol{\beta}_2 + \varepsilon_i, \tag{4}$$

where the parameter of interest now is $\boldsymbol{\beta}_1 = (\beta_{11}, \ldots, \beta_{14})$. The coefficient β_{1j} measures the average causal effect on the outcome Y of being randomized into treatment arm j relative to the control group.

We observe the outcomes for all students in our sample, so there is no attrition due to non-response. However, we also observe that a sizeable fraction of students in the treatment group did not engage with the chatbot, and so did not receive the information treatment—they are effectively untreated. We address this issue below.

Treatment on the Treated. The ITT captures the unbiased causal effect of the policy (i.e. of being assigned to treatment), but it does not capture the effect of actually *receiving* the information treatment. In this section we are interested in estimating the effect of the Treatment on the Treated (TOT), that is, the average effect of the information on those who actually received it. In order to do this, we need to account for imperfect compliance in treatment take-up.

We observe substantial non-compliance for those assigned to treatment, as many students in the treatment group did not engage with the chatbot, and so did not receive the information intended for them. The presence of these *never-takers* implies that the TOT effect will be larger in magnitude than the corresponding ITT effect in our setting, as the latter is "diluted" by the presence of those who were in fact not treated (Angrist and Pischke 2009).

Let T_i denote treatment compliance for student i, defined analogously to Z_i: $T_i \in \{0, 1, 2, 3, 4\}$ is a categorical variable indicating whether student i actually interacted with the chatbot enough to receive the information in the treatment arm to which they were assigned. Under perfect compliance, we would have $Z_i = T_i$ for all students. However, in our setting we have that

$$T_i = \begin{cases} Z_i \text{ or } 0 & \text{if } \; Z_i > 0 \\ 0 & \text{if } \; Z_i = 0. \end{cases}$$

The first case reflects the fact that we observe a specific type of non-compliance: if student i has $Z_i > 0$, but does not engage with the chatbot, then $T_i = 0$. Implicitly, we assume that non-compliance can only take the form of not receiving information at all, but not of receiving information intended for a different treatment arm (i.e. $T_i = Z_i$ or $T_i = 0$ if $Z_i > 0$). The second case corresponds to the assumption that non-compliance is one-sided: it only arises from students assigned to treatment failing to receive the information, but not the other way around (i.e. $T_i = 0$ if $Z_i = 0$). In other words, we assume that no student in the control group received the information packet.[23]

[23]These are reasonable assumptions in our setting for three reasons. First, there are practical barriers to students

The main concern for identification is that non-compliance is endogenous: students assigned to the treatment who do not engage with the chatbot and fail to receive the information intended for them may be systematically different from those who do. As it is standard in the literature, we overcome this issue by employing an instrumental variables (IV) approach, where we exploit the randomly assigned treatment as an instrument for the actual treatment received to address the endogeneity problem. Under the IV framework, we can recover the local average treatment effect (LATE): the effect of the treatment on the population of compliers. Furthermore, the LATE recovers the effect of the treatment on the treated in our setting (Angrist and Pischke 2009).

We estimate the LATE by two-stage least squares (2SLS). For treatment arms $j > 0$, let $Z_i(j)$ be an indicator variable that equals 1 if student i is assigned to treatment arm j, and 0 if assigned to the control group. Similarly, let $T_i(j)$ be an indicator variable that equals 1 if student i is a complier in treatment arm j, and 0 otherwise.[24] Therefore, for all treatment arms $j > 0$, the first stage and second stage equations are

$$T_i(j) = \pi_0 + \pi_1 Z_i(j) + \boldsymbol{X}_i \boldsymbol{\pi}_2 + \nu_i \tag{5}$$

$$Y_i = \beta_0 + \beta_1 \widehat{T}_i(j) + \boldsymbol{X}_i \boldsymbol{\beta}_2 + \varepsilon_i. \tag{6}$$

4 Results

IcfesBot Assignment Effects on Application and Take-Up of Icetex Student Loans. Students assigned to receiving the IcfesBot are significantly more likely to apply to and obtain a loan than those in the control group. Overall, the application rate of those assigned to the treatment is 10.3 % higher than that of the control group. This difference in application rates translates into a 8.1 % higher loan take-up rate for those assigned to receive the IcfesBot. This corresponds to the net effect of the treatment on the joint likelihood of applying for and obtaining a loan. Both of these differences are statistically significant at the 1 percent level. See Table 3 for regression estimates and control means.

