Application Mistakes and Information frictions in College Admissions

Anaïs Fabre * Tomás Larroucau† Manuel Martinez‡
Christopher Neilson § Ignacio Rios¶

January 4, 2023

Abstract

We analyze the prevalence and relevance of application mistakes in a seemingly strategyproof centralized college admissions system. We use data from Chile and exploit institutional features to identify a common type of application mistake: applying to programs without meeting all requirements (admissibility mistakes). We find that the growth of admissibility mistakes over time is driven primarily by growth on active score requirements. However, this effect fades out over time, suggesting that students might adapt to the new set of requirements but not immediately. To analyze application mistakes that are not observed in the data, we design nationwide surveys and collect information about students' true preferences, their subjective beliefs about admission probabilities, and their level of knowledge about admission requirements and admissibility mistakes. We find that between 2% - 4% of students do not list their true most preferred program, even though they face a strictly positive admission probability, and only a fraction of this skipping behavior can be rationalized by biases on students' subjective beliefs. In addition, we find a pull-to-center effect on beliefs, i.e., students tend to attenuate the probability of extreme events and under-predict the risk of not being assigned to the system. In addition, we estimate that at least 1% of students would be better off by listing more programs in their application lists. We use these insights to design and implement a large-scale information policy to reduce application mistakes. We find that showing personalized information about admission probabilities has a causal effect on improving students' outcomes, significantly reducing the risk of not being assigned to the centralized system and the incidence of application mistakes. Our results suggest that information frictions play a significant role in affecting the performance of centralized college admissions systems, even when students do not face clear strategic incentives to misreport their preferences.

*Toulouse School of Economics. Email: anais.fabre@tse-fr.eu
†Arizona State University, Department of Economics. Email: Tomas.Larroucau@asu.edu
‡University of Chicago, Harris School of Public Policy. Email: manmart@uchicago.edu
§Princeton University. Email: cneilson@princeton.edu
¶University of Texas at Dallas, Naveen Jindal School of Management. Email: ignacio.riosuribe@utdallas.edu

1 Introduction

Centralized admission systems are widely used in the world. Examples include the school choice systems in NYC, Chicago, Boston, New Haven, Paris, Turkey, Ghana, Chile, and the college admissions systems in Turkey, Taiwan, Tunisia, Hungary, and Chile. The most common allocation mechanism in place is the Deferred Acceptance (DA) Algorithm (Gale and Shapley, 1962), which is known to be strategy-proof for students; that is, students face no incentives to misreport their true preferences when submitting their applications. Even though truthful reporting is a dominant strategy for students under DA, recent evidence has shown that students misreport their preferences (Chen and Sönmez, 2006; Rees-Jones, 2018; Hassidim et al., 2017). One possible explanation is that students behave strategically and consider their beliefs on admission probabilities to decide where to apply (Fack et al., 2019; Larroucau and Ríos, 2018; Chen and Sebastián Pereyra, 2019). Another potential reason is that students do not fully understand the mechanism and cannot identify the optimal strategy, which may explain why low cognitive-ability students are more likely to misreport their preferences (Rees-Jones and Skowronek, 2018). In some cases, misreporting may still be weakly optimal (e.g., if students skip programs where they believe that their admission probability is equal to zero or negligible), but in other cases, misreporting may be a dominated strategy. In the latter case, we say that students make an *application mistake*.

The literature on centralized assignment mechanisms has recently focused on understanding the prevalence and relevance of application mistakes. For instance, Rees-Jones (2018) shows that a significant fraction of residents do not report their preferences truthfully in the National Resident Matching, even though they face no incentives to misreport. In a follow-up paper, Rees-Jones and Skowronek (2018) show that this misreporting behavior may be due to several factors, including students' scores, access to advice and information, and optimism. Artemov et al. (2017) study the Australian college admissions system and find that a non-negligible fraction of students makes obvious mistakes. More specifically, some students apply to programs with both full-fee and reduced-fee options but only include the former in their preference list. Nevertheless, the authors show that the vast majority of these mistakes are payoff irrelevant. Shorrer and Sóvágó (2021) study the Hungarian college admissions process and find a similar pattern. Moreover, they estimate the causal effect of selectivity on making dominated choices, and they show that the prevalence of these mistakes is higher in more selective programs. Finally, Hassidim et al. (2020) analyze the Israeli Psychology Master's Match and show that students often report that they prefer to avoid receiving funding. The authors refer to these as *obvious misrepresentations* and argue that there are other kinds of preference misrepresentation. As in previous studies, the authors find that these mistakes are more common among weaker applicants and argue that this may be due to misunderstanding of the instructions (due to lower cognitive ability) and beliefs that assign low admission probabilities.

To analyze the prevalence and relevance of application mistakes, researchers must overcome significant challenges. First, it is not always clear how to identify application mistakes using administrative data. Without access to data on students' true preferences and subjective beliefs on admission probabilities, researchers typically resort to analyzing unambiguous application mistakes that are idiosyncratic to their settings, achieving little external validity. Second, even if we can identify some application mistakes in the data, assessing their relevance to students' welfare is particularly challenging. To do so, we need to understand the effects of mistakes on outcomes and predict counterfactual behavior twhen students face changes to the system.

Understanding the drivers of students' application mistakes and addressing them—especially if they are payoff-relevant—is still an open question. For instance, recent evidence in school choice systems shows that application mistakes can be driven by families having incorrect beliefs over their assignment probabilities (Bobba and Frisancho (2019); Kapor et al. (2020); Arteaga et al. (2022)). However, we do not know how much biased beliefs contribute to students' college admissions mistakes. Moreover, there could be other potential drivers for student mistakes that have not being explored, such as lack of understanding about the admission and assignment process, information frictions, or even other behavioral biases.[1]

This paper analyzes the prevalence and relevance of application mistakes in the Chilean centralized college admissions system and investigates the effects of information policies to reduce their incidence. The Chilean system uses a variant of the DA algorithm, which allows us to understand the prevalence of mistakes in similar settings worldwide. We exploit two characteristics of the Chilean system to identify the prevalence and relevance of application mistakes. First, a type of application mistake is observed in the administrative data: students can apply to programs even if they do not meet all the admission requirements. We refer to these as *admissibility* mistakes. Second, there is a substantial variation in admission requirements and *admissibility* mistakes over time: the fraction of students who make an *admissibility* mistake has grown from 17% to more than 33% in the last 12 years.

Our results show that the growth of *admissibility* mistakes over time is mainly driven by growth on active score requirements both in the extensive and intensive margins. Although changes in admission requirements over time seem to increase *admissibility* mistakes, this effect fades out over time, suggesting that students adapt to the new set of requirements but not immediately. Also, we find that a significant fraction of students is not aware of their *admissibility* mistakes and does not understand the consequences of making such mistakes, as they believe there is a positive probability of being admitted to those programs. Finally, we find that *admissibility* mistakes are likely welfare-relevant, as close to 25% of students who only list programs with *admissibility* mistakes could have been assigned in the centralized system if they had included programs in which they were eligible.

In addition, we analyze application mistakes that are not directly observed in the administrative data and assess their relevance. We refer to these mistakes as *strategic* mistakes. To achieve this, we design nationwide surveys and collect novel data on students' true preferences for programs, their subjective beliefs about admission probabilities, and their

[1]For instance, Dreyfuss et al. (2019) show that some application mistakes can be rationalized if we account for loss aversion. Taking into account both biased beliefs about admission probabilities and optimization errors, de Haan et al. (2023) find that 8.3% of the secondary-school applicants in Amsterdam make strategic mistakes.

level of knowledge about admission requirements and *admissibility* mistakes. This information also helps us to identify which information frictions are the most relevant to explain students' mistakes and design effective information policies to address application mistakes.

We find that between 2% - 4% of students in our sample do not list their top-true preference, even though they face a strictly positive admission probability and would have unambiguously increased the expected value of their application lists by reporting it as their top preference. Moreover, only a fraction of this skipping behavior can be rationalized by bias on students' subjective beliefs. In addition, we find that students' subjective beliefs are closer to *adaptive* beliefs than *rational expectations* and that students' subjective beliefs are subject to a pull-to-the-center effect, i.e., students' beliefs are biased towards the middle, assigning an attenuated probability to extreme outcomes compared to *rational expectations* beliefs. This pattern implies that students tend to under-predict the risk of not being assigned to the centralized system. Indeed, we estimate that at least 1% of students could have been better off by listing more programs in their application list. In addition, consistent with previous literature, we find substantial differences in the magnitude of the bias depending on students' characteristics, with high score students from private schools having more accurate beliefs than low score students.

Finally, we evaluate the effects of a large-scale outreach intervention designed to decrease information frictions and reduce the incidence of students' application mistakes. In collaboration with MINEDUC and using partial information about students' applications, we created personalized websites with general information about programs included in the student's application list, personalized information on admission probabilities and applications' risk, and personalized recommendations about other majors of potential interest. We randomized the information shown to students to evaluate the effects of reducing information frictions on different margins. We find that showing personalized information about admission probabilities and risk has a causal effect on improving students' outcomes. Students who received safety messages significantly increased their chances of getting assigned to the centralized system (close 50% from their baseline value). Our results suggest that information frictions play a significant role in affecting the performance of centralized college admissions systems, even when students do not face clear strategic incentives to misreport their preferences. Policy interventions that reduce these frictions are then necessary to reduce the incidence of application mistakes.

The paper is organized as follows. In Section 2, we describe the Chilean college admissions system and our sources of data. In Section 3, we define the types of application mistakes that we analyze in the paper: *admissibility* and *strategic* mistakes. In Section 4, we analyze the prevalence, relevance, and drivers of *admissibility mistakes* and analyze their growth over time. In Section 5, we study the prevalence and relevance of *strategic* mistakes and shed light on their potential drivers. In Section 6, we describe the information policy to reduce application mistakes and report the results. Finally, in Section 7 we conclude.

2 Background and Data

2.1 Background

We focus on the centralized part of the Chilean tertiary education system, which includes the 41 most selective universities.[2] From now on, we refer to this as the admission system.

To participate, students must undergo a series of standardized tests (*Prueba de Selección Universitaria* (PSU) until 2020, and *Prueba de Transición* (PDT) starting from 2021). These tests include Math, Language, and a choice between Science or History, providing a score for each of them. The performance of students during high school gives two additional scores, one obtained from the average grade during high school (*Notas de Enseñanza Media* (NEM)), and a second that depends on the relative position of the student among his/her cohort (*Ranking de Notas* (Rank)).

Before the start of the admissions process, the institutions that participate in the admission system must release the number of seats offered by each of their programs,[3] the weights they will consider in each admission factor to compute application scores, and the set of requirements that students must satisfy to be eligible. For instance, some programs require a minimum application score, a minimum average score between the Math and Verbal tests, or require students to take additional specific exams. Some requirements are common to all programs that participate in the admission system (e.g., a minimum average score of Math and Verbal of 450), while others are optional and depend on each program (e.g., some programs require a minimum application score of 450, 500 or 600, while others do not include this requirement). If a student does not satisfy all the requirements imposed by a program, they are not admissible, and thus their chances of admission to that program are equal to zero. In Table 2.1 we show all the admission requirements imposed in the application process of 2019.

Table 2.1: Admission requirements

Requirement	Mistake
Requires High-school GPA (NEM)	Missing NEM, Missing NEM from foreign country
Restricts the number of applications to the Institution of the program	Exceeds the number of applications to the Institution of the program
Restricts province of graduation	Does not satisfy province of graduation
Restricts applicants' gender	Does not satisfy gender restriction
Requires minimum weighted score	Does not satisfy minimum weighted score
Requires special test (exclusion)	Did not take or pass special test (exclusion)
Requires special test (weighting)	Did not take or pass special test (weighting)
Requires a specific year for High-school graduation	Does not satisfy year for High-school graduation
Restricts number of enrollments via Regular Process	Exceeds number of allowed enrollments via Regular Process
Restricts academic qualifications to enroll in the program	Academic qualifications do not allow to enroll in the program
Requires mandatory test of Verbal	Missing score in mandatory test of Verbal
Requires mandatory test of Math	Missing score in mandatory test of Math
Requires History and Social Sciences test	Missing score in History and Social Sciences
Requires Sciences test	Missing score in Sciences
Requires minimum average score Math-Verbal	Does not satisfy minimum average score Math-Verbal
Requires either History and Social Sciences test or Sciences test	Did not take History and Social Sciences test nor Sciences test
Requires minimum average score Math-Verbal ≥ 450	Average score Math-Verbal is below 450
Requires minimum weighted score for special test (weighting)	Does not satisfy minimum weighted score for special test (weighting)
Requires Education prerequisites	Does not meet Education prerequisites

[2]See Larroucau and Rios (2021) for a more general description of tertiary education in Chile and more institutional details.

[3]Students apply directly to programs, i.e., pairs of university-major.

After scores are published, students can access an online platform to submit their applications, where they can list up to ten programs in decreasing order of preference. We refer to these lists as Rank Order Lists (ROLs), and in Section 2.1.1 we discuss more details about the application process. DEMRE collects all these applications, checks students' eligibility in each of their listed programs and, if eligible, computes their application scores and sorts them in decreasing order. Then, considering the preferences of students and the preferences and vacancies of programs, DEMRE runs an assignment algorithm to perform the allocation. The mechanism is a variant of the DA algorithm, where ties on students' scores are not broken.[4] As a result, the algorithm assigns each student to at most one program, and programs may exceed their capacities only if there are ties for their last seat. We refer to the score of the last admitted student as the *cutoff* of each program.

It is important to highlight that, due to the large nature of the market, students do not face strategic incentives to misreport their preferences when the constraint on the length of the list is not binding (Rios et al., 2020). However, the empirical evidence shows that some students still misreport their preferences, even when this constraint is not binding. As discussed in (Larroucau and Ríos, 2018), it may be weakly optimal for students to misreport their preferences if they face degenerate admission probabilities. Moreover, when the constraint on the length of the list is binding might also be strategizing (Haeringer and Klijn (2009)). In both cases, the information provided by previous years' cutoffs could be relevant for students to form correct beliefs about their admission probabilities (Agarwal and Somaini (2018)) and avoid application mistakes due to biases in their beliefs.

2.1.1 Information Access.

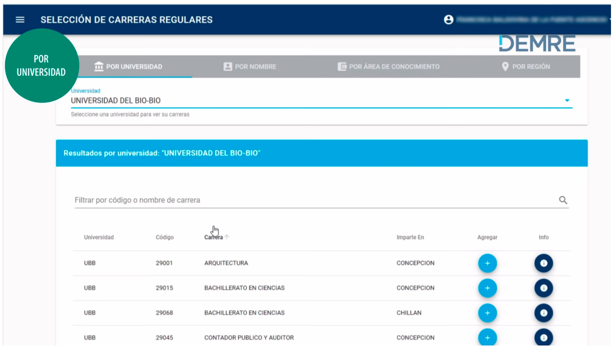
Although the information about programs' seats, weights, requirements, and past cutoffs is public, no platform collects and summarizes it for students. Instead, each institution publishes its information and, in many cases, they do not display all the relevant details on the same website. As a result, it is hard for students to collect all the relevant information and compare programs before starting the application process.

Part of this information—namely, the application scores and whether the student satisfies the requirements imposed by each program—is included in the platform that students use to submit their application. More specifically, the platform displays three types of information:

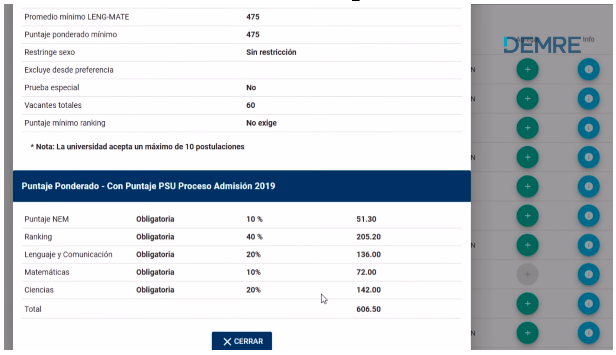
1. **Academic information:** students receive information about their scores, high-school grades, and other academic credentials.
2. **Information about programs:** students can search for information about the programs' characteristics and requirements, as illustrated in Figures 2.1a and 2.1b.

[4]See Rios et al. (2020) for a detailed description of the mechanism used and its properties.

(a) Searching stage



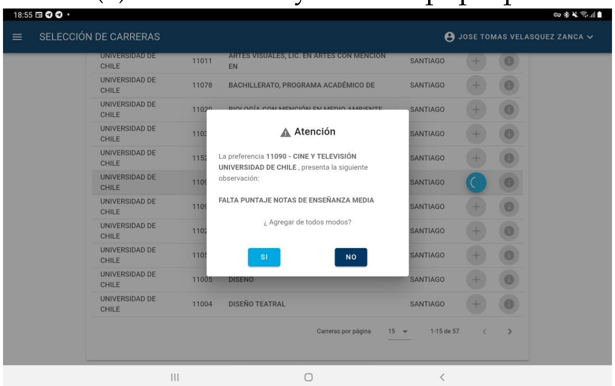
(b) Admission requirements



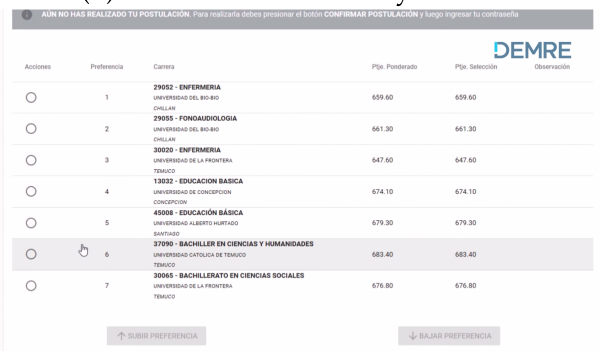
3. **Information about application:** for each of the programs included in their list, students can see their application score and whether they satisfy the requirements imposed by the program.

   Starting from 2019, DEMRE includes a message to warn students if they do not meet an admission requirement when adding a program to their application list, as illustrated in Figure 2.2a, specifying the admission requirements not satisfied by the student while students are adding and sorting their options, as shown in Figure 2.2b.

(a) Admissibility mistake pop-up



(b) Potential admissibility mistake



Even though DEMRE displays precise information about admission requirements, it still allows students to include programs for which they do not meet the admission requirements. As we will show in Section 5, this feature contributes towards generating confusion and introducing biases on students' beliefs. Moreover, the system does not provide information about cutoffs in previous years or students' admission probabilities, potentially increasing the biases on students' beliefs.

2.2 Data

We combine a panel of administrative data on the admissions process with two novel datasets that we collect to analyze students' mistakes. We now provide details on each of these data sources.

ADMISSIONS PROCESS. To characterize the historical evolution of the admissions process and how it affects mistakes, we combine information on the admissions processes from 2004 to 2020. This dataset includes information about students (socio-economic characteristics, scores, and applications), programs (weights, seats available, and admission requirements), and also the results of the admissions process (i.e., for each student and each program they applied to, whether the application was valid, and whether the student was assigned to that program or wait-listed).

SURVEYS - 2020 - 2022. In 2020, 2021, and 2022, we designed and conducted three nationwide surveys to gather information on students' preferences for programs and their beliefs on admission probabilities. In these surveys, we ask students about their beliefs on the cutoffs and their admission probabilities for the programs included in their application list and their top-true preference (even if they did not include it in their list of preferences). These surveys were sent to students after the application process and before the assignment results were published. As shown in Section 5, we use this information to evaluate whether biased beliefs explain the application mistakes that we observe in the data.

INTERVENTIONS- 2021 & 2022. In collaboration with MINEDUC, DEMRE, and ConsiliumBots, we designed and implemented an intervention to evaluate whether information provision can help to reduce application mistakes. As previously mentioned, we created a personalized website for each student and randomized the information included to measure the effect of different types of information on their chances of making a mistake. Then, by comparing students' application lists before and after the intervention, we can assess the effect of the information displayed on different outcomes, such as their probability of making a mistake and their probability of being admitted in the system, among others. In Section 6.1 we describe the 2022 version of this intervention in detail.