The increase in loan take-up rate for the treatment group is entirely driven by the increase in application rates, as the effect on loan take-up conditional on application is not statistically different from zero. This finding is consistent with the idea that assignment to the chatbot is effective at increasing the likelihood of students applying for a loan, but not at changing the likelihood of being approved for a loan, conditional on applying. However, we do not have

accessing the information intended for other treatment arms, as the only means for them to do so it to use the cellphone of a student assigned to a different treatment arm, and the intervention happened when school had finished. Second, there is a lack of incentives to do so, as all the data employed by the chatbot is public and readily available, and the information relied in conversations is tailored to the targeted student's characteristics only. Third, we are able to observe whether a student uses the chatbot multiple times, which would be indicative of them sharing their cellphone with other students. We find only very few instances of this, and we drop these students from the analysis.

[24] Note that this implies that the instrument is undefined for students assigned to any treatment arm $k \neq j, k > 0$. That is, we estimate the LATE for students assigned to treatment arm j relative to the control group, and we drop from the analysis students assigned to any other treatment arm.

Table 3: Estimated effect of IcfesBot assignment on loan application and take-up rates

	Applies for loan		Obtains a loan			
	(1)	(2)	(3)	(4)	(5)	(6)
Assigned to chatbot	0.006***	0.006***	0.004**	0.003**	−0.001	−0.001
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Applies for loan					0.710***	0.708***
					(0.001)	(0.002)
Assignment effect as %	11.4	10.3	9.7	8.1	−1.6	−2
Control group mean	0.057	0.058	0.041	0.042	0.041	0.042
Control variables	No	Yes	No	Yes	No	Yes
Observations	103,831	95,347	103,831	95,347	103,831	95,347

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

direct experimental evidence of the effect of the chatbot on the approval rate conditional on application, because the treatment is not randomly assigned among those who apply for a loan. Comparing conditional take-up rates across the control and treatment groups, we observe that 71.8 % of the 2,621 students in the control group who applied for a loan obtained one, and 70.5 % of the 3,636 students in the treated group who applied for a loan obtained one. As a reference, the unconditional take-up rate for the entire 2021 cohort is 4.1 %, with a conditional take-up rate of 70.9 %. Taken together, these results suggest that the students in our sample are fairly representative of the average student in the 2021 cohort in terms of loan approval rates.

Simply being assigned to the chatbot might nudge students to apply for a loan, irrespective of the information they receive. However, we find evidence of heterogeneity in the effect on loan application and take-up by the type of information students were assigned to receive, with the largest effects for students assigned to personalized information treatment arms. In our preferred specification (with controls), assignment to receive personalized information about loan eligibility and loan repayment increases application rates by 13 %. The treatment effects of non-personalized information are smaller, with an increase in application rates of 9 % for generic loan availability information, which is only statistically significant at the 10 percent level. Assignment into a treatment arm that does not provide information about student loans does not have any significant effect on loan application or take-up rates. See Table 4 for all assignment estimates by treatment arm.[25]

[25]The estimated coefficients for personalized loan eligibility ($Z = 3$) and loan repayment ($Z = 4$) information on loan applications are statistically different from the estimated coefficients for information unrelated to student loans ($Z = 1$) at the 5 percent level, but not from the estimated coefficients for generic loan availability information.

Table 4: Estimated effects of information assignment by type on loan application and take-up rates

	Applies for loan		Obtains a loan	
	(1)	(2)	(3)	(4)
Degrees costs ($Z = 1$)	0.001	0.001	0.001	0.001
	[2%]	[1%]	[2%]	[2%]
	(0.003)	(0.003)	(0.002)	(0.002)
Loan availability ($Z = 2$)	0.006*	0.005+	0.003	0.002
	[10%]	[9%]	[8%]	[6%]
	(0.003)	(0.003)	(0.002)	(0.002)
Loan eligibility ($Z = 3$)	0.008***	0.008***	0.005**	0.005*
	[15%]	[13%]	[13%]	[11%]
	(0.002)	(0.002)	(0.002)	(0.002)
Loan repayment ($Z = 4$)	0.007***	0.007***	0.005**	0.004*
	[13%]	[13%]	[11%]	[10%]
	(0.002)	(0.002)	(0.002)	(0.002)
Control group mean	0.057	0.058	0.041	0.042
Control variables	No	Yes	No	Yes
Observations	103,831	95,347	103,831	95,347

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: The numbers in brackets correspond to estimates as percentage of the control group mean.