3 BACKGROUND

Consider a finite set of students N and a finite set of programs M. Each student $i \in N$ is characterized by a vector of indirect utilities $u_i \sim f_u$, a vector of scores $\vec{s}_i = \left\{s_i^k\right\}_{k\in\mathcal{K}}$, where $\mathcal{K}$ is a set of admission factors considered in the application process, and a submitted list of preferences $R_i \in \mathcal{R}$, where $\mathcal{R}$ is the set of all possible rank-ordered lists. Each program $j \in M$ is characterized by its number of vacancies $q_j \in \mathbb{N}_+$, by a vector of admission weights $\omega_j = \left\{\omega_j^k\right\}_{k\in\mathcal{K}}$, and by a set of eligibility rules that define whether a student is admissible. Let $A_j \subseteq N$ be the set of students that satisfy these additional requirements and thus are admissible in program j.

The application score of a student $i \in A_j$ in program j, s_{ij}, is given by:[5]

$$s_{ij} = \sum_{k \in \mathcal{K}} \omega_j^k s_i^k. \tag{3.1}$$

These application scores are used by programs to rank their applicants in decreasing order. Let $\bar{s}_j$ be the application score of the last admitted student to program j; we refer to it as the cutoff. Let $p_i \in [0,1]^M$ be the vector of rational-expectations admission probabilities of student i, i.e., for each $i \in N$ and $j \in M$, $p_{ij} = \mathbb{P}(s_{ij} \geq \bar{s}_j)$. Similarly, let $\tilde{p}_i \in [0,1]^M$ be the vector of subjective beliefs on admission probabilities for student i. We now formalize the different types of mistakes.

Definition 1 (Application mistake)**.** Given u and p, $R \in \mathcal{R}$ involves an *application mistake* if $\exists R' \in \mathcal{R} \setminus R$ such that reporting R' weakly dominates reporting R in expected utility, i.e.,

$$EU(R'|u,p) \geq EU(R|u,p)$$

Definition 2 (Obvious mistake)**.** $R \in \mathcal{R}$ involves an *obvious mistake* if $\exists R' \in \mathcal{R} \setminus R$ such that reporting R' weakly dominates–in expected utility–reporting R for any $u \in \text{supp}(f_u)$ and $p \in [0,1]^M$, i.e.,

$$EU(R'|u,p) \geq EU(R|u,p) \, \forall u \in \text{supp}(f_u), p \in [0,1]^M$$

First, notice that obvious mistakes are a special case of application mistakes, in which there exists an alternative ROL R_i' that dominates R_i for all possible utilities and admission probabilities. Second, the concept of *weakly dominated* implies that mistakes may or may not be welfare relevant. Third, as Definition 1 considers both rational expectations beliefs and expected utility maximization, mistakes might be explained by behavioral reasons (without departing from rational expectations),[6] or by biased beliefs (without departing from rationality).

Given their empirical relevance, we focus on two types of mistakes: (1) *admissibility* mistakes, which are a special case of *obvious mistakes*; and (2) *strategic* mistakes, which are not obvious mistakes but play an important role in the Chilean system. Further, we separate *strategic* mistakes in (i) *underconfidence*, (ii) *overconfidence*, and (iii) *ordering* mistakes.

Definition 3 (Admissibility mistake)**.** Program $j \in R_i$ for $R_i \in \mathcal{R}$ involves an *admissibility mistake* for student $i \in N$ if $i \notin A_j$

Notice that an *admissibility* mistake is a particular case of an *obvious mistake* because students face zero admission probability in a program where they are not admissible; thus,

[5]Without loss of generality, we assume that $s_{ij} = 0$ for $i \notin A_j$.

[6]Several behavioral models could fit this definition, such as bounded rationality, non-expected utility maximization, among others.

regardless of students' preferences or beliefs, not including programs with admissibility mistakes weakly dominates including them in the application list. This type of application mistake is observed in the Chilean setting, allowing us to analyze its drivers and relevance (see Section 4.1). To analyze application mistakes that are not *obvious* mistakes, we exploit the data collected in the surveys. We label these mistakes as *strategic* mistakes and analyze them in Section 5.

Definition 4 (Underconfidence mistake). $R_i \in \mathcal{R}$ involves an *underconfidence mistake* for student $i \in N$ if $\exists j' \notin R_i$ such that $p_{ij'} > 0$, $u_{ij'} > \min_{j \in R_i} \{u_{ij}\}$ and

$$EU(R_i \cup \{j'\} \,|u, p) > EU(R_i|u, p).$$

Given a ROL R and admission probabilities $\vec{p_i}$, let $\Pi(R_i)$ be the probability that student i results unassigned, i.e.,

$$\Pi(R_i, p_i) = \prod_{j \in R_i} (1 - p_{ij})\,.$$

We refer to $\Pi(R_i, p_i)$ as the *risk* of submitting a ROL R_i given admission probabilities $\vec{p_i}$.[7]

Definition 5 (Overconfidence mistake). $R_i \in \mathcal{R}$ involves an *overconfidence mistake* for student $i \in N$ if $\exists j' \notin R_i$ such that $u_{ij'} > 0$ and

$$\Pi(R_i \cup \{j'\}\,, \vec{p_i}) < \Pi(R_i, \vec{p_i}).$$

Finally, we also analyze ordering mistakes, whereby students may obtain a higher expected utility by simply changing the order of the programs listed in their ROL.

Definition 6 (Ordering mistake). $R_i \in \mathcal{R}$ involves an *ordering mistake* for student $i \in N$ if $\exists R_i' \in \mathcal{R} \setminus R_i$ such that $\{j\}_{j \in R_i} = \{j\}_{j \in R_i'}$ and

$$EU(R_i'|u, p) > EU(R_i|u, p).$$

4 Admissibility Mistakes

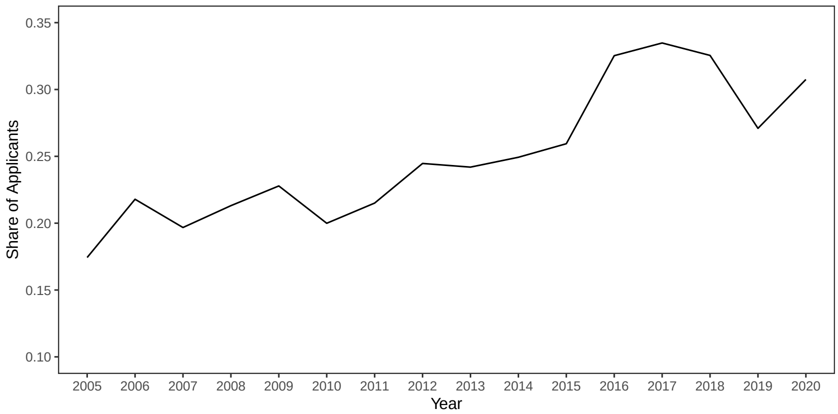
In this section, we focus on admissibility mistakes. As previously discussed, we say that a student makes an *admissibility mistake* if she includes a program in her preference list for which she does not fulfill all the requirements and, thus, her admission probability to that program is equal to zero. We first explore the prevalence and growth of these mistakes over time, the drivers and causes, and then we analyze their relevance for welfare.

[7]This definition of risk assumes independence of admisssion probabilities across programs.

4.1 Prevalence, Growth, and Drivers

In Figure 4.1 we show the evolution of the share of students with at least one *admissibility* mistake between 2005 and 2020. We observe a high increase in the fraction of students with at least one admissibility mistake. Indeed, this fraction has almost doubled in the last 12 years (from close to 17% to more than 33% in 2017).

Figure 4.1: Share of students with *admissibility* mistakes in their ROL



Notes: The share is computed as the total number of students who submitted a ROL with at least one *admissibility* mistake, over the total number of applicants.

4.1.1 Evolution of requirements

Time series of mistakes. To analyze whether changes in the admission requirements over time can have an effect on *admissibility* mistakes, we run a time series analysis on the share of *admissibility* mistakes by program and year. To accomplish this we consider the following specification:

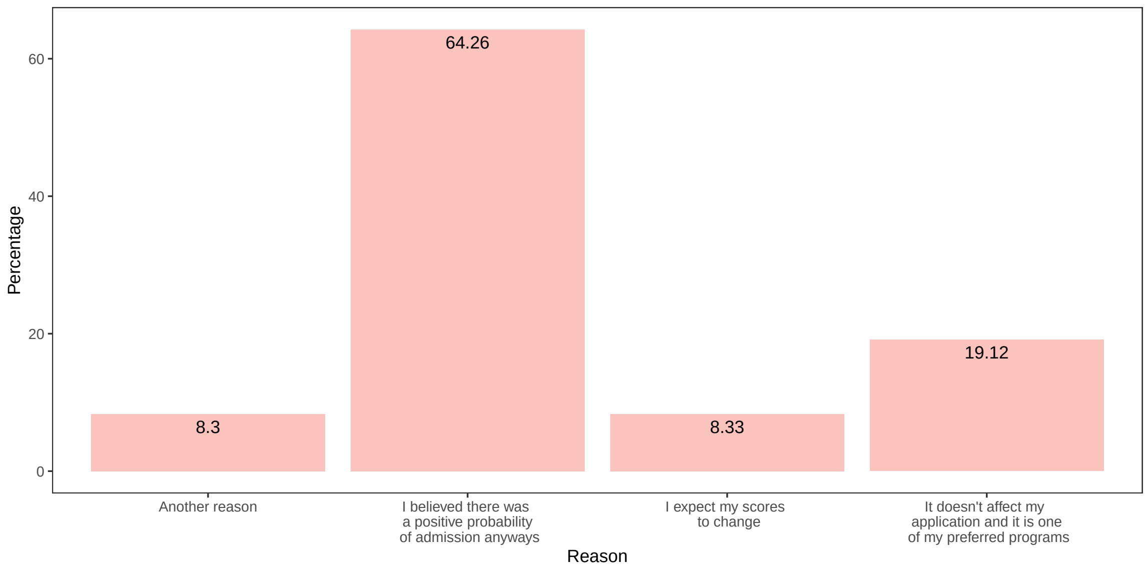
$$z_{jt} = \alpha_j + \lambda_t + \beta_1 z_{jt-1} + \beta_2 z_{jt-2} + \beta_3 \Delta_{jt} + \varepsilon_{jt} \tag{4.1}$$

where z_{jt} is the share of *admissibility* mistakes by program j in year t; α_t and α_j are time and program fixed-effect, respectively; $\Delta_{jt} = \left\{\Delta^+_{jtl}, \Delta^-_{jkl}\right\}_{l\in\mathcal{L}}$ is a matrix of dummy variables, where $\Delta^+_{jtl} = 1$ if program j increased the admission requirement l in period t, and $\Delta^+_{jtl} = 0$ otherwise; similarly, $\Delta^-_{jtl} = 1$ if program j decreased the admission requirement l in period t, and $\Delta^-_{jtl} = 0$ otherwise. We also include lags for the variables Δ^+_{jtl} and Δ^-_{jtl} to capture the evolution of the effect of the change in requirements over years. Finally, ε_{jt} is an i.i.d shock.

Table 4.1 shows the estimation results. We observe that increasing an admission requirement increases the share of *admissibility* mistakes. Depending on the requirement, the effect ranges form 3.3% (Min Math-Verbal) to 4.7% (limiting the position of programs in the ROL). On the other hand, reducing the admission requirements decreases significantly the share of admissibility mistakes (from 2.3% to 5.1%). In addition, we observe that the lag variables of the changes in the admission requirements are consistent in sign, and their magnitude is decreasing over time. For instance, increasing the minimum Math-Verbal requirement increases by 4.04% the share of mistakes in the current year, by 0.76% in the following year, and by 0.27% two years later. These results are consistent with students having adaptive beliefs about admission requirements, i.e., a share of students who make *admissibility* mistakes might be unaware of the changes in requirements in the current year, but this share decreases as time goes by. Under this hypothesis, students might adapt to changes in the rules of the admission process, but this adaptation is not immediate. The lack of immediate awareness of students about admission requirements suggests that changes in admission requirements can introduce a negative externality in the centralized system. If *admissibility* mistakes are welfare-relevant, this externality could affect students' outcomes.

Awareness. To understand the level of awareness of students about their *admissibility* mistakes and how they interpret the information about admission requirements, we leverage the information elicited through the survey implemented by DEMRE in 2020. Overall, 86% of respondents who made an *admissibility* mistake declare to be aware of it at the time of applying. Figure 4.2 shows the reasons why students applied with an *admissibility* mistake conditional on being aware of it. We observe that the majority (close to 64% of students) think that there is a positive probability of admission to a program with an *admissibility* mistake. This lack of understanding about the rules of the admission system could be payoff relevant in some cases. For instance, if a student does not apply to feasible programs besides her application with an *admissibility* mistake, she faces zero probability of admission to the centralized system.

Figure 4.2: Reasons for making admissibility mistakes conditional on being aware



On the other hand, we observe that close to 14% of respondents who made an admissibility mistake declare not to be aware of it when submitting their application.

In Appendix A.1, we analyze which are the specific requirements that the respondents

Table 4.1: Effect of Changes in Admission Requirements

	(1)	(2)	(3)	(4)
Min. average score (P0)	0.029***	0.037***	0.034***	0.040***
	(0.006)	(0.008)	(0.005)	(0.006)
Min. average score (N0)	-0.020**	-0.028***	-0.043***	-0.044***
	(0.007)	(0.007)	(0.004)	(0.005)
Min. application score (P0)	0.019***	0.026***	0.036***	0.035***
	(0.005)	(0.007)	(0.005)	(0.006)
Min. application score (N0)	-0.013*	-0.017*	-0.023***	-0.029***
	(0.006)	(0.009)	(0.006)	(0.007)
Special test (N0)	-0.075	-0.130***	-0.304***	-0.425***
	(0.063)	(0.040)	(0.092)	(0.058)
Restricts application rank (P0)	0.056**	0.015	0.047***	0.013
	(0.022)	(0.025)	(0.011)	(0.019)
Restricts application rank (N0)	-0.021	-0.038**	-0.051***	-0.059***
	(0.018)	(0.014)	(0.008)	(0.007)
Min. average score (P1)		0.015***		0.008**
		(0.004)		(0.003)
Min. average score (N1)		-0.027***		-0.024***
		(0.008)		(0.006)
Min. average score (P2)		0.011**		0.003
		(0.004)		(0.002)
Min. average score (N2)		-0.015**		-0.010*
		(0.006)		(0.005)
Min. application score (P1)		0.017**		0.011*
		(0.006)		(0.005)
Min. application score (N1)		-0.007		-0.007*
		(0.006)		(0.004)
Min. application score (P2)		0.008**		0.001
		(0.004)		(0.005)
Min. application score (N2)		-0.008		-0.004
		(0.006)		(0.004)
Special test (N1)		-0.102		-0.139**
		(0.067)		(0.051)
Special test (N2)		-0.092		-0.057
		(0.084)		(0.065)
Restricts application rank (P1)		0.066**		0.051**
		(0.026)		(0.019)
Restricts application rank (N1)		-0.035***		-0.029***
		(0.010)		(0.006)
Restricts application rank (P2)		0.009		-0.011
		(0.013)		(0.011)
Restricts application rank (N2)		-0.028		-0.015
		(0.017)		(0.011)
Share mistakes (1)			0.466***	0.422***
			(0.026)	(0.039)
Share mistakes (2)			0.082***	0.127***
			(0.018)	(0.022)
Program	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
Lags - Dependent	No	No	Yes	Yes
Lags - Others	No	Yes	No	Yes
Observations	18,951	14,814	16,799	14,814
R^2	0.839	0.873	0.895	0.904
Within R^2	0.058	0.097	0.312	0.316

Note: P0 (N0) represents the variables Δ^{+}_{jtl} (Δ^{-}_{jtl}), while P1 and P2 (N1, N2) capture the first and second lags of these variables. Standard errors clustered at the program and year level reported. Significance: $^{*}p < 0.1;^{**} p < 0.05;^{***} p < 0.01$

know and do not know. Overall, we observe heterogeneity in the level of knowledge by requirement type and significant differences between the groups of students who did and did not make an *admissibility* mistake. Indeed, among the students who did not make an *admissibility* mistake, between 60% to 75% declare to know the requirements of minimum scores and specific tests. In contrast, this number is between 59% to 63% among students who made an *admissibility* mistake. In addition, we observe that students who made an *admissibility* mistake are significantly less correct about programs' vacancies (17% compared to 28%). However, we do not observe substantial differences for other requirements.

In summary, there is poor understanding of the admission requirements and, as expected, students who make *admissibility* mistakes tend to be less aware of these requirements than students who do not make mistakes. This fact suggests that *admissibility* mistakes might be payoff relevant if they are driven by a lack of understanding about admission requirements.

4.2 Relevance

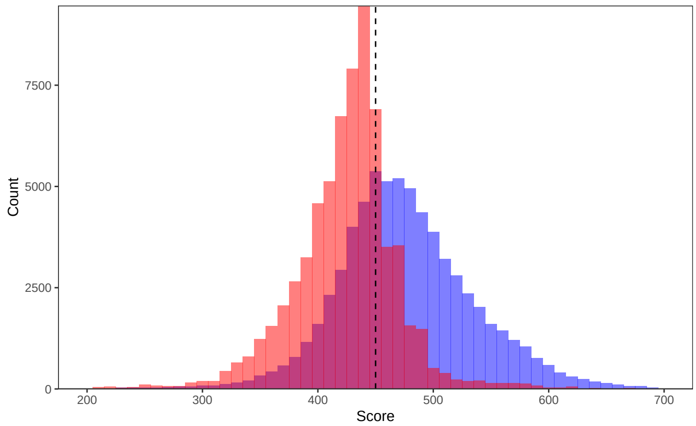
In this section, we analyze the relevance of *admissibility mistakes*, i.e., whether making this type of mistake can affect students' outcomes and welfare.

Admissibility mistakes could be payoff relevant for several reasons. First, since the Chilean system only allows students to apply to at most ten programs, making an admissibility mistake results in a wasted preference, which could potentially limit students' chances of applying to other programs where they are admissible. Second, even for students who apply to less than the maximum number of programs allowed, a high fraction of admissibility mistakes reflects a poor understanding of how the application process and the assignment mechanism work, affecting how students decide to apply.

To analyze the relevance of admissibility mistakes, we consider the probability of being assigned to the centralized system as a proxy for welfare. Even though this is not a precise measure for welfare, resulting assigned to a program can have a significant impact on students' future outcomes due to the high returns of higher education (Rodriguez et al., 2016). As previously discussed, not all admissibility mistakes are payoff relevant. For instance, if a student applies to a program where she faces low admission probability (*reach* program) and then includes programs with high admission probability (*safety* programs), making an admissibility mistake would have no impact on her admission chances.

One case in which admissibility mistakes are likely to be payoff relevant is when they affect all student applications. In 2020, among the students that applied to at least one program (146,438), 18,586 students made application mistakes in all their submitted preferences. In Figure 4.3 we plot the distribution of the average score between Math and Verbal (in red) and the application scores (in blue) among students who made mistakes in all their applications. We observe that 25.05% of these students have scores that would enable them to be admissible in some programs, which would be enough for them to be admitted. Hence, these students could have applied to different programs and being

Figure 4.3: Distribution Average Math-Verbal and Weighted Score for Students with All Mistakes



assigned to the centralized system.

Discussion. Welfare-relevant mistakes may also be present among students that submit valid applications. For instance, students may include valid applications but may result unassigned, and thus not including more valid preferences prevented them from obtaining a better assignment. However, it is difficult to estimate the causal effect of an admissibility mistake on the probability that the student is assigned, as students who make admissibility mistakes may not be comparable to those who do not.

For this reason, it may be the case that two students with similar scores and observable characteristics but different eligibility statuses differ on unobservable characteristics that push them to apply. For instance, students who make admissibility mistakes may have a lower understanding of the system's rules than students with similar characteristics but who do not make admissibility mistakes. Then, identifying the effect of the mistake from the unobservable characteristics is not possible using observational data.

Overall, although we cannot directly estimate the effect of admissibility mistakes, we know that an important fraction of students is not aware of their mistakes. Also, a significant fraction of students make admissibility mistakes in all their applications when they could have included programs for which they are eligible. Hence, we conclude that admissibility mistakes play an important role, and reducing their incidence is a relevant goal that can be achieved by providing students more information.

5 Strategic Mistakes

In this section, we focus on *strategic mistakes*. As discussed in Section 3, we focus on three types of strategic mistakes:

1. *Under-confidence*: students make an under-confidence mistake if, despite having valid applications, they do not apply to their top-true choice as their top-reported

preference, even though their score is high enough to be admitted with positive probability and the constraint in the length of the list is not binding.