Table 5: Estimated effects of information assignment by type on loan application and take-up rates

	Applies for loan				Obtains a loan			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Degrees ($T = 1$)	0.003				0.002			
	(0.008)				(0.006)			
Availability ($T = 2$)		0.014*				0.006		
		(0.007)				(0.006)		
Eligibility ($T = 3$)			0.021***				0.013**	
			(0.006)				(0.005)	
Repayment ($T = 4$)				0.020***				0.011**
				(0.006)				(0.005)
Effect as %	0.045	0.242	0.376	0.36	0.054	0.155	0.311	0.277
Control group mean	0.057	0.057	0.057	0.057	0.041	0.041	0.041	0.041
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	51,230	51,301	60,059	60,206	51,230	51,301	60,059	60,206

Note: *p<0.1; **p<0.05; ***p<0.01

Information Effects on Loan Application and Take-Up. Students who engaged with the chatbot and actually *received* information about student loans exhibit substantially higher loan application and take-up rates compared to those in the control group. The magnitude of the effects are larger than those found for simply being *assigned* to an information treatment arm, a difference that arises from the large share of non-compliers in our experiment. Consistent with the ITT results, larger effects are found for students who receive personalized information about loans: loan application rates are 36 % higher for students who receive personalized repayment information, and 37.6 % higher for students who receive personalized eligibility information. These treatments also increase take-up rates by 31 % and 28 %, respectively. See Table 5 for regression estimates and control means.

Generic information about loan availability appears to have an overall smaller effect on loan application behavior, with a 24 % increase in application rates and a 15 % increase in take-up rates. Moreover, information about college degrees does not seem to have a significant impact on loan application behavior, with effects on both outcomes that are close to 5 %. None of these results are statistically significant at the 10 percent level, except for the effect of loan availability information on application rates.

Heterogeneity in Effects of Information by Baseline Beliefs. Results from the previous section show that personalized information about student loans has a significant effect on increasing loan application and take-up rates. However, we are interested in understanding whether the effect of the information treatment varies by students' baseline beliefs about loan conditions. For instance, we would expect that the effect of information about loan repayment would be larger for students who initially overestimate the monthly payments of a loan than for those who have an accurate perception of these conditions. Conversely, information about loan repayment for students who underestimate the repayment conditions of a loan might not have any effect, or even a negative one. Observing heterogeneity in the effects of information treatments by baseline beliefs would provide evidence that the *information* is actually interacting with students' prior beliefs, and that we are not simply encouraging all students to apply for a loan regardless of the information they already have about them.

Students randomized into receiving personalized information about loan eligibility and loan repayment were surveyed about their baseline beliefs regarding these topics. For the loan eligibility information treatment, we ask students how many credit lines they think they are eligible for, and for the loan repayment information treatment, we ask students how much they think they would have to pay monthly for a loan (see section 2 for additional details). Both of these questions are asked before the information treatment is provided. We use these responses to construct indicator variables for whether a student over- or under-estimates their eligibility for a loan, and whether they over- or under-estimate the repayment conditions of a loan.

We find that the effect of personalized loan repayment information on loan application and take-up rates varies substantially by students' baseline beliefs about loan repayment conditions (see Table 6). First of all, we find that in spite of the fact that the overall effect of the treatment is positive for the sample assigned to that treatment, the net effect is *negative* for the subgroup of students who initially underestimate the monthly payments of a loan, who see their application and and take-up rates decrease by 70 %. In contrast, the net effect on application and take-up rates for treated students who initially overestimate the monthly loan payments is more than 150 % than that of the control group. We also find effect heterogeneity that is consistent with the direction of students' baseline beliefs for the personalized loan eligibility information treatment (see Table 7): Treated students who overestimate their loan eligibility see their application rates decrease by 51 % compared to the control group, while treated students who underestimate their eligibility see their application rates increase by 111 %. These results strongly support the idea that the information provided by the chatbot is effective at changing students' behavior in a way that is consistent with their prior beliefs.

5 Conclusion

In this paper, we highlight the importance of personalized information in correcting biased beliefs about student loans, and explore how this impacts loan application and take-up. In particular, we analyze the impact of personalized information on the demand for student loans in Colombia through a nationwide experiment using an interactive chatbot. Our findings reveal significant

Table 6: Effects of personalized repayment information on loan application and take-up rates by baseline beliefs

	Applies for loan	Obtains a loan
	(1)	(2)
Repayment info.	0.031***	0.016**
	(0.008)	(0.007)
Repayment info. × Overestimate (baseline belief)	0.059***	0.055***
	(0.012)	(0.010)
Repayment info. × Underestimate (baseline belief)	−0.071***	−0.045***
	(0.010)	(0.008)
Control group mean	0.057	0.041
Controls	Yes	Yes
Observations	60,206	60,206
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 7: Effects of personalized eligibility information on loan application and take-up rates by baseline beliefs

	Applies for loan	Obtains a loan	
	(1)	(2)	(3)
Eligibility info.	0.035***	0.021***	−0.004
	(0.009)	(0.007)	(0.004)
Eligibility info. × Overestimate (baseline belief)	−0.064***	−0.049***	−0.003
	(0.010)	(0.009)	(0.005)
Eligibility info. × Underestimate (baseline belief)	0.028**	0.039***	0.019***
	(0.012)	(0.010)	(0.006)
Applies for loan			0.712***
			(0.002)
Control group mean	0.057	0.041	0.041
Controls	Yes	Yes	Yes
Observations	60,059	60,059	60,059
Note:			*p<0.1; **p<0.05; ***p<0.01

positive effects on loan application and take-up rates, particularly among students who received more personalized information. The intervention increased overall loan application rates by 10.3% and loan take-up rates by 8.1%, with the largest effects observed in the most personalized treatment arms.

The heterogeneity in treatment effects underscores the importance of correcting biased beliefs. Personalized information about loan repayment costs, for instance, significantly influences student decisions based on their initial misconceptions. Students who underestimate loan repayment costs were less likely to apply after receiving accurate information, while those who overestimate costs show a substantial increase in loan take-up. This pattern is consistent across different types of information provided, highlighting the effectiveness of personalized, dynamic interventions in addressing information frictions.

Our results contribute to the growing literature on the role of information in educational decisions, emphasizing the potential of low-cost, scalable interventions to improve the efficiency of financial aid. By leveraging technology to provide tailored information, policymakers can better address the persistent barriers that low-income students face in accessing higher education.

In ongoing work, we aim to investigate the long-term effects of this intervention on college enrollment, choice of institution and major, as well as academic performance and completion. Additionally, we plan to examine labor market outcomes and loan repayment behaviors, providing a comprehensive understanding of the broader impacts of improved information dissemination on educational and economic trajectories.

References

Adamopoulou, Eleni and Lefteris Moussiades (2020). “An Overview of Chatbot Technology.” *Artificial Intelligence Applications and Innovations*. Ed. by Ilias Maglogiannis, Lazaros Iliadis, and Elias Pimenidis. Vol. 584. Cham: Springer International Publishing, pp. 373–383.

Angrist, Joshua D. and Jörn-Steffen Pischke (2009). *Mostly Harmless Econometrics: An Empiricist’s Companion*. 1 edition. Princeton: Princeton University Press. 392 pp.

Bailey, Martha and Susan Dynarski (2011). *Gains and Gaps: Changing Inequality in U.S. College Entry and Completion*. w17633. Cambridge, MA: National Bureau of Economic Research, w17633.

Bettinger, Eric P. et al. (2012). “The Role of Application Assistance and Information in College Decisions: Results from the H&R Block Fafsa Experiment*.” *The Quarterly Journal of Economics* 127.3, pp. 1205–1242.

Bleemer, Zachary and Basit Zafar (2018). “Intended College Attendance: Evidence from an Experiment on College Returns and Costs.” *Journal of Public Economics* 157, pp. 184–211.

Bridgeman, Brent (1992). “A Comparison of Quantitative Questions in Open-Ended and Multiple-Choice Formats.” *Journal of Educational Measurement* 29.3, pp. 253–271.

Camacho, Adriana, Julian Messina, and Juan Pablo Uribe Barrera (2017). “The Expansion of Higher Education in Colombia: Bad Students or Bad Programs?” *SSRN Electronic Journal*.