2. *Over-confidence*: students make an over-confidence mistake if, despite having valid applications, they do not apply to programs they: (i) prefer to be unassigned and (ii) face a positive probability of assignment, even though they face a positive probability of being unassigned to the centralized system and the constraint in the length of the list is not binding.

3. *Ordering*: students make an ordering mistake if they do not rank programs with a positive admission probability in decreasing order of utility. As a result, the student would benefit from submitting a ranked ordered list with the same subset of programs but in different order.

Notice that these types of mistakes are, by definition, payoff relevant. In addition, to properly analyze these mistakes, we need to understand students' application behavior and, more specifically, how they form their beliefs on admission probabilities and the expected utilities from attending each program. For this reason, we start by characterizing the application behavior of students and their subjective beliefs. Then, we document the prevalence and relevance of strategic mistakes and analyze their main drivers.

5.1 Application behavior

As part of our surveys, we ask students about their most desired program, aiming to elicit their top-true preference and to understand their application behavior.[8] This question allows us to classify students into three groups: (i) *Truth-tellers*, i.e., students who include their top-true preference as their top-reported preference in their application list, (ii) *Misreporting Exclusion*, i.e., students who do not include their top-true preference in their list, and (iii) *Misreporting Ordering*, i.e., students who include their top-true preference in their list but not as their top reported preference.

To properly classify students into these groups, we analyze the reasons why students did not include their top-true preference as top-reported preference. Table A.1 in Appendix A.2, shows the reasons students give to not list their top-true preference as top-reported preference. We observe that a significant fraction of students give inconsistent answers to this question. For instance, close to 14% of *truth-tellers* do not declare to have listed their top-true preference as top-reported preference. In addition, a significant fraction of students who are classified as *misreporting exclusion* or *misreporting ordering* declare to not list their top-true preference as top-reported preference because they do not have the monetary resources to pay for that program (26% and 20%, respectively). However, the

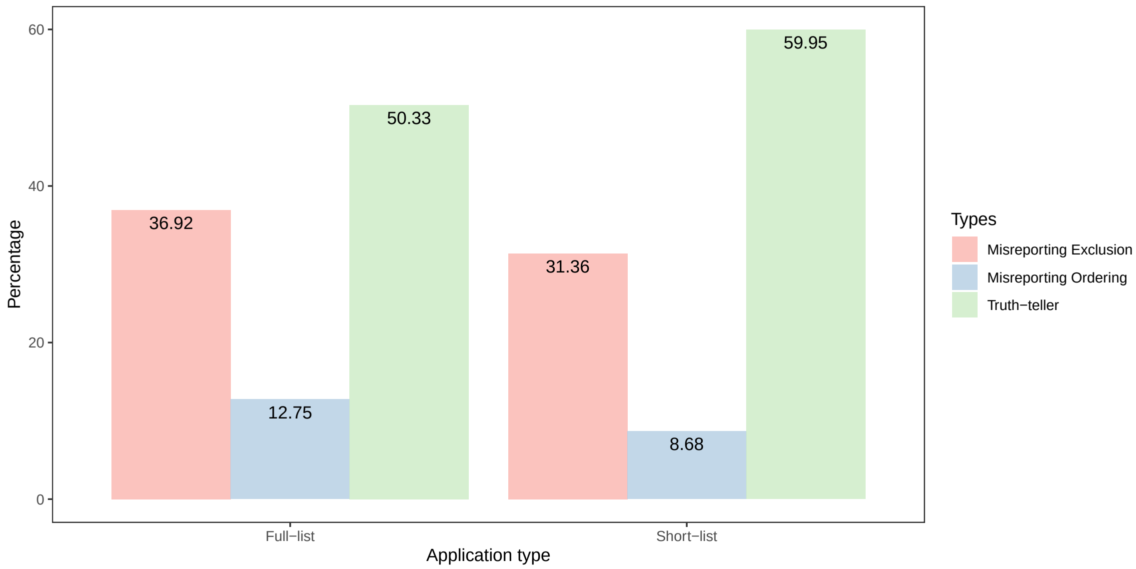
[8] In particular, we ask students the following question:
This question aims to know where you would have applied to in the hypothetical case in which your admission did not depend on your scores. We remind you that this is only a hypothetical question and will not affect your application or admission probabilities. If the Admissions Process did not depend on your PSU scores, nor your NEM or Ranking scores. To which program would you have applied?

survey question we are analyzing is intended to elicit students' ideal program taking into account their monetary costs. To avoid over-estimating the share of students who misreport their preferences, we consider only students who give consistent answers regarding their application type.[9]

Figure 5.1, shows the percentage of students in each group who give consistent answers. We further divide these groups between *short-list* (students who report less than 10 programs) or *full-list* (students who list exactly 10 programs). We observe that, among *short-list* students (88% of applicants), close to 60% of applicants report their top-true preference as their top-reported preference, and 31% exclude this program from their application list. This statistic contrasts to the close to 50% for *full-list* students who include their top-true preference as their top-reported preference. A potential explanation for these differences is that students who submit full lists might face strategic incentives to exclude their top-true preferences if their beliefs assign a low admission probability to that program.

In addition, we observe that a significant fraction of students misreports the order of their top-true preference (*Misreport Ordering*). This percentage is close to 8% for *short-list* students, while it is close to 13% for full-list students.

Figure 5.1: Application types



5.1.1 Subjective beliefs

We now characterize students' subjective beliefs on admission probabilities. We ask students in the survey about their beliefs over the realization of cutoff scores and the probability they assign to their application score being above the cutoff score for every program

[9] We consider as inconsistent answers, students who are classified as *truth-tellers* and do not give reason (a) or give reasons (c) or (d), and students who are classified as *misreporting exclusion* or *misreporting ordering* and give reason (a) or reasons (c) or (d).

in their application list.[10]

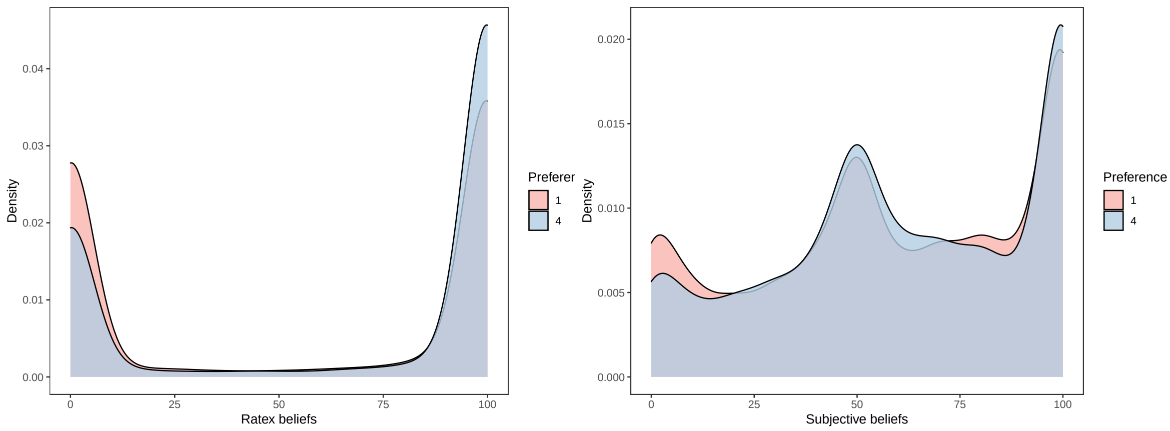
Rational expectations and biased beliefs. To understand if students have correct beliefs regarding their admission chances, we compute their Rational expectation beliefs (*Ratex*) for every listed preference and also for their top-true preference. *Ratex* beliefs are computed following the approach described in Larroucau and Ríos (2018)[11].

Figure 5.2a shows the distribution of *Ratex* beliefs for the first and fourth reported preferences. We observed peaks around 0% and 100%, with little mass in the middle of the distribution's support. This pattern is explained by the fact that a significant fraction of students faces almost degenerate admission probabilities for the programs listed in their application lists.

Figure 5.2b shows the distribution of subjective beliefs for the first and fourth reported preferences. We observe peaks at 0%, 50%, and 100%, and we also observe a significant mass between these points of the distribution. The mass at 50% suggests that students' subjective beliefs could be subject to a pull-to-the-center effect, i.e., students' beliefs are biased towards the middle, assigning an attenuated probability to extreme outcomes compared to *Ratex* beliefs.[12]

Figure 5.2: Distribution of Ratex and subjective beliefs by preference

(a) Ratex beliefs (b) Subjective beliefs



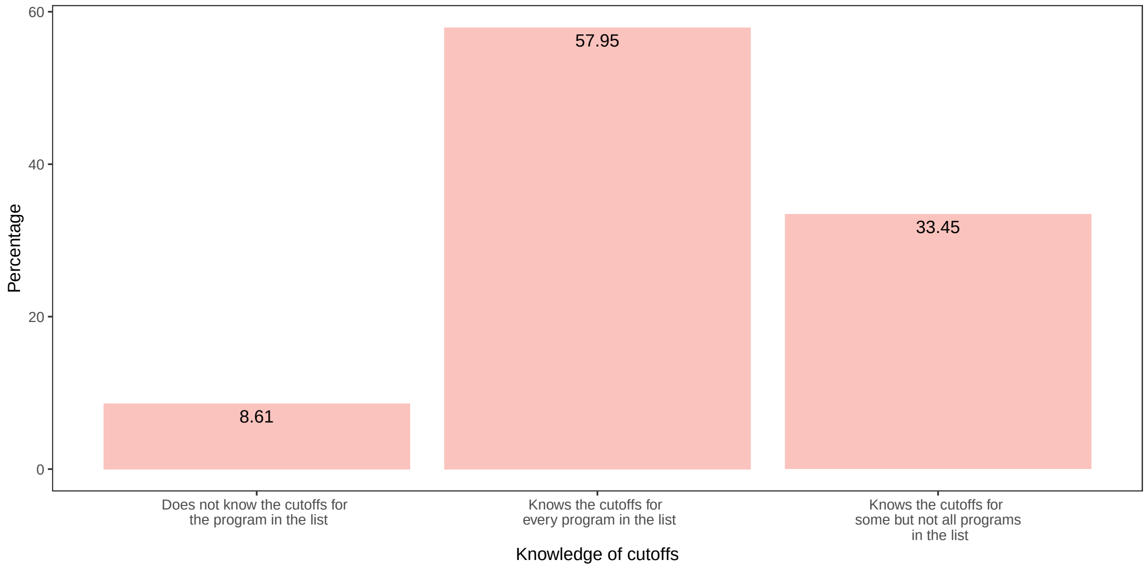
[10]In particular, we asked the following question:
We show you now a list of the programs you applied to, in strict order of preference. For each of them, please tell us which do you think will be the value of the cutoff score for the CURRENT Admission Process and how likely do you think your application score will be above the cutoff score. We remind you that this is only a survey, and it DOES NOT affect in any way your application nor your admission probabilities. What do you think will be the value of the cutoff score for the current Admission Process for each of these programs?
How likely do you think your application score for the following programs will be above the current admission process's cutoff score?
On a scale from 0 to 100, where 0 is "completely sure that your application score WILL NOT be above the cutoff score for this program" and 100 is "completely sure that your application score WILL BE above the cutoff score for this program".

[11]The only difference to the approach followed by Larroucau and Ríos (2018), is that after obtaining the marginal distribution of cutoffs for every program, we smooth these distributions by fitting Truncated Normal distributions doing a standard MLE procedure.

[12]Similar pull-to-center effects are found in more general belief elicitation tasks, and also in newsvendor problems (see Bostian et al. (2008)).

Knowledge of cutoffs. To understand whether knowledge about cutoff scores could explain biased beliefs, we ask students in the survey whether they know the value of cutoff scores for the previous year.[13] Figure 5.3 shows students' knowledge level about the previous year's cutoff scores. Close to 58% declare to know the previous year's cutoffs for all of the programs listed in their application list. In contrast, close to 9% declare to ignore the previous year's cutoffs of all the programs listed in their application.[14]

Figure 5.3: Knowledge of previous year's cutoff scores



Although previous year cutoffs are informative for the current process, cutoffs are random variables that may vary from year to year. To assess if students know this and understand how they use past information to build their beliefs, we ask them to predict the expected cutoff for the current admission process for every program listed in their application list. On the one hand, Figure 5.4a shows the distribution of the difference between the standardized expected cutoff (subjective) and the standardized realized cutoffs (*Ratex*) by position in the preference list. First, we observe that bias distributions are centered around zero. Second, we observe that students tend to be more accurate about the expected cutoffs of their top-reported preferences, as the distributions become significantly more spread for programs listed in lower reported preferences. This heterogeneity implies that there is a significant fraction of students with high positive bias, and a significant share with a high negative bias. We refer to these groups of students as *pessimistic* and *optimistic*, respectively. On the other hand, Figure 5.4b shows that the distribution of

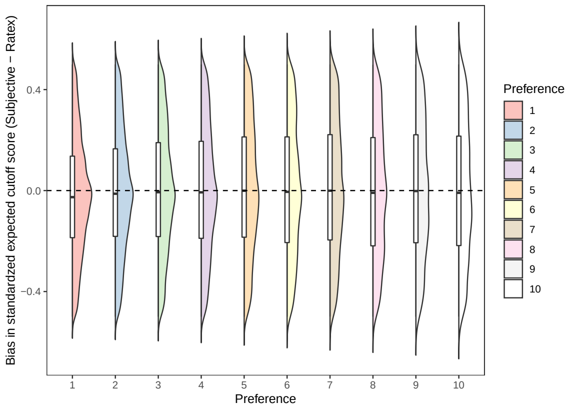
[13]In particular, we ask the following question: *It is referred to a cutoff score as the application score of the last admitted students to a given program. Each student is assigned to the highest reported preference for which her application score is greater than or equal to the cutoff score that realizes in the current Admission Process. Do you know which was the cutoff score for the PREVIOUS YEAR for each of the programs you applied to?*

[14]DEMRE does not provide any information about programs' cutoffs during the application process. However, this information can be typically found on universities' websites. One reason behind the lack of centralized information about cutoff scores is the concern that some students might not understand what a cutoff score exactly means. For instance, they might believe that cutoffs are predetermined by programs and do not understand that they may vary from year to year. This discussion stresses the importance of providing not only information that is necessary for students to forecast their admission chances but also to educate them about the meaning of this information.

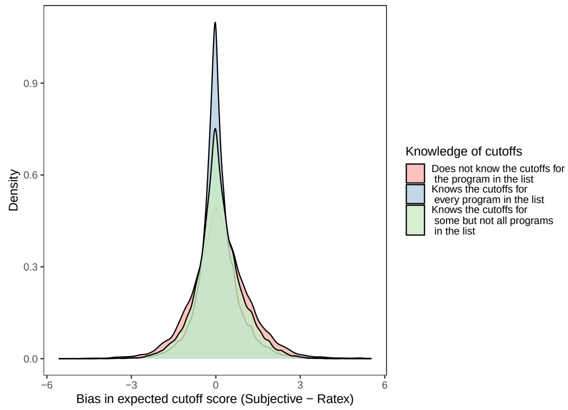
bias is more spread for students who do not know the previous-year cutoffs for some or all of the programs in their lists. This pattern suggests that giving information to students about previous year cutoff scores could be an effective policy to decrease their bias.

Figure 5.4: Distributions of bias in standardized expected cutoff

(a) By preference



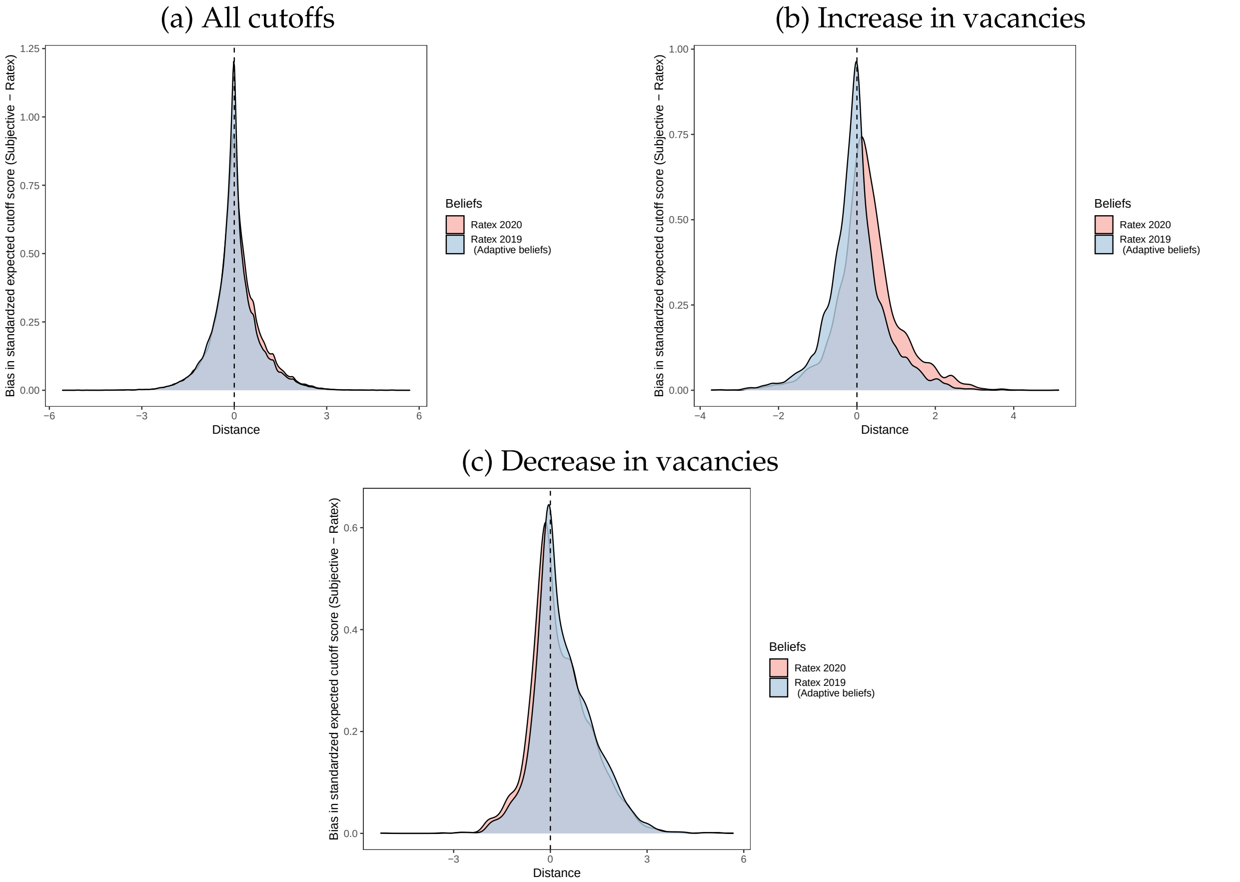
(b) By knowledge of cutoff scores



Adaptive beliefs. We now analyze whether students anticipate current changes in the distribution of admission cutoffs (*Ratex* beliefs) or believe that cutoff distributions for the current admission process are the same as the distributions of the previous admission process (*Adaptive* beliefs).

Figure 5.5 shows the distributions of the difference between subjective expected cutoffs– for every listed program–and expected cutoffs given by *Ratex* (red) and *Adaptive* beliefs (blue). We measure the difference in expected cutoffs in standard deviations of application scores. Panel 5.5a shows these distributions for all programs. We observe that both distributions are centered around zero, i.e., there is no evidence of aggregate *optimism* or *pessimism*. However, the distribution with *Adaptive* beliefs is more concentrated towards zero. To analyze whether students' beliefs are closer to *Ratex* or *Adaptive* beliefs, Panel 5.5b shows the distribution of bias for programs that increased their vacancies in at least 25% compared to the previous year, and Panel 5.5c for programs that decreased their vacancies in at least 25%. In both cases we observe that the distributions of bias with *Adaptive* beliefs are more centered around zero, even though the distribution of bias with *Ratex* beliefs are more displaced to the sides. This suggests that students do not correctly anticipate changes in cutoffs, even for programs that change significantly their vacancies from year to year, and that their beliefs are closer to *Adaptive* than *Ratex* beliefs.

Figure 5.5: Distributions of the standardized difference between subjective expected cutoffs and *Ratex* and *Adaptive* beliefs



Modeling bias. We consider a simple model of subjective beliefs to capture the previous data patterns, i.e., (i) that students have biased beliefs that are centered around *Adaptive* beliefs, (ii) that beliefs are subject to the pull-to-the-center effect, and (iii) that students are more biased if they do not know previous years' cutoff scores. Formally, we introduce the following definitions and assumptions:

Definition 7 (Consideration sets)**.** For each student $i \in N$, define by $M_{it} \subseteq M$ the set of programs such that the student knows the cutoff for year $t-1$, $\bar{s}_{jt-1}$.

Assumption 1 (Subjective beliefs as a deviation)**.** We denote by $\tilde{p}_{ijt}$ the subjective belief of student i in program j in period t, and we compute it as

$$\tilde{p}_{ijt} \equiv \mathbb{P}\left(s_{ij} \geq \tilde{s}_{ijt}\right) \tag{5.1}$$

where s_{ij} is the application score of student i in program j, and $\tilde{s}_{ijt}$ is a random variable given by

$$\tilde{s}_{ijt} = \begin{cases} \bar{s}_{jt-1} + \nu_{ijt} & \text{if } j \in M_{it} \\ \eta_{ijt} & \text{otherwise} \end{cases} \tag{5.2}$$

where $\bar{s}_{jt-1}$ is the realized cutoff score for program j in year $t-1$, $\nu_{ijt} \sim g_{ijt}\left(\nu_{ijt}\right)$ is an idiosyncratic shock that induces bias over the cutoff distribution for program j in year

t, and $\eta_{ijt} \sim h_{ijt}(\eta_{ijt})$ is the prior beliefs of student i when she is uninformed about the expected cutoff score for program j in year $t-1$.

To capture the pull-to-the-center effect, we allow the bias to depend on the distance between students' application scores and the cutoff score of the previous year. For instance, if students above the previous year cutoff scores tend to be pessimistic, and students below tend to be optimistic, we would observe gravitation to the middle.

To test whether students' bias are correlated with their preferences and whether the pull-to-the-center effect is driven by differences in the mean of the bias shock, we decompose students' bias on admission probabilities relative to *Adaptive* beliefs and estimate the following regression:

$$\underbrace{\frac{\mathbb{E}\left[\tilde{s}_{ijt}\right]-\bar{s}_{jt-1}}{\bar{s}_{jt-1}/100}}_{\text{Bias in expected cutoff}} = \alpha_i+\beta_1 \underbrace{\left[\frac{s_{ij}-\bar{s}_{jt-1}}{\bar{s}_{jt-1}/100}\right]^+}_{\text{Distance (if positive)}} +\beta_2 \underbrace{\left[\frac{s_{ij}-\bar{s}_{jt-1}}{\bar{s}_{jt-1}/100}\right]^-}_{\text{Distance (if negative)}} +\gamma_{rank} rank\left(R_i(j)\right)+\epsilon_{ijt}, \quad (5.3)$$

where $\mathbb{E}\left[\tilde{s}_{ijt}\right]$ is the subjective expectation of the cutoff score $\bar{s}_{jt}$ by student i, α_i is a student fixed-effect, s_{ijt} is the application score of student i in program j, $rank_{R_i(j)}$ is a vector with a one in the position of program j in student's i ROL, and zero otherwise, and ϵ_{ijt} is an i.i.d error term. The function $[\cdot]^+$ ($[\cdot]^-$) returns the absolute value of the argument if positive (negative), and returns zero otherwise.

Column (1) in Table 5.1 reports the estimation results. We observe that students whose application scores are above the previous year's cutoffs have, on average, an additional upward bias of near 0.5 percentage points per unit of distance (per one percentage point above the cutoff). This statistic suggests that students above the cutoffs tend to be more pessimistic as the distance from the cutoff increases. Similarly, students who are below the previous year's cutoffs have on average an additional downward bias of near 0.4 percentage points per unit of distance, i.e., students below the previous year's cutoffs tend to be more optimistic. These effects are consistent with the pull-to-the-center effect. In addition, we observe a negative correlation between the preference rank and the proportional bias in expected cutoffs relative to the top-reported preference. For instance, programs listed in the fifth reported preferences exhibit 1.1 additional percentage points of downward bias than programs listed in the top-reported preference. This result suggests that students tend to be slightly more optimistic for programs listed at the bottom of their application lists.

To understand how the magnitude of the bias differs with students' observable characteristics, in column (2) of Table 5.1 we report the results of considering the logarithm of the norm 2 of the bias as a dependent variable and replacing the fixed effects with students' observable characteristics, including their gender, normalized application score, the type of high school they graduated from (relative to Private schools), whether the program is their most desired preference, whether they know someone at the program, among others. First, we find that females are significantly more biased than males. Second, we observe that students from public and voucher schools are significantly more biased than students from private schools. Third, we observe that the application score has a negative and significant effect. These results are consistent with previous literature and suggests that students with high SES might have more accurate beliefs than students with low SES (potentially due to differential access to information). Fourth, we observe that students' beliefs about their admission chances in their most desired program are significantly more accurate than in other programs. This result is intuitive, as students may collect more information regarding their most desired preference. Finally, we observe that knowing someone at the program also helps students to have more accurate beliefs.

Table 5.1: Regression Results on Bias

	(1)	(2)
Distance score to cutoff (positive)	0.546***	0.055***
	(0.005)	(0.0005)
Distance score to cutoff (negative)	-0.412***	0.062***
	(0.010)	(0.001)
Score	-	-0.377^{***}
	-	(0.005)
Female	-	0.058***
	-	(0.010)
Public	-	0.111***
	-	(0.015)
Voucher	-	0.123***
	-	(0.013)
Most Preferred	-	-0.106^{***}
	-	(0.019)
Knows Someone	-	-0.099^{***}
	-	(0.012)
Preference 2	-0.418***	0.114***
	(0.050)	(0.017)
Preference 3	-0.692***	0.237***
	(0.061)	(0.018)
Preference 4	-1.042***	0.326***
	(0.071)	(0.019)
Preference 5	-1.169***	0.403***
	(0.089)	(0.021)
Preference 6	-1.167***	0.467***
	(0.105)	(0.023)
Preference 7	-1.068***	0.522***
	(0.129)	(0.026)
Preference 8	-0.840***	0.540***
	(0.154)	(0.030)
Preference 9	-1.017***	0.605***
	(0.182)	(0.034)
Preference 10	-1.588***	0.626***
	(0.227)	(0.039)
Constant	-	0.113***
	-	(0.019)
Observations	78,095	77,409

Note: In column (1), the dependent variable is the bias, and we include student fixed-effects. In column (2), the dependent variable is the log of the norm 2 of the bias, and we do not consider student fixed-effects.

Discussion: For every student i, we elicited a measure for $\mathbb{E}\left[\tilde{s}_{ijt}\right]$ and $\tilde{p}_{ijt}$ at her application score s_{ij} but only for programs that were listed in the application or declared as top-true preference. We then face a selection problem: the sample of observed beliefs might not come from a random sample of programs from within the consideration set of the student. The potential bias could come from at least two sources: (i) correlation between preferences and bias and (ii) correlation between bias and the decision to rank a program in the list. In the first case, students' preferences could be correlated with their bias if they follow a search process to form their subjective beliefs and tend to search more information for programs they like the most. In the second case, the ranking strategy could be correlated with bias if, for instance, students maximize their expected utility over their assignment and face–even small–application costs. Under this scenario, if a student has a positive bias in her subjective beliefs for a given program that she likes, she may be more likely to include that program in her list than a similar program where she has a negative bias in her beliefs. To address this selection issue, we redo our previous analysis considering only the bias for students' top-true preference. We obtain similar results concerning the pull-to-the-center effect.

5.2 Prevalence and Relevance

Under-confidence and Ordering. As previously defined, we say that students make an under-confidence mistake if, despite them having valid applications, they do not apply to their top-true choice as their top-reported preference, even though their score is high enough to be admitted with positive probability and the constraint in the length of the list is not binding. In addition we say that students make an ordering mistake if by changing the order of a program in their list they can improve their expected utility.

Table 5.2 shows the percentage of students who make under-confidence and ordering mistakes. We compute this statistic by the level of knowledge the student declares about last year's cutoff scores. Overall, we observe that between 1.7% and 3.2% of students make an under-confidence mistake *ex-post* and that between 0.9% and 2.0% of students make an ordering mistake *ex-post*, i.e., they do not include their top-true preference as their top-reported preference, and their application score was above the realized cutoff score for that program. This percentages increase to 2.3%-4.2% if we consider *ex-ante* under-confidence mistakes and to 1.8%-3.5% if we consider *ex-ante* ordering mistakes, i.e., all students who face a strictly positive probability of admission to their top-true preference but did not include that program as their top-reported preference. We also observe that students who do not know any of the cutoff scores for their listed programs experience a higher prevalence of under-confidence and ordering mistakes, suggesting that a driver of these mistakes could be the lack of information about past cutoff scores.

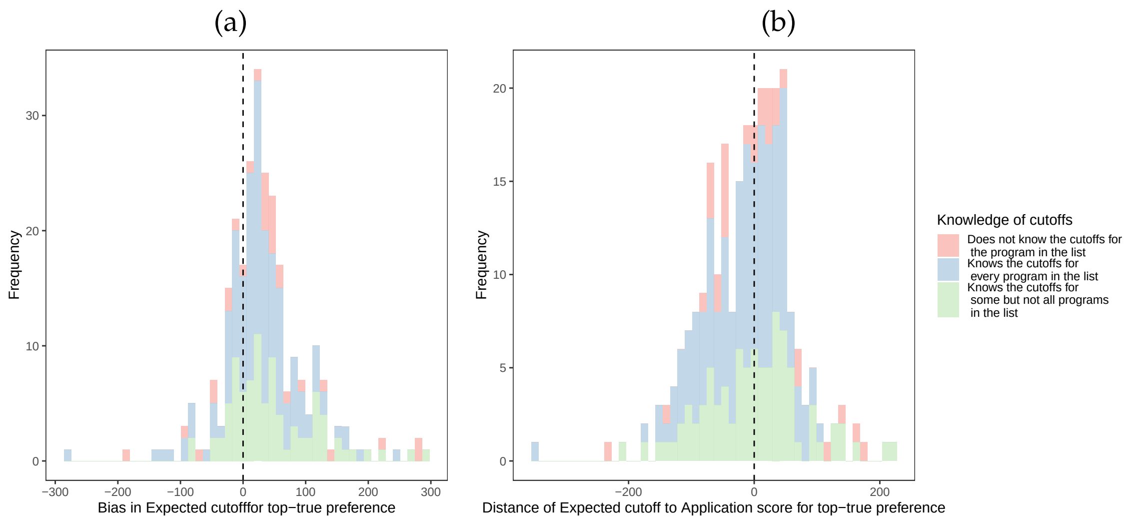
Table 5.2: Under-confidence and Ordering mistakes

	Under-confident mistake		Ordering mistake	
Knowledge of cutoffs	Ex-post [%]	Ex-ante [%]	Ex-post [%]	Ex-ante [%]
Does not know the cutoffs	3.19	4.16	1.94	3.47
for the program in the list	(0.65)	(0.74)	(0.51)	(0.68)
Knows the cutoffs for	1.73	2.34	0.94	1.82
every program in the list	(0.17)	(0.19)	(0.12)	(0.17)
Knows the cutoffs for some	2.00	2.74	1.35	2.55
but not all programs in the list	(0.25)	(0.29)	(0.2)	(0.28)

Note: standard errors are computed in parenthesis.

We now analyze whether students' subjective beliefs can fully explain under-confidence mistakes. Figure 5.6 shows frequency histograms for the bias in expected cutoffs for the top-true preference relative to *Ratex* beliefs (panel (a)), and the distance between the application score of the student and their subjective expected cutoff (panel (b)). We compute these distributions only for students who made an *ex-ante* under-confidence mistake and give consistent answers to the survey (see Section 5.1). From panel (a), we see that close to 75% of the observations fall at the right of zero, i.e., students who make an under-confidence mistake tend to be pessimistic about the expected cutoff for their top-true preference. From panel (b) we observe that only close to 50% of students declare to believe that the realization of the cutoff—for the current admission process—will be higher than their application score. This last result implies that only 50% of under-confidence mistakes could be explained by bias in subjective beliefs about admission cutoffs (without considering potential measurement errors).[15]

Figure 5.6: Distributions of bias in expected cutoff by knowledge of cutoffs for under-confidence mistakes



OVER-CONFIDENCE. Students make an over-confidence mistake if, despite having valid applications, they do not apply to programs they: (i) prefer to being unassigned, and (ii) face a positive probability of assignment, even though they face a positive probability of being unassigned to the centralized system and the constraint in the length of the list is not binding.

To estimate the prevalence and relevance of over-confidence mistakes, we included in our 2022-survey the following question:

Imagine a HYPOTHETICAL scenario in which you were NOT admitted to any program in your application list. Is there any program in the centralized system that you have NOT included in your application but you would prefer than being unassigned?

[15]In Appendix A.2, Table A.3, we detail the reasons students–with *ex-ante* under-confidence mistakes–give for not listing their top-true preference.

This question allows us to measure bounds on over-confidence mistakes. We estimate that around 5% of survey respondents declare *yes* to the previous question and had strictly positive risk of being unassigned to the centralized system. We interpret this measure as an upper bound on over-confidence mistakes.

On the other hand, by analyzing the program students declare after answering *yes* in the previous question, we estimate that at least 0.9% and 1.2% of students make an *ex-ante* and *ex-post* over-confidence application mistake respectively.

To understand the drivers of these mistakes, we analyze students' beliefs. Notice that, for any student i, a necessary condition for making an over-confidence mistake given her application list R_i, is that $\exists j \in M \notin R_i$, such that i prefers to be assigned to j than to be unassigned, i.e., $j \succ_i \emptyset$. If such a program exists, over-confidence mistakes could result from students having biased beliefs and facing small application costs. For instance, suppose students face small application costs. In this scenario, a student might not include a program of her preference if her subjective beliefs: (i) assign a low admission probability to that program or (ii) assign a low risk of being unassigned to the centralized system given her application list. Therefore, if these beliefs are biased, students' might be making over-confidence mistakes.

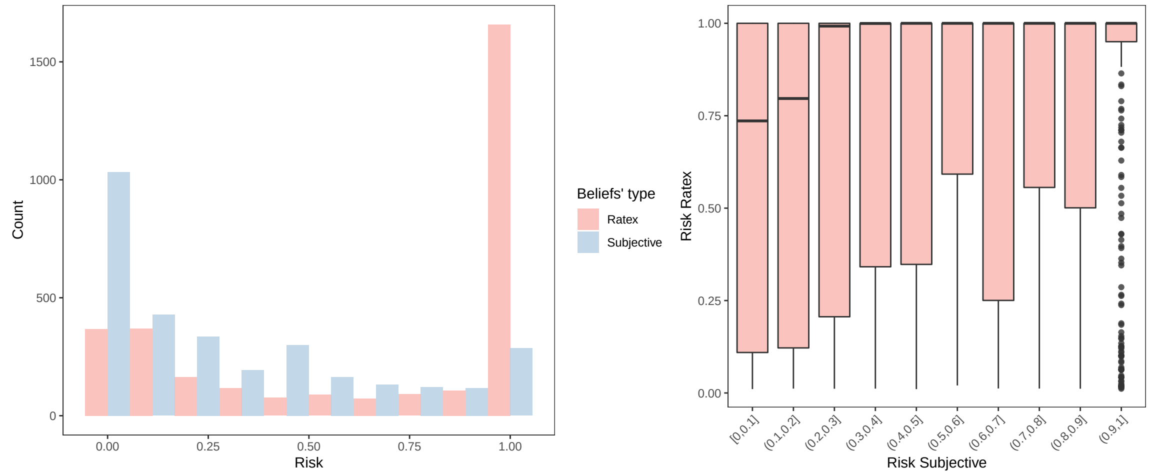
Figure 5.7 shows the distributions of *Ratex* and subjective application *Risk*. Panel 5.7a shows the histograms for the risk of application lists given *Ratex* and subjective beliefs, and Panel 5.7b shows boxplots of *Risk* given *Ratex* beliefs by level of subjective beliefs. We compute these statistics for all students who face a non-zero risk given *Ratex* beliefs and whose average scores are above the minimum required to be admitted to a program in the centralized system. These students are likely to face a strictly positive probability of admission to some program in the centralized system.

We observe that students who face some risk tend to under-predict how risky are their application lists. In fact, close to 10% of the sample faces a $Risk$ given *Ratex beliefs* greater than 1%, and close to 80% of this group under-predicts their $Risk$. This pattern is particularly severe at the extremes, where only a small fraction of students believe to be facing a risk equal to 1 relative to what *Ratex* beliefs predict. Indeed, close to 20% of students with positive $Risk$ and average scores above the minimum requirements believe they have a $Risk$ lower than 10% when they face a $Risk$ given $Ratex$ greater than 70%.

These results suggest that biased beliefs result in payoff-relevant over-confidence mistakes. In Section 6, we analyze if it is possible to reduce these mistakes by giving information to students about the correct risk of their application lists.

Figure 5.7: Distribution of Subjective and Ratex application Risk

(a) Histograms of Risk by beliefs' type (b) Ratex Risk by Subjective Risk levels



6 Field Experiment

Based on our previous results, information frictions could affect students' applications and outcomes. This section reports the design and results of an intervention aiming to reduce information frictions and application mistakes.[16]

6.1 Description

In collaboration with MINEDUC, we designed and implemented an intervention to provide information to students during the application process. Specifically, using partial information about the applications, we created a personalized website for each student, including information about the programs in the student's preference list and recommendations to improve the application.

6.1.1 Background.

As discussed previously, the application process starts when the scores of the PDT are published. Students have five days to submit an application list—in the admission process of 2022, from January 11 to January 15—and they are allowed to modify and update it as many times as they want within this time window.

To personalize the information provided, we use the applications received up to January 12 at 8 pm CT, which included 107,837 students, representing 72.48% of the total number

[16] Results for the 2021 intervention are available upon request.

of students who applied. For each student in this group, we used their submitted application to create a personalized website, and we sent them an email on January 13 at 9 am CT. The email included each student's personalized link and a general message inviting them to open it and get more information to improve their application.

6.1.2 INFORMATION.

We carefully designed the information included in the personalized websites to address the causes of mistakes outlined in the previous sections, namely, lack of information and biased beliefs. Specifically, the intervention had four main modules:

- M1 General information about programs included in the applicant's list.
- M2 Personalized information about scores for the programs included in the applicant's list.
- M3 Personalized alerts depending on the admission probabilities.
- M4 Personalized recommendations about other majors of potential interest.

GENERAL INFORMATION ABOUT PROGRAMS. In Figure 6.1 we show an example of M1. Figure 6.1a shows how the list of programs included in the student's list is displayed. Students can click on each program to see detailed information, as shown in Figure 6.1b. Specifically, this information includes:

- Location: campus and university to which the program belongs.
- Accreditation: number of years that the institution is accredited.[17]
- Benefits: benefits and types of student aid for which the student is eligible in that program.
- Duration: formal duration of the program, measured in semesters.
- Tuition: yearly tuition measured in Chilean pesos.

We provide this information to all students who received the intervention, i.e., applied before January 12 at 8 pm CT.

[17] The years of accreditation is a signal of the quality of the institution. If the institution is not accredited, enrolled students cannot receive public student aid. See details in `https://www.cnachile.cl/`.