Carril, Alvaro (2017). “Dealing with Misfits in Random Treatment Assignment.” *Stata Journal* 17.3, pp. 652–667.

Darolia, Rajeev (2016). *An Experiment on Information Use in College Student Loan Decisions*. Working paper (Federal Reserve Bank of Philadelphia) 16-18. Federal Reserve Bank of Philadelphia, pp. 16–18.

DNP (2016). *Declaración de Importancia Estratégica de Identificación de Potenciales Beneficiarios (Sisbén IV)*. 3877. Departamento Nacional de Planeación, p. 60.

Dobronyi, Christopher R., Philip Oreopoulos, and Uros Petronijevic (2019). “Goal Setting, Academic Reminders, and College Success: A Large-Scale Field Experiment.” *Journal of Research on Educational Effectiveness* 12.1, pp. 38–66.

Dynarski, Susan et al. (2021). “Closing the Gap: The Effect of Reducing Complexity and Uncertainty in College Pricing on the Choices of Low-Income Students.” *American Economic Review* 111.6, pp. 1721–1756.

Evans, Brent J. and Angela Boatman (2019). “Understanding How Information Affects Loan Aversion: A Randomized Control Trial of Providing Federal Loan Information to High School Seniors.” *The Journal of Higher Education* 90.5, pp. 800–832.

Ferreyra, Maria Marta et al. (2017). *At a Crossroads: Higher Education in Latin America and the Caribbean*. World Bank, Washington, DC.

Hastings, Justine S. et al. (2016). “(Un)Informed College and Major Choice: Evidence from Linked Survey and Administrative Data.” *Economics of Education Review* 51, pp. 136–151.

Hoxby, Caroline and Christopher Avery (2013). “The Missing "One-Offs": The Hidden Supply of High-Achieving, Low-Income Students.” *Brookings Papers on Economic Activity* 44 (1 (Spring)), pp. 1–65.

Hoxby, Caroline Minter and Kevin M. Stange (2019). *Productivity in Higher Education*. National Bureau of Economic Research Conference Report. Chicago (Ill.): The University of Chicago Press.

Lochner, L. and A. Monge-Naranjo (2016). “Chapter 8 - Student Loans and Repayment: Theory, Evidence, and Policy.” *Handbook of the Economics of Education*. Ed. by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann. Vol. 5. Elsevier, pp. 397–478.

Londoño-Vélez, Juliana, Catherine Rodriguez, et al. (2023). *Financial Aid and Social Mobility: Evidence from Colombia's Ser Pilo Paga*. w31737. Cambridge, MA: National Bureau of Economic Research, w31737.

Londoño-Vélez, Juliana, Catherine Rodríguez, and Fabio Sánchez (2020). “Upstream and Downstream Impacts of College Merit-Based Financial Aid for Low-Income Students: Ser Pilo Paga in Colombia.” *American Economic Journal: Economic Policy* 12.2, pp. 193–227.

Lozano-Rojas, Felipe (2018). “Effects of Cash Subsidies Embedded in Student Loans.” *SSRN Electronic Journal*.

Mackenzie, Lee (2022). “A Case of Capability Expansion or Capability Reduction? An Exploration of Loan Recipients' Views of ICETEX: Colombia's Higher Education Student Loans Institute.” *Journal of Hispanic Higher Education* 21.1, pp. 82–111.

OECD (2016). *Education in Colombia*. Reviews of National Policies for Education. OECD.

– (2022). *Education at a Glance 2022: OECD Indicators*. Education at a Glance. OECD.

Oreopoulos, Philip and Ryan Dunn (2013). “Information and College Access: Evidence from a Randomized Field Experiment.” *The Scandinavian Journal of Economics* 115.1, pp. 3–26.

Riehl, Evan, Juan Saavedra, and Miguel Urquiola (2016). *Learning and Earning: An Approximation to College Value Added in Two Dimensions*. w22725. Cambridge, MA: National Bureau of Economic Research, w22725.

Serrano Mora, Sandra Milena et al. (2023). “Percepciones Sociales Frente al Crédito y El Acceso a La Educación Superior En Colombia: Una Aproximación Polifónica.” *Revista Pedagogía Universitaria y Didáctica del Derecho* 10.2, pp. 227–248.