Figure 6.1: Information on Programs Included in Application

(a) General

(b) Detailed

Personalized information about scores. In Figure 6.2 we show an example of M2. As in M1, students first see a list of the programs they included in their application (as shown in Figure 6.2a), and they can click on each program to see their personalized information, which includes:

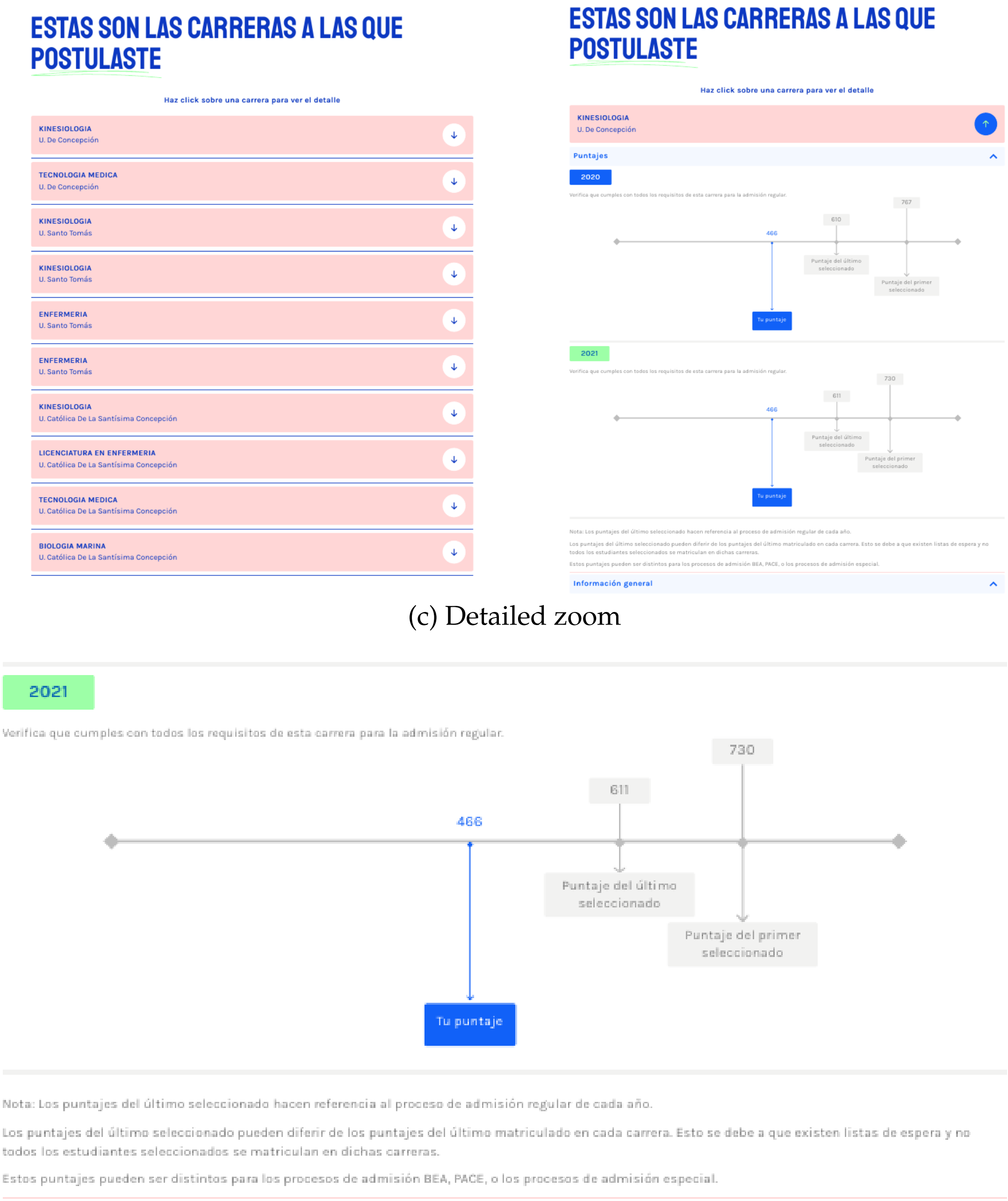
- Scores: application score of the first and last student admitted in the admission processes of 2020 and 2021. We also include a graphical representation of where the student stands relative to these scores.[18]
- Validity: if the student does not fulfill the requirements of the program (i.e., makes an admissibility mistake), we display an alert that includes the following message:[19]

  Please verify that you satisfy the admission requirements for this program.

[18] To provide more relevant information, the score of the last admitted student displayed depends on the admission tracks where the student is participating. Hence, if the student is BEA, we display the score of the last admitted student in the BEA process.

[19] As requested by MINEDUC, we display the application score if it can be computed, despite the admissibility mistake. However, if one of the scores is missing (and thus we cannot compute the application score), then we display the message *Score not computed*.

Figure 6.2: Feedback on Programs' Admission Chances



(a) General

(b) Detailed

(c) Detailed zoom

Personalized alerts depending on admission probabilities. Considering the programs listed and the application scores of all students who applied before January 12 at 8 pm CT, this module computes the probability of each student's admission in each of their preferences. In Appendix B.1 we described in detail how we calculate these probabilities. In short, we estimate two sets of probabilities:[20]

[20]We consider these two sets of probabilities to reduce the risk of displaying misleading information to students.

- Interim: we compute these probabilities using the applications received before January 12 at 8 pm CT. To do so, we first estimate the total number of applicants that will apply during the application process and then use this information to estimate the admission probabilities via Bootstrap.
- Adaptive: we compute these probabilities via Bootstrap using students' applications in the admission process of 2021.

Based on these probabilities, this module displays personalized alerts to help students prevent strategic mistakes. As shown in Figure 6.3, these alerts are embedded in M2, adding the following information:

- Alert by program: if the estimated admission probabilities (both interim and adaptive) are below 1%,[21] we display a red alert that includes the following message:

  > *Based on the applications received up to January 12th at 11 pm, we find that your admission probability in this program is low. Nevertheless, you can still apply, as the cutoff of this program may change from year to year and also there are waitlists.*

  We illustrate this alert in Figure 6.4.

- Overall alert: depending on the admission probabilities of the top preference and overall,[22] we display an alert nudging students to consider additional programs in their application list. Figure 6.5 shows the different message types. There are three groups:

  1. If the estimated probability of being assigned to the top preference is above 99%, we recommend students to add *reach* programs to their lists, i.e., programs that are generally more preferred, that the student may be interested in[23], and for which the student faces positive admission probability (Figure 6.5b).
  2. If the overall probability of being assigned is below 99%, we recommend students to add *safety* programs, i.e., programs for which the student faces a positive admission probability (Figure 6.5a);
  3. If none of the cases above holds, we display a message inviting students to *explore* and get information about other programs.

Notice that we recommend students to add *safety* programs to reduce over-confidence mistakes, while we recommend adding *reach* programs to reduce under-confidence mistakes. As requested by MINEDUC, none of our interventions encourages students to

[21] Admission probabilities are bimodal and highly concentrated in the two extremes (i.e., probability equals to 0 or 1). Hence, any threshold between 1% and 99% leads to similar results. Nevertheless, MINEDUC opted to use 1% to be more conservative.

[22] Notice that this probability considers all programs included in the list. More specifically, if the student applied to a subset R of programs and p_r represents the probability of being assigned to program $r \in R$, then this probability can be computed as $1 - \prod_{r \in R}(1 - p_r)$.

[23] To determine potential *reach* programs, we use the information on students' top-true preferences in the survey of 2021. We compute transition matrices for programs that are typically declared to be top-true preferences conditional on the top-reported preference submitted by the student.

remove programs from their lists (even in the presence of admissibility mistakes) or alter the order of the programs initially included.

Figure 6.3: Feedback on Programs' Admission Chances

(a) Red Alert (b) No Alert

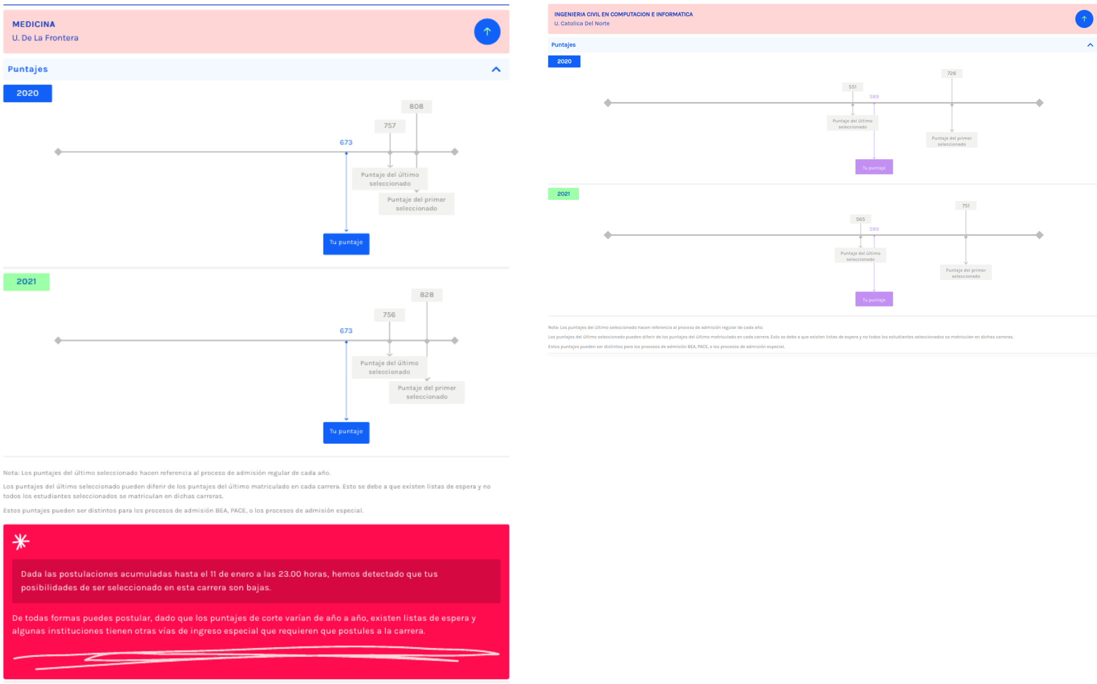


Figure 6.4: Feedback on Programs' Admission Chances: red warning

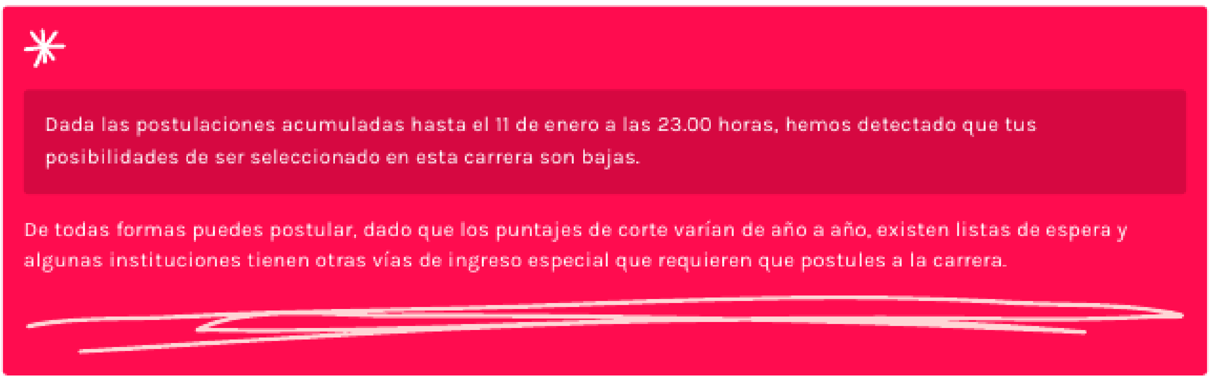
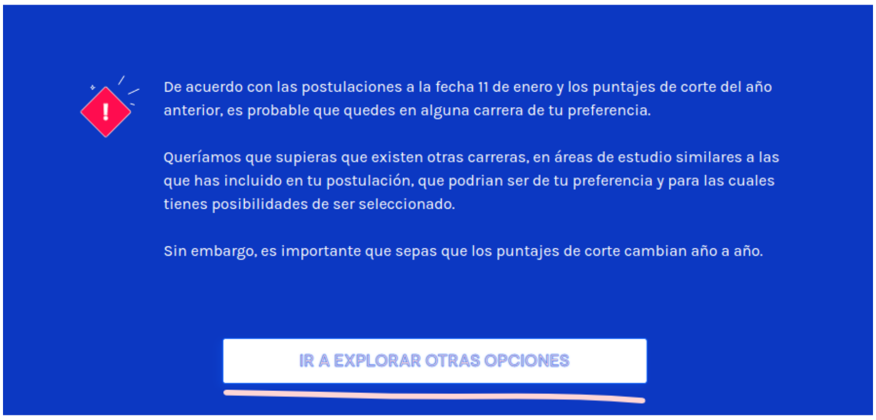


Figure 6.5: Feedback on Application list's and potential strategic mistakes

(a) Add Safety

(b) Add Reach



PERSONALIZED RECOMMENDATIONS. Based on students' scores and their reported preferences, we compute four personalized major recommendations to encourage students to consider other options. In Appendix B.2 we described in detail how these recommendatios are obtained. In summary, the four recommmendations are:

1. The most preferred major according to the student's list of preferences,
2. The second most preferred major according to the student's list of preferences,
3. The major with the highest expected wage among all majors belonging to IPs or CFTs,
4. The major with the highest expected wage among all feasible majors (i.e., majors for

which the student has a positive probability of assignment) considering the transition matrix previously computed.

For each recommended major we include the following information:

- Duration: average duration of the programs belonging to that major.
- Employment: average employment rate among the programs belonging to that major.
- Wage: average wage (four years after graduation) among the programs belonging to that major.
- Scores: minimum and maximum application score of the last admitted student to any of the programs belonging to that major.

In Figure 6.6 we provide an example of this module.

Figure 6.6: Recommendation of Other Majors

(a) General

(b) Detailed

Note that the recommendatios are made at the major level, while students apply to specific programs.[24] However, by providing a range of scores for the last admitted students,

[24]MINEDUC did not allow us to make program-specific recommendatios to avoid favoring some schools/universities.

we aim to extend students' consideration sets and encourage them to find more information about these majors. Hence, we believe that the recommendation module serves two purposes: (i) reduce potential information frictions about programs' characteristics and (ii) affect students' beliefs on admission probabilities for programs that are not in their consideration sets.

6.2 Experimental Design

To properly evaluate the impact of each module, we assigned each student (who applied before January 12 at 8pm CT) to one of four treatments groups:

- T1 General information: these students received only M1.
- T2 General information + scores: these students received module M1 and M2.
- T3 General information + scores + alerts: these students received module M1, M2 and M3.
- T4 General information + recommendations: these students received module M1 and M4.

We perform the assignment of students to treatments in a stratified way to achieve balance on observables across groups. In Appendix B.3 we describe the variables used for stratificationn and we report the results of several balance checks.

As previously discussed, each student that applied before January 12 at 8pm CT received an email with a link to their personalized website. In addition, using the same stratification discussed in Appendix B.3, we randomly chose 30,000 students and sent them an SMS encouraging them to open their personalized website. As we discuss in Section 6.3, we use this as an instrument for the decision of opening the email.

6.3 Results

In this section, we evaluate the results of the intervention. Table 6.1 shows aggregate statistics by group. Among the four treatment groups of interest (T1, T2, T3, and T4), we observe that close to 26,000 students received the email, and around 28% of them opened their personalized website. As expected, we do not observe significant differences across groups in opening the email. We use T1 as a control group because all treatment arms received M1. We observe that close to 77% of the students in each group are assigned to the system. In addition, students in T2 and T3 increase the length of their application lists more than students in T4 and T1. This translates into more students entering the centralized system or changing their program of assignment after the intervention.

Table 6.1: Summary Statistics by Group

			Application			Assignment				
Treatment	Total	Opened [%]	Modified [%]	Increased [%]	Decreased [%]	Assigned [%]	Entered [%]	Left [%]	Changed status [%]	Changed program [%]
T1	25758	28.325	10.839	4.158	1.452	77.739	2.677	0.827	1.254	3.762
		(0.281)	(0.194)	(0.124)	(0.075)	(0.259)	(0.209)	(0.064)	(0.069)	(0.119)
T2	25932	28.679	11.33	4.454	1.542	77.553	3.274	0.85	1.415	3.59
		(0.281)	(0.197)	(0.128)	(0.077)	(0.259)	(0.229)	(0.065)	(0.073)	(0.116)
T3	25889	28.36	11.7	4.558	1.622	77.736	3.47	0.794	1.414	3.948
		(0.28)	(0.2)	(0.13)	(0.079)	(0.259)	(0.236)	(0.063)	(0.073)	(0.121)
T4	25826	28.692	11.314	4.306	1.642	77.643	3.176	0.828	1.375	3.473
		(0.281)	(0.197)	(0.126)	(0.079)	(0.259)	(0.226)	(0.064)	(0.072)	(0.114)

Note: Opened is a binary variable equal to 1 if the student opened the personalized website, 0 otherwise. Modified is a binary variable equal to 1 if the student modified its application after the personalized websites were sent, 0 otherwise. Increased (decreased) is a binary variable equal to 1 if the student increased (decreased) the number of valid applications in their list, 0 otherwise. Assigned is a binary variable equal to 1 if the student resulted assigned at the end of the process, 0 otherwise. Entered (left) is a binary variable equal to 1 if the student resulted unassigned (assigned) given their list of preferences before the intervention and (un)assigned given their preferenes after the intervention, 0 otherwise. Changed status (program) is a binary variable equal to 1 if the student changed their status (program) of assignment considering the list of preferences submitted before and after the intervention. Standard errors reported in parenthesis.

6.3.1 Effect of Modules on Outcomes.

General Results. In Table 6.2 we replicate Table 6.1 separating students who did and did not open their personalized website. We focus on students who opened the email and treat T1 as a control group. In this way we can assess which interventions have the largest impact on student outcomes. Table 6.3 formalizes the analysis showing logistic regressions and their odd-ratios of the outcomes on the different treatment arms. We observe that students in T2—who received information about their scores, cutoffs, and admissibility mistakes—and students in T3—who in addition received personalized alerts—are significantly more likely to modify their application list after receiving the intervention (around 14%-15% higher odds relative to the control group). Moreover, students in T3 are significantly more likely to increase the length of their application lists (16% higher odds relative to the control group), being assigned to the centralized system after modifying their list (50% higher odds relative to the control group), and changing their assignment status (46% higher odds relative to the control group) after they received the intervention. These results suggests that providing information about scores and previous years' cutoffs, together with warning messages about students' risk, is significantly more effective at reducing information frictions and application mistakes, compared to providing only administrative information about scores and previous year's cutoff scores.

Table 6.2: Summary Statistics by Group and Reception

			Application			Assignment			
Treatment	Opened	N	Modified [%]	Increased [%]	Decreased [%]	Entered [%]	Left [%]	Changed status [%]	Changed program [%]
T1	No	18462	9.999	3.884	1.316	2.498	0.887	1.267	3.618
			(0.221)	(0.142)	(0.084)	(0.236)	(0.079)	(0.082)	(0.137)
T1	Yes	7296	12.966	4.852	1.796	3.173	0.682	1.22	4.126
			(0.393)	(0.252)	(0.155)	(0.442)	(0.109)	(0.129)	(0.233)
T2	No	18495	10.024	4.098	1.319	2.909	0.851	1.341	3.412
			(0.221)	(0.146)	(0.084)	(0.253)	(0.077)	(0.085)	(0.133)
T2	Yes	7437	14.576	5.338	2.098	4.25	0.846	1.6	4.034
			(0.409)	(0.261)	(0.166)	(0.497)	(0.12)	(0.146)	(0.228)
T3	No	18547	10.476	4.146	1.483	3.01	0.74	1.272	3.661
			(0.225)	(0.146)	(0.089)	(0.259)	(0.072)	(0.082)	(0.138)
T3	Yes	7342	14.792	5.598	1.975	4.687	0.93	1.771	4.672
			(0.414)	(0.268)	(0.162)	(0.522)	(0.127)	(0.154)	(0.246)
T4	No	18416	10.339	4.045	1.558	3.129	0.819	1.368	3.312
			(0.224)	(0.145)	(0.091)	(0.263)	(0.076)	(0.086)	(0.132)
T4	Yes	7410	13.738	4.953	1.849	3.301	0.849	1.39	3.873
			(0.4)	(0.252)	(0.157)	(0.442)	(0.121)	(0.136)	(0.224)

Note: Opened is a binary variable equal to 1 if the student opened the personalized website, 0 otherwise. Modified is a binary variable equal to 1 if the student modified its application after the personalized websites were sent, 0 otherwise. Increased (decreased) is a binary variable equal to 1 if the student increased (decreased) the number of valid applications in their list, 0 otherwise. Entered (left) is a binary variable equal to 1 if the student resulted unassigned (assigned) given their list of preferences before the intervention and (un)assigned given their preferenes after the intervention, 0 otherwise. Changed status (program) is a binary variable equal to 1 if the student changed their status (program) of assignment considering the list of preferences submitted before and after the intervention. Standard errors reported in parenthesis.

Table 6.3: Regression Results among Openers

	Application			Assignment			
Treatment	Modified	Increased	Decreased	Entered	Left	Changed status	Changed program
T2	0.136^{***}	0.101	0.159	0.304	0.218	0.275^{*}	-0.023
	(0.048)	(0.075)	(0.120)	(0.189)	(0.215)	(0.141)	(0.083)
T3	0.153^{***}	0.151^{**}	0.097	0.406^{**}	0.313	0.378^{***}	0.130
	(0.048)	(0.074)	(0.122)	(0.185)	(0.212)	(0.139)	(0.081)
T4	0.067	0.022	0.030	0.041	0.221	0.132	-0.066
	(0.049)	(0.076)	(0.123)	(0.200)	(0.215)	(0.146)	(0.084)
Constant	-1.904^{***}	-2.976^{***}	-4.002^{***}	-3.418^{***}	-4.981^{***}	-4.394^{***}	-3.146^{***}
	(0.035)	(0.054)	(0.088)	(0.144)	(0.161)	(0.107)	(0.059)
				Odd-Ratios			
T2	1.145	1.106	1.172	1.355	1.243	1.317	0.977
T3	1.165	1.163	1.102	1.501	1.367	1.46	1.139
T4	1.069	1.022	1.03	1.042	1.247	1.141	0.977
Observations	29,485	29,485	29,485	6,502	22,983	29,485	29,485

Note: Logistic regression results. Modified is a binary variable equal to 1 if the student modified its application after the personalized websites were sent, 0 otherwise. Increased (decreased) is a binary variable equal to 1 if the student increased (decreased) the number of valid applications in their list, 0 otherwise. Entered (left) is a binary variable equal to 1 if the student resulted unassigned (assigned) given their list of preferences before the intervention and (un)assigned given their preferences after the intervention, 0 otherwise. Changed status (program) is a binary variable equal to 1 if the student changed their status (program) of assignment considering the list of preferences submitted before and after the intervention. Significance reported: *p<0.1; **p<0.05; ***p<0.01

EFFECT OF SPECIFIC WARNINGS. To understand the effect of the different warning messages, in Table 6.4 we show the previous statistics conditional on opening the personalized website and grouping by message type. Group 1 was assigned to potentially receive *reach* messages; group 2, *safety* messages; and group 3, *explore* messages. [25] We now analyze the results for each type of warning.

Reach. Students who faced an admission probability to their top-reported preference above 99% were eligible to receive a message recommending them to include additional *reach* programs. The purpose of recommending *reach* programs is to decrease their likelihood of making an underconfidence mistake. By design, this group of students faces a low risk of not being assigned to the centralized system, as the column Assigned [%] in Table 6.4 shows. We do not find any statistically significant effect of the treatmeants on the outcomes of interest. This suggests that *reach* meassages are not highly effective at reducing *under-confidence* mistakes.

Safety. Students who faced an application risk greater than 1% were eligible to receive a message recommending them to include additional *safety* programs. *Safety* programs are less selective programs than the ones listed in the student's application—where the student faces positive admission probability—but that the student might prefer to her outside option. The purpose of recommending *safety* programs is to decrease their likelihood of making an overconfidence mistake. As Table 6.4 shows, students in the *safety* group have a significantly lower probability of being assigned, ranging from 11% to 13%. In Table 6.5 we focus on students in the *safety* group who opened their emails and measure the treatment effects on the same outcomes. We observe that students in T2 and T3 have 53% and 62% higher odds to enter the centralized system than students in T1. In addition, we do not observe that students in this group are more likely to change the length of their application lists or program of assignment.

Explore. Students who faced an application risk below 1% and a probability of not being assigned to their top-reported preference below 99% were eligible to receive a message recommending them to explore additional programs. In Table 6.6 we focus on students in the *explore* group who opened their emails and measure the treatment effects on the same outcomes. We observe that students in T3 have close to 16% higher odds to modify their lists relative to the control group. In addition, students in T3 have close to 17% higher odds to increase the length of their lists and 15% higher odds to change their assigned program. However, these effects are only statistically significant at a 10% confidence level.

[25] Students receive these type of messages only if they are assigned to treatment T3.

Table 6.4: Summary Statistics by Treatment, Reception and Message Group

				Application			Assignment				
Treatment	Open	Group	N	Modified [%]	Increased [%]	Decreased [%]	Assigned [%]	Entered [%]	Left [%]	Changed status [%]	Changed program [%]
		1	1544	8.031	3.044	1.101	99.87	0	0.195	0.194	2.979
				(0.692)	(0.437)	(0.266)	(0.092)	(0)	(0.112)	(0.112)	(0.433)
T1	No	2	4271	7.82	2.95	0.468	11.192	2.231	7.034	2.599	0.258
				(0.411)	(0.259)	(0.104)	(0.482)	(0.235)	(1.416)	(0.243)	(0.078)
		3	12647	10.975	4.301	1.629	96.513	5.036	0.809	0.949	4.831
				(0.278)	(0.18)	(0.113)	(0.163)	(1.072)	(0.081)	(0.086)	(0.191)
		1	604	10.43	4.636	1.987	99.834	0	0.166	0.166	2.98
				(1.245)	(0.856)	(0.568)	(0.166)	(0)	(0.166)	(0.166)	(0.692)
T1	Yes	2	1529	11.118	4.186	0.327	11.511	2.593	10.784	3.139	0.392
				(0.804)	(0.512)	(0.146)	(0.816)	(0.421)	(3.086)	(0.446)	(0.16)
		3	5163	13.81	5.075	2.208	97.211	8.844	0.538	0.775	5.365
				(0.48)	(0.305)	(0.205)	(0.229)	(2.35)	(0.103)	(0.122)	(0.314)
		1	1538	7.932	3.706	1.105	99.74	0	0.326	0.325	1.886
				(0.689)	(0.482)	(0.267)	(0.13)	(0)	(0.145)	(0.145)	(0.347)
T2	No	2	4290	8.065	3.17	0.186	11.538	2.619	5.643	2.844	0.21
				(0.416)	(0.268)	(0.066)	(0.488)	(0.253)	(1.294)	(0.254)	(0.07)
		3	12667	10.942	4.46	1.729	96.376	5.621	0.792	0.955	4.681
				(0.277)	(0.183)	(0.116)	(0.166)	(1.116)	(0.08)	(0.086)	(0.188)
		1	688	10.174	3.779	1.599	100	0	0	0	2.762
				(1.153)	(0.728)	(0.479)	(0)	(NA)	(0)	(0)	(0.625)
T2	Yes	2	1591	14.205	4.525	0.566	13.074	3.992	6.195	4.148	0.126
				(0.875)	(0.521)	(0.188)	(0.845)	(0.509)	(2.278)	(0.5)	(0.089)
		3	5158	15.277	5.797	2.637	96.51	6.548	0.842	1.028	5.409
				(0.501)	(0.325)	(0.223)	(0.256)	(1.914)	(0.129)	(0.14)	(0.315)
		1	1577	9.702	4.502	1.141	99.873	0	0.064	0.063	2.917
				(0.746)	(0.522)	(0.268)	(0.09)	(0)	(0.064)	(0.063)	(0.424)
T3	No	2	4273	7.559	2.996	0.398	11.233	2.68	5.346	2.879	0.164
				(0.404)	(0.261)	(0.096)	(0.483)	(0.257)	(1.263)	(0.256)	(0.062)
		3	12697	11.554	4.489	1.89	96.771	6.361	0.707	0.882	4.93
				(0.284)	(0.184)	(0.121)	(0.157)	(1.233)	(0.076)	(0.083)	(0.192)
		1	627	12.759	5.104	2.552	99.681	NaN	0.319	0.319	3.987
				(1.333)	(0.88)	(0.63)	(0.225)	(NA)	(0.225)	(0.225)	(0.782)
T3	Yes	2	1582	12.705	4.804	0.506	12.705	4.15	12.5	4.741	0.19
				(0.838)	(0.538)	(0.178)	(0.838)	(0.52)	(3.139)	(0.534)	(0.109)
		3	5133	15.683	5.903	2.357	96.571	9.249	0.746	1.033	6.137
				(0.508)	(0.329)	(0.212)	(0.254)	(2.209)	(0.122)	(0.141)	(0.335)
		1	1508	9.682	4.907	0.862	100	16.667	0	0.066	2.586
				(0.762)	(0.556)	(0.238)	(0)	(16.667)	(0)	(0.066)	(0.409)
T4	No	2	4260	7.793	3.31	0.305	11.667	2.634	5.788	2.864	0.446
				(0.411)	(0.274)	(0.085)	(0.492)	(0.255)	(1.326)	(0.256)	(0.102)
		3	12648	11.275	4.19	2.064	96.497	7.565	0.793	1.02	4.364
				(0.281)	(0.178)	(0.126)	(0.163)	(1.287)	(0.08)	(0.089)	(0.182)
		1	667	10.495	4.348	1.499	99.85	0	0	0	2.699
				(1.188)	(0.79)	(0.471)	(0.15)	(0)	(0)	(0)	(0.628)
T4	Yes	2	1569	11.09	3.314	0.637	11.217	2.802	10.377	3.314	0.319
				(0.793)	(0.452)	(0.201)	(0.797)	(0.432)	(2.976)	(0.452)	(0.142)
		3	5174	14.959	5.528	2.261	96.637	7.647	0.759	0.986	5.102
				(0.496)	(0.318)	(0.207)	(0.251)	(2.044)	(0.123)	(0.137)	(0.306)

Note: Opened is a binary variable equal to 1 if the student opened the personalized website, 0 otherwise. Modified is a binary variable equal to 1 if the student modified its application after the personalized websites were sent, 0 otherwise. Increased (decreased) is a binary variable equal to 1 if the student increased (decreased) the number of valid applications in their list, 0 otherwise. Assigned is a binary variable equal to 1 if the student resulted assigned at the end of the process, 0 otherwise. Entered (left) is a binary variable equal to 1 if the student resulted unassigned (assigned) given their list of preferences before the intervention and (un)assigned given their preferenes after the intervention, 0 otherwise. Changed status (program) is a binary variable equal to 1 if the student changed their status (program) of assignment considering the list of preferences submitted before and after the intervention. Standard errors reported in parenthesis.

Table 6.5: Regression Results among Openers in Safety group

	Application			Assignment			
Treatment	Modified	Increased	Decreased	Entered	Left	Changed status	Changed program
T2	0.280***	0.082	0.550	0.422*	−0.605	0.289	−1.141
	(0.109)	(0.176)	(0.559)	(0.218)	(0.504)	(0.193)	(0.817)
T3	0.151	0.144	0.438	0.485**	0.167	0.429**	−0.729
	(0.111)	(0.174)	(0.571)	(0.217)	(0.428)	(0.188)	(0.708)
T4	−0.003	−0.243	0.670	0.025	−0.043	0.056	−0.209
	(0.114)	(0.190)	(0.549)	(0.234)	(0.451)	(0.203)	(0.607)
Constant	−2.079***	−3.131***	−5.720***	−2.738***	−2.113***	−3.429***	−5.537***
	(0.081)	(0.128)	(0.448)	(0.170)	(0.319)	(0.147)	(0.409)
Odd-Ratios							
T2	1.324	1.085	1.734	1.525	0.546	1.335	0.319
T3	1.164	1.155	1.549	1.624	1.182	1.536	0.482
T4	0.997	0.785	1.955	1.026	0.958	1.058	0.319
Observations	6,271	6,271	6,271	2,545	433	6,271	6,271

Note: Logistic regression results. Modified is a binary variable equal to 1 if the student modified its application after the personalized websites were sent, 0 otherwise. Increased (decreased) is a binary variable equal to 1 if the student increased (decreased) the number of valid applications in their list, 0 otherwise. Entered (left) is a binary variable equal to 1 if the student resulted unassigned (assigned) given their list of preferences before the intervention and (un)assigned given their preferences after the intervention, 0 otherwise. Changed status (program) is a binary variable equal to 1 if the student changed their status (program) of assignment considering the list of preferences submitted before and after the intervention. Significance reported: *p<0.1; **p<0.05; ***p<0.01

Table 6.6: Regression Results among Openers in Explore group

	Application			Assignment			
Treatment	Modified	Increased	Decreased	Entered	Left	Changed status	Changed program
T2	0.118**	0.141	0.182	−0.325	0.378	0.285	0.009
	(0.056)	(0.087)	(0.129)	(0.426)	(0.251)	(0.210)	(0.087)
T3	0.149***	0.160*	0.067	0.049	0.330	0.290	0.143*
	(0.056)	(0.087)	(0.132)	(0.391)	(0.254)	(0.210)	(0.085)
T4	0.093*	0.090	0.024	−0.158	0.319	0.243	−0.053
	(0.056)	(0.088)	(0.133)	(0.409)	(0.254)	(0.212)	(0.088)
Constant	−1.831***	−2.929***	−3.791***	−2.333***	−5.204***	−4.853***	−2.870***
	(0.040)	(0.063)	(0.095)	(0.290)	(0.193)	(0.159)	(0.062)
Odd-Ratios							
T2	1.125	1.151	1.199	0.722	1.459	1.33	1.009
T3	1.161	1.173	1.069	1.05	1.391	1.336	1.153
T4	1.098	1.095	1.025	0.854	1.375	1.275	1.009
Observations	20,628	20,628	20,628	658	19,664	20,628	20,628

Note: Logistic regression results. Modified is a binary variable equal to 1 if the student modified its application after the personalized websites were sent, 0 otherwise. Increased (decreased) is a binary variable equal to 1 if the student increased (decreased) the number of valid applications in their list, 0 otherwise. Entered (left) is a binary variable equal to 1 if the student resulted unassigned (assigned) given their list of preferences before the intervention and (un)assigned given their preferences after the intervention, 0 otherwise. Changed status (program) is a binary variable equal to 1 if the student changed their status (program) of assignment considering the list of preferences submitted before and after the intervention. Significance reported: *p<0.1; **p<0.05; ***p<0.01

7 Conclusions

We analyze the prevalence and relevance of application mistakes in the Chilean centralized college admissions system. We exploit institutional features to identify a common type of application mistake: applying to programs without meeting all requirements (*admissibility* mistakes). We exploit the fact that *admissibility* mistakes are observed in the Chilean data. Moreover, there is a significant variation in admission requirements and *admissibility* mistakes over time.

We find that the growth of *admissibility* mistakes over time is driven primarily by growth on active score requirements. Also, we find that changes in admission requirements over time increase *admissibility* mistakes. However, this effect fades out over time, suggesting that students might adapt to the new set of requirements but not immediately. In addition, admissibility mistakes are likely welfare-relevant. Indeed, close to 25% of students who only list programs with *admissibility* mistakes could have been assigned in the centralized system if they had included programs in which they were eligible. As students are not fully aware of admission requirements and *admissibility* mistakes can be welfare-relevant, changes in requirements can affect students' outcomes. In this sense, increasing the complexity of the admission process can generate a negative externality in the system.

To analyze application mistakes not directly observed in the data, we design nationwide surveys and collect information about students' true preferences, their subjective beliefs about admission probabilities, and their level of knowledge about admission requirements and *admissibility* mistakes. Using this data, we shed light on which information frictions are the most relevant to explain students' mistakes.

We find that between 2% - 4% of students do not list their top-true preference of program, even though they face a strictly positive admission probability, and only a fraction of this skipping behavior can be rationalized by bias on students' subjective beliefs. Also, we find a pull-to-center effect on beliefs, i.e., students tend to attenuate the probability of extreme events. This effect translates into students under-predicting the risk of being unassigned to the system. Indeed, we estimate that at least 1% of students could have been better off by listing more programs in their application list. Finally, we also find that the magnitude of the bias considerably changes depending on students' characteristics. High-score students from private schools have significantly more accurate beliefs than students from public schools or low-score students.

Using the previous insights, we design and implement a large-scale outreach policy to reduce application mistakes. We find that showing personalized information about admission probabilities and information about the risk of application lists has a causal effect on improving students' outcomes, significantly reducing the risk of not being assigned to the centralized system and the incidence of over-confidence mistakes.

Our results suggest that information frictions play a significant role in affecting the performance of centralized college admissions systems, even when students do not face clear strategic incentives to misreport their preferences. In this sense, policy interventions that reduce these frictions can significantly reduce application mistakes and improve students' welfare. In future work, we aim at quantifying the welfare cost of the information frictions described in this paper by combining our administrative, survey and experimental data together with a structural model. Appendix C describes this model, our identification and estimation strategy, as well as the counterfactual policies we plan to evaluate.

REFERENCES

Agarwal, N. and Somaini, P. (2018). Demand analysis using strategic reports: An application to a school choice mechanism. *Econometrica*, 86(2):391–444.

Agarwal, N. and Somaini, P. J. (2022). Demand analysis under latent choice constraints. Technical report.

Arteaga, F., Kapor, A. J., Neilson, C. A., and Zimmerman, S. D. (2022). Smart matching platforms and heterogeneous beliefs in centralized school choice. *The Quarterly Journal of Economics*, 137(3):1791–1848.

Artemov, G., He, Y., and Che, Y.-K. (2017). Strategic " Mistakes ": Implications for Market Design.

Bobba, M. and Frisancho, V. (2019). Perceived ability and school choices.

Bostian, A. A., Holt, C. A., and Smith, A. M. (2008). Newsvendor "pull-to-center" effect: Adaptive learning in a laboratory experiment. *Manufacturing & Service Operations Management*, 10(4):590–608.

Chen, L. and Sebastián Pereyra, J. (2019). Self-selection in school choice. *Games and Economic Behavior*, 117:59–81.

Chen, Y. and Sönmez, T. (2006). School choice: An experimental study. *Journal of Economic Theory*, 127(1):202–231.

de Haan, M., Gautier, P., Oosterbeek, H., and Klaauw, B. V. d. (2023). The performance of school assignment mechanisms in practice. *Journal of Political Economics*.

Dreyfuss, B., Heffetz, O., and Rabin, M. (2019). Expectations-based loss aversion may help explain seemingly dominated choices in strategy-proof mechanisms.

Fack, G., Grenet, J., and He, Y. (2019). Beyond truth-telling: Preference estimation with centralized school choice and college admissions. *American Economic Review*, 109(4):1486–1529.

Gale, D. and Shapley, L. S. (1962). College admissions and the stability of marriage. *American Mathematical Monthly*, 69(1):9–15.

Haeringer, G. and Klijn, F. (2009). Constrained school choice. *Journal of Economic Theory*, 144(144):1921–1947.

Hassidim, A., Marciano, D., Romm, A., and Shorrer, R. I. (2017). The mechanism is truthful, Why aren't you? In *American Economic Review*, volume 107, pages 220–224.

Hassidim, A., Romm, A., and Shorrer, R. I. (2020). The Limits of Incentives in Economic Matching Procedures. *Management Science*, (October):1–13.

Kapor, A. J., Neilson, C. A., and Zimmerman, S. D. (2020). Heterogeneous beliefs and school choice mechanisms. *American Economic Review*, 110(5):1274–1315.

Larroucau, T. and Ríos, I. (2018). Do "Short-List" Students Report Truthfully? Strategic Behavior in the Chilean College Admissions Problem.

Larroucau, T. and Rios, I. (2021). Dynamic college admissions and the determinants of students' college retention.

Rees-Jones, A. (2018). Suboptimal behavior in strategy-proof mechanisms: Evidence from the residency match. *Games and Economic Behavior*, 108:317–330.

Rees-Jones, A. and Skowronek, S. (2018). An experimental investigation of preference misrepresentation in the residency match. *Proceedings of the National Academy of Sciences of the United States of America*, 115(45):11471–11476.

Rios, I., Larroucau, T., Parra, G., and Cominetti, R. (2020). Improving the Chilean College Admissions System. *Operations Research (forthcoming)*.

Rodriguez, J., Urzua, S., and Reyes, L. (2016). Heterogeneous Economic Returns to Post-Secondary Degrees: Evidence from Chile. *Journal of Human Resources*, 51(2):416–460.

Shorrer, R. I. and Sóvágó, S. (2021). Dominated Choices in a Strategically Simple College Admissions Environment : The Effect of Admission Selectivity.