Smyth, Jolene D. et al. (2009). “Open-Ended Questions in Web Surveys.” *Public Opinion Quarterly* 73.2, pp. 325–337.

Wiseman, William David (2021). *Appraisal Stage Program Information Document (PID) - Support to the Development of Colombia's Integrated Social Registry - P174341*. Program Information Document PIDA235429. The World Bank.

World Bank (2011). *Chile's State-Guaranteed Student Loan Program (CAE)*. 70209. World Bank.

– (2012). *Tertiary Education in Colombia*. Working Paper 75472. World Bank Group.

– (2024). *Investing in Colombia's Future: How Innovative Financing Unlocked Higher Education*. World Bank. URL: `https://www.worldbank.org/en/news/feature/2024/03/01/investing-in-`

colombia-s-future-how-innovative-financing-unlocked-higher-education (visited on 04/19/2024).

Acronyms

2SLS two-stage least squares. 19

HEI higher education institution. 4–6

Icfes the Colombian Institute for the Evaluation of Education. 2, 3, 5, 12, 14, 16, 17

ITT Intent to Treat. 17, 18, 22

IV instrumental variables. 19

LATE local average treatment effect. 19

OLS ordinary least squares. 17

RCT randomized controlled trial.

SES socioeconomic status. 7, 10, 13, 15

TOT Treatment on the Treated. 18

A Tables

Table A.1: Balance Table: Mean Characteristics for Students by Treatment Status

	Population (N=436924)		Sample (N=103831)	
	Mean	Std. Dev.	Mean	Std. Dev.
Individual				
Share Female	0.54	0.50	0.54	0.50
Share Minority Ethnic Group	0.07	0.25	0.05	0.22
Share Foreign	0.01	0.10	0.01	0.10
Share Has Job	0.39	0.49	0.37	0.48
Share Takes Entry Exam*	1.00	0.00	1.00	0.00
Entry Exam Percentile*	49.62	28.98	51.57	28.70
Household				
SES index	2.30	1.37	2.28	1.28
Mother's education (years)	9.93	4.52	10.12	4.48
High School				
Share Public	0.78	0.42	0.77	0.42
Share Rural	0.17	0.38	0.14	0.35
Share Vocational	0.29	0.35	0.29	0.34

Table A.2: Aggregate Statistics of Student Loans in Colombia

Period	New Credit Beneficiaries	Credit Amount	Maintenance Subsidy	Total Disbursed	Disbursements	Avg. per Enrollment
2015-1	35,308	115,440,792,235	9,095,610,831	124,536,403,066	45,831	3,346,922
2015-2	23,813	104,886,565,748	1,220,818,897	106,107,384,645	25,176	4,510,990
2016-1	34,714	164,466,456,512	7,059,229,808	171,525,686,320	43,060	4,790,217
2016-2	22,397	115,655,916,032	4,300,529,958	119,956,445,990	28,026	4,984,316
2017-1	27,195	142,457,832,402	7,819,365,051	150,277,197,453	36,274	5,074,140
2017-2	25,251	142,297,798,151	3,001,293,606	145,299,091,757	29,109	5,245,177
2018-1	14,871	99,768,646,631	2,820,220,661	102,588,867,292	18,228	6,418,088
2018-2	20,534	136,746,286,945	4,801,693,713	141,547,980,658	26,336	6,119,696
2019-1	29,653	175,010,061,932	7,596,379,041	182,606,440,973	38,329	5,750,324
2019-2	10,507	70,112,435,818	1,899,321,910	72,011,757,728	12,775	6,416,845
2020-1	21,951	137,370,773,856	8,605,648,181	145,976,422,037	31,552	6,104,040
2020-2	20,149	116,103,111,868	5,403,146,150	121,506,258,018	26,015	5,707,252
2021-1	24,639	146,619,578,056	9,518,522,490	156,138,100,546	35,075	5,796,358
2021-2	21,545	153,078,482,760	6,721,882,291	159,800,365,051	28,727	6,598,062
2022-1	24,305	173,625,098,559	9,342,003,850	182,967,102,409	33,224	6,717,108
2022-2	24,482	200,016,642,727	8,012,115,349	208,028,758,076	32,094	7,265,038
2023-1	28,535	254,166,220,742	9,419,485,853	263,585,706,595	35,805	8,051,075