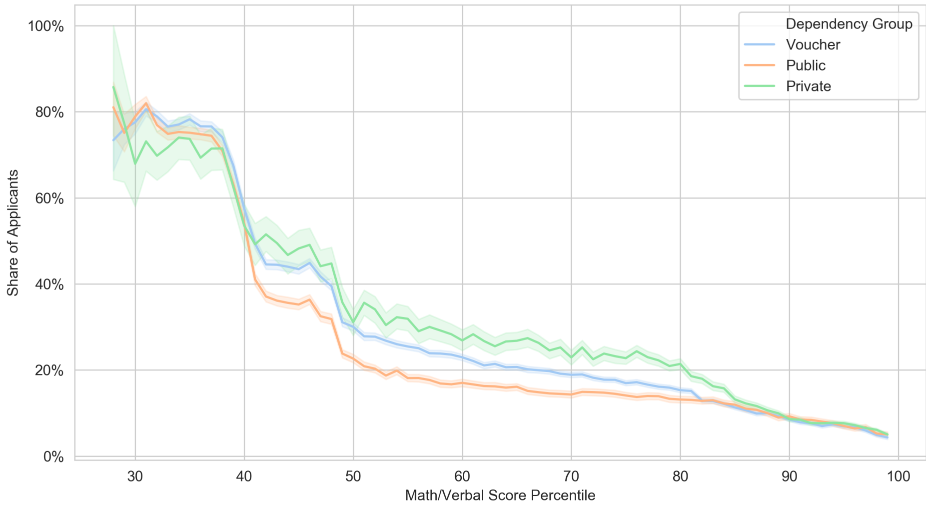
Tincani, M. M., Kosse, F., and Miglino, E. (2022). The effect of preferential admissions on the college participation of disadvantaged students: The role of pre-college choices.

Appendix

A Appendix Admissibility Mistakes, Misreporting & Subjectuve Beliefs

A.1 Admissibility Mistakes

Figure A.1: Share of students with *admissibility* mistakes by average score and school type



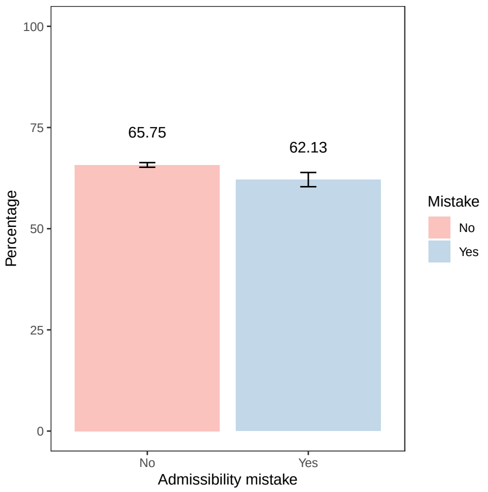
Notes: The share is computed as the total number of students in admission process 2005-2018 who submitted a ROL with at least one *admissibility* mistake, over the total number of applicants per bin of score percentiles and school type. The solid line is a conditional mean computed with a bandwidth of 1 score percentiles and shaded region corresponds to its 95% confidence interval. The score percentiles are computed with respect to the population of students who participated in the admission process and had a valid average Math/Verbal score.

Figure A.2 shows the percentage of students who declare to know the admission requirements for a subset of the programs listed in their applications. We compute this statistic only for programs with an active requirement and group it by different admission requirements and whether the student made an *admissibility* mistake or not. Overall, we observe heterogeneity in the level of knowledge by requirement type and significant differences between the group of students who made an *admissibility* mistake or not. Between 60% to 75% of students who did not make an *admissibility* mistake declare to know the minimum scores' requirements and specific tests. However, close to 50% to 60% declare to know these requirements from the group who made an *admissibility* mistake.

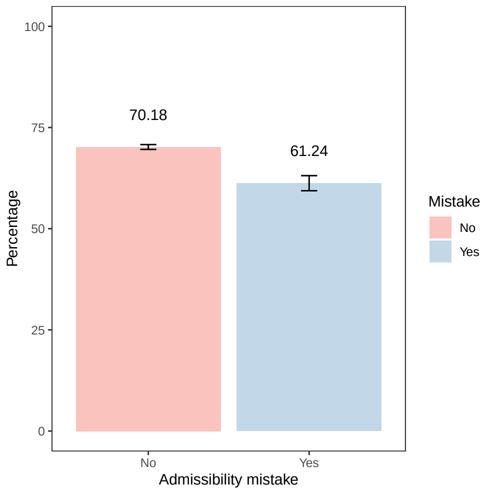
Figure A.3 shows the percentage of students who gave a correct answer for each requirement where they declared to have the correct knowledge. We observe a low level of correct knowledge about requirements and heterogeneity by requirement type. The requirements for the restriction in the preference ranking and number of vacancies are the lowest.

Figure A.2: Knows admission requirements

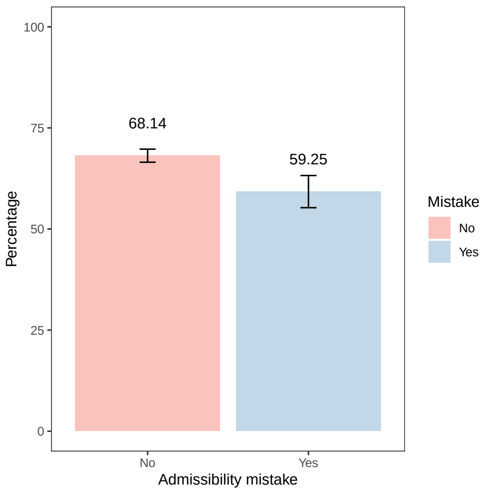
(a) Minimum Math-Verbal Score



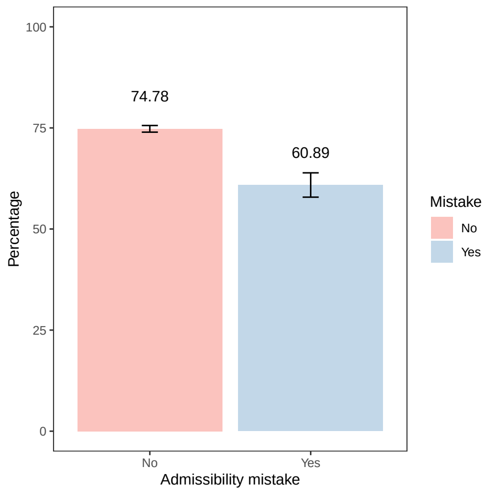
(b) Minimum Weighted Score



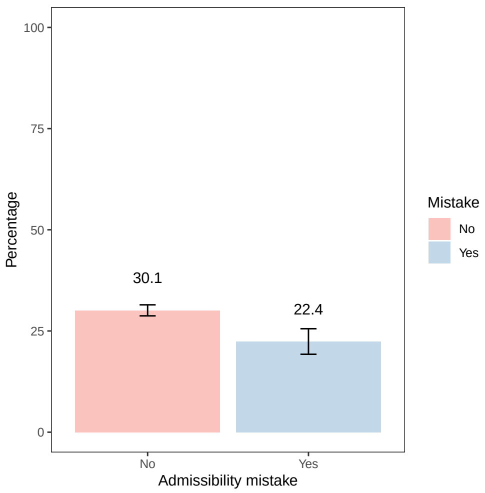
(c) Requires History



(d) Requires Science



(e) Restricts Preference Ranking



(f) Number of Vacancies

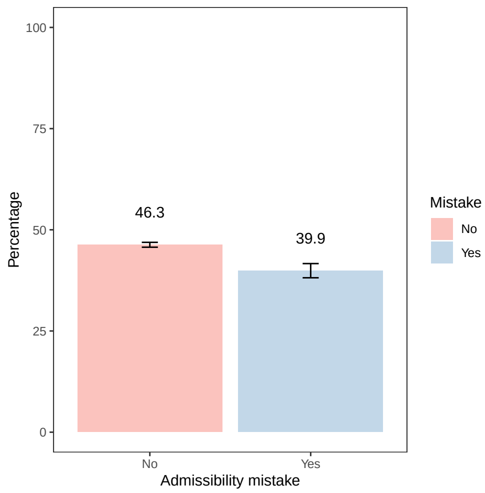
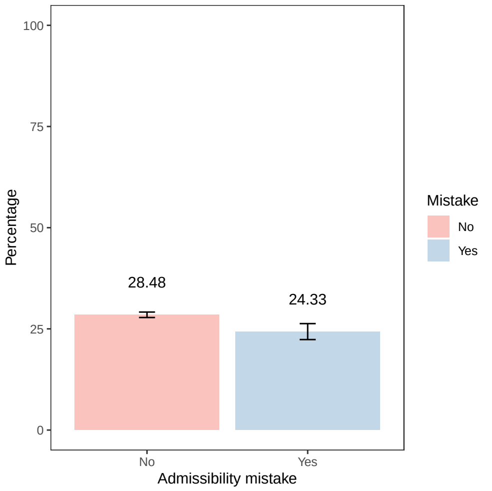
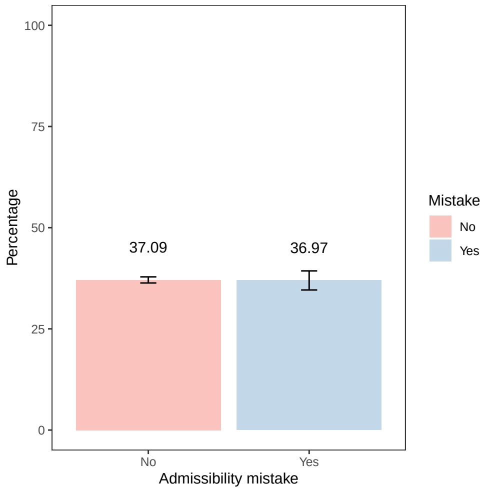


Figure A.3: Knows correctly admission requirements

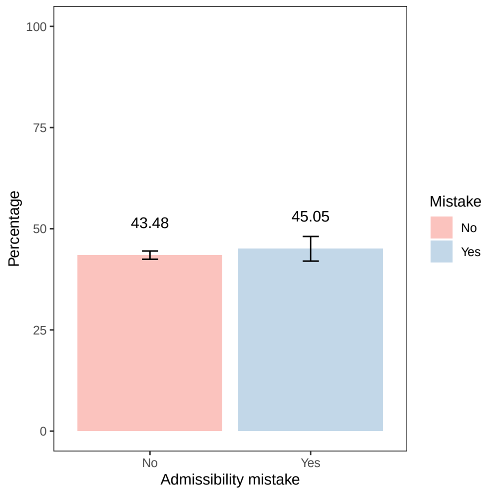
(a) Minimum Math-Verbal Score



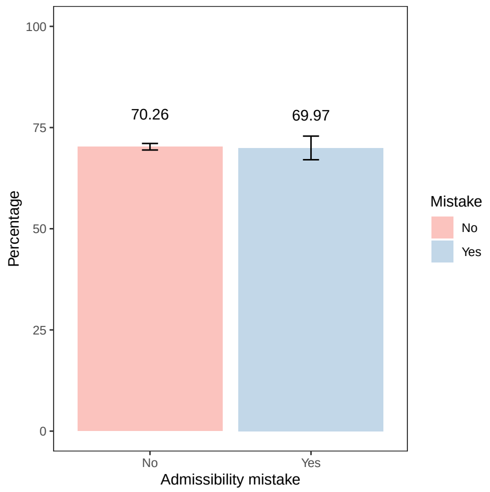
(b) Minimum Weighted Score



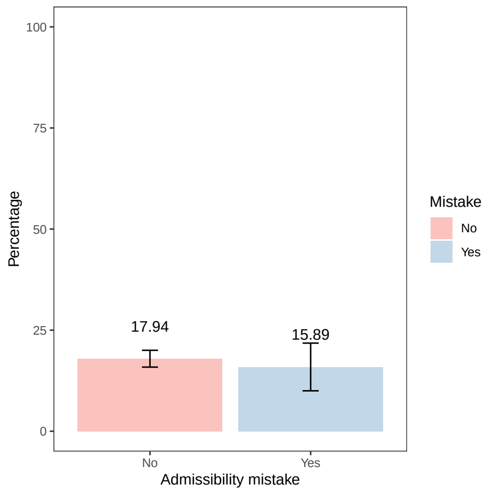
(c) Requires History



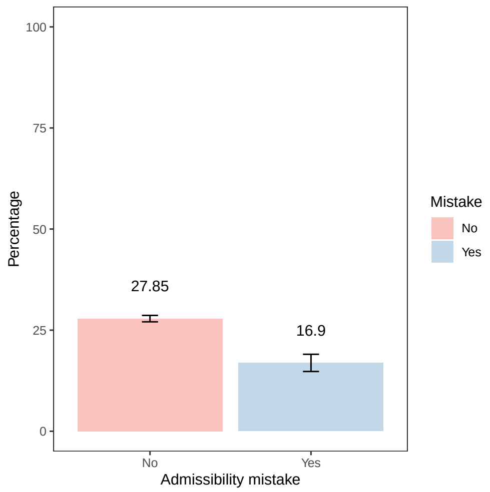
(d) Requires Science



(e) Restricts Preference Ranking



(f) Number of Vacancies



Discussion: the low levels of correct knowledge might suggest a high measurement error in the survey, mainly because the questions about the survey requirements were at the end of the survey. Doing an in-depth analysis of the responses about the admission requirements, we observe that a significant fraction of students seems to confuse the requirement types. For instance, (i) responding to the value of the minimum weight score requirement when asked about the value of the minimum average math-verbal score requirement; (ii) responding to the current preference of the program in the list instead of the preference restriction of the program, and (iii) confusing when a program allows the student to take the Science (History) test with when the program requires that one of these tests is taken. In this sense, we attribute these low levels of correct knowledge to be a combination of students not understanding the system's rules and misinterpreting the survey question. However, we do not find evidence that responses are randomly selected.

A.2 Misreporting

Table A.1: Reasons for misreporting

Application type	Misreporting Exclusion [%]	Misreporting Ordering [%]	Truth-teller [%]
Reasons			
(a) YES, I did apply to my ideal program as a top-reported preference	20.33	29.02	86.22
(b) My admission probability to that program is too low	50.21	46.53	9.08
(c) The program is too hard and I don't think I would be able to graduate from it	3.06	1.56	0.37
(d) I do not have the monetary resources to pay for the program	25.44	20.2	4.73
(e) To include my ideal program, I would have to exclude some program from my list	2.2	2.53	0.36
(f) The decision to where to apply did not depend only on me, and it was influenced by other people (family, friends, etc.)	6.5	8.28	1.17
(g) I thought that by including this program in my list I would have reduced my chances of being admitted to the other listed programs	6.6	4.78	0.53
(h) Given that my admission chances are too low, I prefer to do not list this program and being assigned to a higher reported preference	36.3	14.02	1.22
Other	13.91	12.74	1.74
Total	6184	1861	6939

Note: respondents can choose multiple reasons.

A.3 Subjective beliefs

Figure A.4a shows a heat-map of the average subjective beliefs on admission probabilities by preference (P1-P10) and application length (L1-L10). We observe that given an application length, there is a positive gradient in the average admission probability by preference (position of the program in the list). This pattern could be explained if students tend to list programs with low admission probabilities as their most preferred ones

Table A.2: Reasons for misreporting (consistent responses)

Application type	Misreporting Exclusion [%]	Misreporting Ordering [%]	Truth-teller [%]
Reasons			
(b) My admission probability to that program is too low	64.2	70.92	-
(e) To include my ideal program, I would have to exclude some program from my list	2.12	3.43	-
(f) The decision to where to apply did not depend only on me, and it was influenced by other people (family, friends, etc.)	5.77	9.66	-
(g) I thought that by including this program in my list I would have reduced my chances of being admitted to the other listed programs	7.58	5.36	-
(h) Given that my admission chances are too low, I prefer to do not list this program and being assigned to a higher reported preference	46.98	18.99	-
Other	18.51	19.74	-
Total	3257	932	5983

Note: respondents can choose multiple reasons. Percentages are computed among the fraction of consistent respondents.

Table A.3: Reasons for misreporting conditional on making an ex-ante under-confidence (consistent responses)

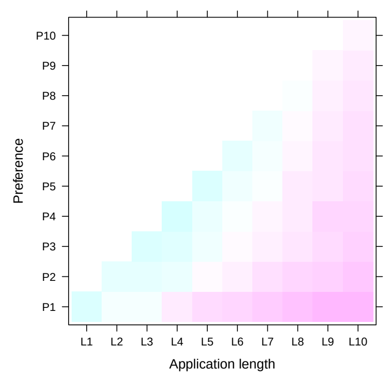
Application type	Misreporting Exclusion [%]
Reasons	
(b) My admission probability to that program is too low	28.52
(e) To include my ideal program, I would have to exclude some program from my list	1.9
(f) The decision to where to apply did not depend only on me, and it was influenced by other people (family, friends, etc.)	21.29
(g) I thought that by including this program in my list I would have reduced my chances of being admitted to the other listed programs	7.22
(h) Given that my admission chances are too low, I prefer to do not list this program and being assigned to a higher reported preference	17.11
Other	49.81
Total	263

Note: respondents can choose multiple reasons. Percentages are computed among the fraction of consistent respondents.

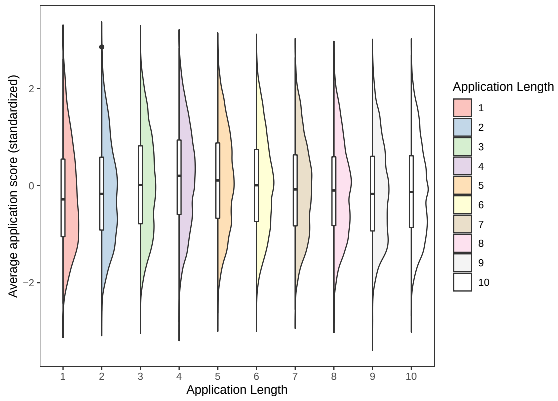
(*reach*) and programs with high admission probabilities at the bottom (*safety*). Furthermore, students who submit longer application lists tend to face lower admission chances at the top and the bottom of their application lists. To understand if students' scores explain this pattern, Figure A.4b shows the distribution of application scores by the length of application lists. We observe a non-monotonic relation between the median application score and the length of the application list, showing its peak for lists of length equal to 4. In addition, we observe that within the length of application lists, there is significant variation in the average application score. These data patterns suggest that the correlations observed in Figure A.4a cannot be explained only by systematic differences in scores for students who submit lists of different lengths.

Figure A.4: Subjective beliefs and preference of assignment

(a) Average subjective beliefs by preference and application length



(b) Distribution of application scores (standardized) by application length



B Appendix for Section 6

B.1 Admission Probabilities

To compute the admission probabilities, we use a bootstrap procedure similar to that in Agarwal and Somaini (2018) and Larroucau and Ríos (2018). The main difference is that these approaches use complete information regarding the applications. In our case, we only have the application list of close to 2/3 of the students that ended up applying, so running the bootstrap procedure on this sample would considerably underestimate the cutoffs. For this reason, our first task is to estimate the total number of students that would apply in 2022 based on the applications received so far. To accomplish this, we divide the population into three segments based on their average score between Math and Verbal (the two mandatory exams of the PSU/PDT). Then, using data from 2020 and 2021, we estimate which fraction of all students that take the national exam would apply to at least one program in the centralized system taking the average between these two years. Finally, comparing this number with the actual fraction of students in each score bin that have applied so far, we quantify the number of students that have not applied yet.

Based on the number of applicants missing, we perform 1000 bootstrap simulations, each consisting of the following steps:

1. Sample with replacement the number of students missing in each bin score, and incorporate the sampled students to the pool of applications received so far.
2. Run the assignment mechanism used in the Chilean system. See Rios et al. (2020) for a detailed description of the mechanism used in Chile to solve the college admissions problem.
3. Compute the cutoff of each program for both the regular and BEA admission processes.

As a result of this procedure, we obtain two matrices (for the regular and BEA processes) with 1000 cutoffs for each program. Hence, the next step is to estimate the distribution of the cutoff of each program in each admission track. To accomplish this, we estimate the parameters of a truncated normal distribution for each program and admission track via maximum likelihood. Then, using the estimated distributions, we evaluate the CDF on the application score of the student to obtain an estimate of the admission probability, taking into account whether the student participates only in the regular process or also in the BEA track.

B.2 Recommendations

The recommendation algorithms works as follows.

1. Find the most and the second most popular majors based on the preferences included in the student's ROL.
2. For each pair of majors, and considering the most and the second most preferred major of each student, compute a transition matrix that returns the probability that a given major is followed by another major as the most preferred ones.
3. For each student, compute the set of feasible majors considering the student's scores and her admission probabilities (obtained as described in the previous section).
4. For students with high scores (i.e., average between Math and Verbal above 600), choose four majors according to the following rule:
   (a) Choose most preferred major according to the student's list of preferences,
   (b) Choose the second most preferred major according to the student's list of preferences,
   (c) Choose the major with the highest average wage[26] among all majors considering the transition matrix previously computed,

[26] Average wages are measured at the fourth year after graduation. This statistic is computed by SIES and provided to us by MINEDUC.

(d) Choose the major with the highest average wage among all feasible majors (i.e., majors for which the student has a positive probability of assignment) considering the transition matrix previously computed.

5. For students with low scores (i.e., average between Math and Verbal below 600), choose four majors according to the following rule:

   (a) Choose the most preferred major according to the student's list of preferences,

   (b) Choose the second most preferred major according to the student's list of preferences,

   (c) Choose the major with the highest expected wage among all majors belonging to IPs or CFTs,

   (d) Choose the major with the highest expected wage among all feasible majors (i.e., majors for which the student has a positive probability of assignment) considering the transition matrix previously computed.

B.3 Treatment Assignment and Stratification

As discussed in Section 6.2, we assign students to treatments in a stratified way to achieve balance. For the stratification we consider the following observables:

- Female: dummy variable equal to 1 if the student is female, and 0 otherwise.
- Region: categorical variable that takes four 3 levels depending on the region where the student graduated from high-school. Specifically, this variable is equal to 1 for students graduating in the north (regions I, II, III, IV and XVII); 2 for students graduating in the center (regions V, XIII, VI, VII); and 3 for students graduating in the south (regions VIII, IX, X, XI, XII, XIV and XVI).
- Score: categorical variable that takes 4 levels depending on the average score between the PDT tests in Math and Verbal. Specifically, this variable is equal to 1 for students with average score below 450; 2 for students which average score between 450 and 600; and 3 for students with score above 600.
- Overall alert: as discussed in Section 6.1.2, there are three types of overall alerts: (i) reach, (ii) safety, and (iii) more information. Each student can be assigned to one of these groups, and thus we also use this assignment as part of the stratification.
- Opened scores' intervention: when the scores of the PDT were published, MINEDUC ran an experiment aiming to provide information regarding the relative position of students among their peers (their high-school and their region). Hence, we use a dummy variable equal to 1 if the student received that intervention (and 0 otherwise) as part of our stratification.
- SMS: dummy variable equal to 1 if the student received an SMS encouraging them to open their personalized website, and 0 otherwise.

In Table B.1 we report the results of a multinomial regression models that consider the treatment assigned as dependent variable and the aforementioned variables as controls. The first three columns report the results considering all observations, while the last three columns report the resulting excluding misfits. We observe that none of the variables considered is significant, which confirms that our treatment assignment is balanced in terms of these covariates.

Table B.1: Treatment Assignment: Balance Checks

	Dependent variable: Treatment					
	All observations			Excluding misfits		
	(1)	(2)	(3)	(4)	(5)	(6)
Region - Center	−0.011	−0.007	−0.003	−0.010	−0.007	−0.003
	(0.026)	(0.026)	(0.026)	(0.027)	(0.027)	(0.027)
Region - South	−0.007	−0.004	−0.002	−0.005	−0.004	−0.002
	(0.029)	(0.029)	(0.029)	(0.029)	(0.029)	(0.029)
Female	−0.002	−0.001	−0.001	−0.002	0.00002	−0.0001
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
Score - Medium	−0.008	−0.004	−0.002	−0.006	−0.004	−0.002
	(0.032)	(0.032)	(0.032)	(0.033)	(0.033)	(0.033)
Score - High	−0.003	−0.001	−0.001	−0.001	0.0004	0.0001
	(0.036)	(0.036)	(0.036)	(0.037)	(0.037)	(0.037)
Overall Alert - Safety	−0.014	−0.008	−0.004	−0.012	−0.008	−0.004
	(0.038)	(0.038)	(0.038)	(0.039)	(0.039)	(0.039)
Overall Alert - Information	−0.017	−0.010	−0.005	−0.017	−0.011	−0.007
	(0.031)	(0.031)	(0.031)	(0.032)	(0.032)	(0.032)
Received SMS	0.007	0.005	0.003	0.007	0.005	0.002
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
Received Previous Intervention	0.037	0.023	0.011	0.034	0.022	0.010
	(0.036)	(0.036)	(0.037)	(0.038)	(0.038)	(0.039)
Constant	0.031	0.018	0.010	0.029	0.018	0.010
	(0.049)	(0.049)	(0.049)	(0.051)	(0.051)	(0.051)
Observations	107,837	107,837	107,837	106,100	106,100	106,100

C RATIONALIZING ROLS UNDER SUBJECTIVE BELIEFS AND MISTAKES

C.1 FRAMEWORK

We construct a model of students' application decisions under subjective and potentially biased beliefs. We show that, under minimal restrictions on students' behavior, one can identify students' preferences when they are strategic, without imposing strong assumptions on the beliefs they hold. The framework we consider allows for over-confidence and under-confidence mistakes.

We define the notations we use throughout the model in what follows. $\mathcal{R}_i$ denotes the ROL submitted by student i, while $\mathcal{R}_i(r)$ is the program ranked by i at rank r. When

i ranks program j before j' in $\mathcal{R}_i$, we write $j\mathcal{R}_ij'$. If program j is weakly preferred to program j' by student i, we write $j \succsim_i j'$. $\tilde{p}_{ij}$ denotes student i's subjective beliefs regarding her admission probability to program j. Finally, let $\mathcal{O}_i$ denote the set of programs to which student i believes to have a strictly positive probability of admission, that is $\mathcal{O}_i = \{j \in J : \tilde{p}_{ij} > 0\}$. There exists an outside option for students, denoted 0, which the student can always access. Thus, we defined student i's acceptance set as $\mathcal{A}_i = \mathcal{O}_i \bigcup \{0\}$.

We impose the following assumptions:

Assumption 2. *Conditional on her subjective beliefs, a student does not play a dominated strategy.*

Assumption 3. *Consider $j \in \mathcal{O}_i$ and $j' \notin \mathcal{O}_i$. If $j'\mathcal{R}_ij$, then $j' \succ_i j$.*

We will start by considering students who rank less than 10 programs. For them, under Assumption 2, the limit on the length of the list is not binding. Indeed, since the student does not play a dominated strategy, she would have included additional programs if it was payoff relevant to do so. Note that the fact that the limit on the length of the list is not binding does not imply that the student prefers the outside option to any unranked programs. Indeed, it could be that the student i) believes she has a probability of one of being admitted to a ranked program, in which case it is payoff irrelevant to rank another program she prefers to the outside option or ii) she prefers some unranked programs to the outside option but believes she has a zero probability of being admitted to these programs.

Assumption 3 simply states that, even if it is payoff-irrelevant to do so, a student will never rank a program she prefers less before a program for which she thinks she has a strictly positive probability of admission.

C.2 IDENTIFICATION

The following proposition can be derived:

Proposition 1. *Under Assumptions 2 and 3, in a constrained student-optimal DA mechanism, when the limit on the length of the list is not binding, $j = \mathcal{R}_i(1) \implies j \succsim_i j', \forall j' \in \mathcal{A}_i$.*

Our identification strategy of students' preferences under subjective beliefs will explicitly exploit Proposition 1. Indeed, it states that, under the above assumptions and in a student-optimal DA mechanism, students for whom the limit on the length of the ROL is not binding should submit a ROL such that the top-ranked program is preferred to any programs for which she believes she has a strictly positive probability of admission.[27]

[27] We observe unique survey data allowing to partially test whether this prediction is consistent with observed students' behavior. We observe that, among short-list students, 64% report ranking their top-true program first. Among those who do not, 75% explain that they anticipated that their probability of admission was too low. This suggests that more than 90% of the students in our sample behave in a way consistent with the predictions of Proposition 1.

We can thus rewrite the students' problem at rank 1 as follows:

$$\mathcal{R}_i(1) = \underset{j \in \mathcal{A}_i}{\operatorname{argmax}}\, v_{ij} \tag{C.1}$$

where v_{ij} is the indirect utility of student i from enrolling in program j. In this framework, the conditional choice probability (CCP) to apply to a given program at the top of the ROL only depends on the subjective beliefs of student i through the set $\mathcal{A}_i$. This implies that knowing this set is sufficient to be able to identify students' preferences. One challenge thus remains: how to construct the set $\mathcal{A}_i$?

We overcome this limitation by building on Agarwal and Somaini (2022) who establish identification of a general random utility model for consumer preferences with latent choice sets. We enrich this model with survey data on students' subjective beliefs about their admission probabilities. By doing so, we highlight an alternative use to this framework: it allows to identify students' preferences when (i) agents are strategic, (ii) have subjective beliefs which may differ from rational expectations, and (iii) the econometrician has access to partial information about unobserved choice sets.

Following Agarwal and Somaini (2022), we consider indirect utility of the following form:

$$v_{ij} = u_j(w_i, \omega_i) - g_j(w_i, d_{ij})$$

where w_i is a vector of observed student attributes; d_{ij} is a scalar observed attribute varying at the student-program level and ω_i is a random vector capturing unobserved student-specific unobserved heterogeneity. We rewrite $\mathcal{A}_i$ in terms of σ_{ij}, the acceptance policy functions:

$$\mathcal{A}_i = \{j \in J : \sigma_{ij} = 1\} \bigcup \{0\}$$

where σ_{ij} is defined as follows:

$$\begin{aligned} \sigma_{ij} &= \sigma_{ij}(w_{ij}, \omega_i, z_{ij}) \in \{0, 1\} \\ &= \mathbb{1}\{\tilde{p}_{ij}(w_{ij}, \omega_i, z_{ij}) > 0\} \end{aligned}$$

with z_{ij} a student-program specific observable scalar characteristic, which affects the perception of student i to be admitted to program j. This variable can only affect the subjective probability of student i with respect to program j, not program k. As mentioned by Agarwal and Somaini (2022), this rules out strategic interactions across programs on this dimension, but such interactions can happen through the dependence of the acceptance function to w_i.

We follow Agarwal and Somaini (2022) in imposing some restrictions.

Assumption 4. *The unobserved term ω_i is conditionally independent of the vector (d_i, z_i) given w_i.*

Assumption 4 implies that 1) each component d_{ij} shifts preferences without interacting with consumer-specific unobservables that affect either preferences or choice sets and 2) the unobserved determinants of preferences are independent of z_i given w_i.

Assumption 5. *The function $\sigma_j(w_i, \omega_i, z_{ij})$ is non-increasing in z_{ij}. Moreover, for all j, w_i and ω_i,*
$\lim_{z \longrightarrow -\infty} \sigma_j(w_i, \omega_i, z_{ij}) = 1$ *and* $\lim_{z \longrightarrow \infty} \sigma_j(w_i, \omega_i, z_{ij}) = 0$.

Assumption 5 requires that 1) the student is more likely to have a subjective probability of admission regarding program j which is strictly positive if the value of z_{ij} is lower and 2) that the perception of having a strictly positive probability of admission to a given program changes from 1 to 0 for some value of z, for each value of the other variables. Agarwal and Somaini (2022) show that, when g is known, the joint distribution of u_j and acceptance policy conditional on w is identified.

Note that we did not discuss the case of students who ranked 10 programs. For these students, it might not be a dominated strategy to omit from their ROL their most preferred program for which they think they have a strictly positive probability of admission. Only 13% of the students do submit a ROL where the limit is binding. If these students omit a program they prefer to those ranked, and for which they believe they have a strictly positive probability of admission, we would not be able to exploit the above identification argument.

The richness of our survey allows us to check whether it is the case. Among students for which the limit on the length of the list is binding, we observe that 48% omit their top-true program from the top of their ROL. However, among them, 72% do this because they anticipate that they would not be admitted. Overall, among the students for whom the limit on the ROL is binding, 86% of them either rank their top-true program first, or fail to do so because they expect a too low probability of admission. These behaviors would be consistent with the implications of Proposition 1 that we will exploit in our estimation strategy, such that the latter is valid for 90% of the overall applications observed.

C.3 Parametric Specification & Choice of Excluded Variables

Although the identification results provided by Agarwal and Somaini (2022) are non-parametric, we follow them in our empirical application by assuming a parametric specification. we consider:

$$v_{ij} = \beta w_i x_j - d_{ij} + \epsilon_{ij} \tag{C.2}$$

and

$$\sigma_{ij} = \mathbb{1}\{\tilde{p}^*_{ij} \equiv \alpha w_i - z_{ij} + \nu_{ij} \geq 0\} \tag{C.3}$$

where w_i are student i's observable characteristics, x_j are program j's observable characteristics, ϵ_{ij} and ν_{ij} are idiosyncratic shocks. We allow unobserved match-specific correlation, by allowing for ϵ_{ij} and u_{ij} to be jointly normally distributed with mean zero and estimated covariance matrix Σ. $\tilde{p}^*_{ij}$ is the latent variable governing whether a program is in a student i's acceptance set: $\tilde{p}_{ij} > 0$ if $\tilde{p}^*_{ij} \geq 0$ and $\tilde{p}_{ij} = 0$ if $\tilde{p}^*_{ij} < 0$.

The shifter of students' preferences that is excluded from the acceptance policy function is the distance between the student and program location. This implies that distance matters to student directly in their indirect utility, but does not play a role in determining whether the student thinks she has a strictly positive probability of admission. This restriction would be violated if distance is correlated with students' perception of having a zero-probability of admission. We rule out such correlations. First, it is unlikely that students believe their admission probability will depend on the distance between their residence and a program's location. The different components weighted by programs are made clear on the platform, and they only mention test scores at the different test taken by the student, GPA and high-school rankings. Second, while it may still be true that students have some perception of selectivity which may be different across universities located in different regions, we will capture that by conditioning on the region where the program is located. Conditional on the latter, it is thus plausible that distance per se does not shape students perceived feasible sets.

The shifter of students' perceived feasible set, z_{ij}, that is excluded from students' indirect utility is the distance between the program's cutoff from the year before and the student's weighted score to this program, which we interact with the treatment status of the student.

We inform the specification of our acceptance policy function by exploiting our survey. In particular, we use the reported subjective beliefs, which we observe for a various set of both ranked and unranked programs, to predict the probability that a student excludes a program from her acceptance set $\mathcal{A}_i$, controlling for a rich set of of students' and programs' characteristics.

C.4 Estimation

The model can be estimated using the Gibbs sampler. Let us follow Agarwal and Somaini (2022) in defining the cutoff quantity $\Pi_j(w_i, \omega_i) = sup\{z : \sigma_j(wi, \omega_i, z) = 1\}$. Under Assumption 5, the function $\Pi_j(.)$ determines product j's acceptance policy for almost every z since $z < \Pi(w_i, \omega_i)$ implies $\sigma_j(w_i, \omega_i, z) = 1$ and $z > \Pi(w_i, \omega_i)$ implies $\sigma_j(w_i, \omega_i, z) = 0$.

The idea of the Gibbs sampler is that, once we condition on either the vector $\Pi_i = (\pi_{i1}, ..., \pi_{iJ})$ or $u_i = (u_{i1}, ..., u_{iJ})$, the problem becomes standard and tractable. Indeed, Π_i determines the latent choice set, such that the rest of the problem is a standard discrete choice model, while conditional on u_i, i matches with j if and only if $\pi_{ij} > 0$ and $\pi_{ij'} < 0$ for all j' with $u_{ij'} > u_{ij}$. The sampler will iterate between data augmentation steps for π_i and u_i.

Algorithm: The sampler starts with an initial guess for the parameters (α, β, Σ) and for the latent variables $(v_i, \epsilon_{i0}, \pi_i)$ for every i. This guess is denoted by $\Theta(0)$. For each draw k, the following steps are performed:

1. Data augmentation:
   (a) Draw the latent acceptance index $\pi_{ij}|\Theta^{k-1}$ for every i and j. The posterior distribution of π_{ij} conditional on all the parameters θ^{k-1} is Normal.

If i was allocated to program j, then we draw π_{ij} from the conditional posterior truncated by $\pi_{ij} \geq z_{ij}$. If i was matched to program $j* \neq j$ and $v_{ij}^{k-1} > v_{ij*}^{k-1}$, then π_{ij} is drawn from the conditional posterior truncated by $\pi_{ij} < z_{ij}$. Otherwise, we draw it from the conditional posterior without any truncation.

Let π^k denote the vector of draws and let $\mathcal{A}_i^k = \{j \in J : \pi_{ij} \geq z_{ij}\}$.

(b) Draw the latent utility $v_{ij}|\Theta^{k-1}$ for every i and j. The posterior distribution of v_{ij} conditional on all the parameters θ^{k-1} and on π^k is Normal. Let $j*$ be the program where i enroll. Draw v_{ij*} from the conditional posterior truncated at $v_{ij*} \geq \max_{j \in \mathcal{O}_i^k} v_{ij}$. Denote it by v_{ij}^k. Then, draw v_{ij} for $j \in \mathcal{O}_i^k \setminus \{j*\}$ from the conditional posterior truncated at $v_{ij} \leq v_{ij*}^k$. Lastly, draw v_{ij} for $j \notin \mathcal{O}_\rangle^\|$ from its unconditional posterior without any truncation. Let v^k denote the vector of draws.

(c) Seemingly unrelated Bayesian regression: with the draws of v^k and π^k and for fixed value of ϵ_{i0}^{k-1}, the equations above form a system of seemingly unrelated regressions. The posterior distributions of the parameters α, β are Normal and the posterior distribution of Σ is inverse Wishart. We draw these parameters and obtain the resulting residuals $\hat{\epsilon}_{ij}^k$ and $\hat{\eta}_{ij}^k$.

(d) Update random effects

(e) Update the variance of the random effects

(f) Collect all parameter draws in step k and denote them by θ^k.

C.5 COUNTERFACTUALS

We use our estimated parameters to perform three key sets of counterfactuals. The first set of counterfactuals we will consider is aimed at quantifying the costs of information frictions on students' welfare. Our paper is the first to measure such welfare costs under a DA mechanism and in the high-stakes context of a country-wide college application mechanism. Our second set of counterfactuals will test different mechanisms and information policies which could potentially improve students' welfare given that they have subjective and potentially biased beliefs. Finally, our last set of counterfactuals will allow to understand the sources of students' bias, by proposing a model of students' beliefs formation, while explicitly allowing beliefs to be correlated with preferences. This model will allow us to understand the sources of bias behind students' beliefs.

The framework displayed above allows to identify and estimate students' preferences over programs, allowing them to be strategic and to have subjective beliefs, under minimal restrictions over their ranking behavior and the beliefs they hold. The implementation of our counterfactual exercises requires to impose some additional assumptions over students behavior in order to be able to simulate counterfactual ROL.

First, our counterfactual should be such that they do not imply that the limit on the length of students' ROL is now binding. Second, we assume that, with some probability - that we estimate from the data - a student can decide to rank a program while it is payoff-irrelevant to do so. In particular, we denote q_{ir} the probability that a student ranks a program at rank r when $\tilde{p}_{iR_i(r-1)} = 1$. We also let Π_{ir} denote the probability to rank a program j such that $\tilde{p}_{ij} = 0$ at rank r. From this, we can simulate the choice of student i at rank r as follows:

$$R_i(r) = \begin{cases} \Pi_{ir} \operatorname{argmax}_{\bar{\mathcal{A}}_i} v_{ij} - (1 - \Pi_{ir}) \operatorname{argmax}_{\mathcal{A}_i} v_{ij} & \text{if } \tilde{p}_{iR_i(r-1)} \neq 1 \\ q_{ir} \left[\Pi_{ir} \operatorname{argmax}_{\bar{\mathcal{A}}_i} v_{ij} - (1 - \Pi_{ir}) \operatorname{argmax}_{\mathcal{A}_i} v_{ij} \right] & \text{if } \tilde{p}_{iR_i(r-1)} = 1 \end{cases} \tag{C.4}$$

We can thus correct students' beliefs, which would change the set of programs included in the set $\mathcal{A}_i$. Thanks to the above equation, we can simulate the new ROL submitted by the student under correct beliefs, which allows to quantify the cost of information frictions on students' welfare.