Search and Biased Beliefs in Education Markets

Patrick Agte
Princeton University

Claudia Allende
Stanford GSB

Adam Kapor
Princeton University

Christopher Neilson
Yale University

*

August 2, 2023

Abstract

When learning about schools requires costly search, search decisions depend on families' beliefs about the returns. This paper asks how families' (limited) awareness of schools and (inaccurate) beliefs about schools' prices, quality ratings, and placement chances distort their search efforts and application decisions in the context of Chile's nationwide centralized school choice process. We combine novel data on search activity with a panel household surveys, administrative application data, randomized information experiments, and a model of demand for schools. We find that households are unaware of many relevant schools, and hold inaccurate beliefs about admissions chances, prices, and quality scores. The average household underinvests in search because it is excessively optimistic about the quality rating of the schools it knows. Correcting this bias causes students to match to schools with higher qualitym approaching what can be achieved under a full-information benchmark, but providing this information broadly and early, before search takes place, discourages some households from applying and reduces placement rates.

*Emails: pagte@princeton.edu, callende@stanford.edu, akapor@princeton.edu, christopher.neilson@yale.edu.

We thank JPAL-LAC and the implementation team of data scientists and developers at ConsiliumBots for their help throughout the project. The authors wish to thank the government partners that made this research possible, in particular the leadership at the Sistema de Admission Escolar (SAE) of the Ministry of Education in Chile. The authors would like to thank Sofía Valdivia, Isabel Jacas, and María Elena Guerrero for excellent research assistance, Mariel Beyoda and Eduardo Vargas at J-PAL for their help during the fieldwork. All errors are our own.

1 Introduction

Families participating in markets for schools often choose low-performing schools when higher-performing schools are available.[1] Moreover, many families who have many available options apply to relatively few schools, even when applications are free. Some of these families may find themselves without an offer, an outcome they consider very undesirable, while others regret their decisions and renege on their assignments.

While in principle these outcomes may reflect families' preferences, in practice it may be difficult to learn one's choice set and form preferences over the relevant options. A typical household participating in a school choice process faces the challenge of constructing an application portfolio without knowing which schools will offer a placement. In settings with many available options, discovering the relevant schools, and evaluating their characteristics in order to form a ranking, requires search effort. When this effort is costly, households' search decisions may depend on their beliefs about returns. A household that overestimates its chance of being admitted by a desirable school, underestimates the desrability of schools it has not yet looked into, or overestimates the quality of schools it "knows" may stop searching too early, leading to adverse outcomes.

This paper investigates the nature, extent, and impact of families' limited awareness and imperfect information about schools in a large "school choice" market. We ask the following research questions. First, how do families' (limited) awareness of schools and (inaccurate) beliefs about their characteristics interact with their preferences and search costs to distort their search efforts, application decisions, and school assignments? Second, what do effective information interventions look like: who should be targeted, at what time, with what information, to what effect?

We address these questions in the context of Chile's nationwide centralized school assignment process, in which 461,223 students were assigned to 7,979 schools via a student-proposing deferred acceptance algorithm in 2021, the year of our study. To do so, we exploit novel data and field experiments. We created and provided a "school explorer" web app with personalized information about schools. We used this app to conduct RCTs of two interventions early in households' search process: reducing the cost of finding high-quality schools, and providing information about the distribution of schools' price and quality. In addition, we use the application platform itself to conduct a large-scale randomized "report card" intervention, providing information about the price and quality of schools already known as well as suggestions for further search.[2] We conducted multiple rounds of household surveys to measure households' preferences, beliefs, and awareness of schools, pre- and post-treatment(s). We linked households' survey responses to their search activity on the platform, and to administrative data on demographics and submitted

[1] Hastings and Weinstein (2008)

[2] In addition, we use the platform provide "warnings" to students with low chances of receiving a placement; see Arteaga et al. (2022).

applications.

We use these data and experiments as follows. We first describe households' preferences, their awareness of schools, their search activity, and their beliefs about schools' prices, quality ratings, and admissions chances. Next, we evaluate the impacts of providing information on households' beliefs, search activity, and application decisions. Finally, we set up and estimate a model of households' search and application decisions, motivated by these results. Our model allows households to have limited awareness of schools, to imperfectly observe the price, quality, admissions chance, and match-level econometrician-unobserved characteristics of schools that they are aware of, and to hold subjective beliefs about the distribution of price, quality, and match-level characteristics of schools they have not yet investigated, which may differ from the objectve joint distribution of these objects. Households may discover schools, and develop more accurate beliefs about known schools, via costly search effort as well as passively via "offline" search. Once the costs of further search exceed the subjective expected benefits, households stop searching and submit rank-order lists to the centralized mechanism. With estimates in hand, we conduct counterfactual simulations, comparing the impacts of the interventions we have conducted to those of an ideal information intervention, and investigating the timing and targeting of information provision.

Our main finding is that households search too little, on average, because they are excessively optimistic about the quality of schools they know and like. The policymaker can achieve significant gains in the average quality of schools at which students are placed by correcting this misperception, but faces a tradeoff: given the search technology that is available, providing information broadly and "early", i.e. before search takes place, may discourage some students from engaging in search, and increase the number of students who do not receive a placement. Thus, to the extent that the policymaker wants to encourage more search and greater demand for quality, according to preliminary estimates it may be more effective to provide targeted information about schools to which households have already listed on their application, or to provide information to those households that appear to be making mistakes, rather than to provide general information as early as possible.

We show this finding in three steps. First, we conduct descriptive analyses of the survey data. Households are aware of less than 50% of random schools asked about at baseline. Thus, there is a role for search effort. Surprisingly, the average household is not pessimistic about the distribution of academic quality—according to a known, government-calculated quality index—of schools not yet investigated. However, households perceive the characteristics of "known" schools with noise, which leads them to systematically overestimate the quality of their first-choice school. Their rankings of schools are sensitive to perceived quality and moreover, conditional on their survey responses, do not depend on the schools' "true" quality. That is, we do not find biased beliefs about the distribution from which schools are drawn, but rather excess optimism about the quality of known schools, a characteristic that households value.

Second, we analyze our experiments. Explorer search activity predicts knowledge of schools, and accurate beliefs about their characteristics. However, while our information treatments on the distribution of schools' price and quality shift households' reported beliefs, they have small effects on average on their search effort and application portfolios.[3] In contrast, the "report card" intervention, which corrects misperceptions about known schools, has large impacts on households' awareness sets, application lists, welfare, placement rates, and the characteristics of placed schools, including their academic quality.

Third, in counterfactual simulations in which we provide information at scale, we find that in order to sort students into schools with high quality indices, the policymaker must provide students with information about schools' quality, but the timing of this information matters.[4] A hypothetical intervention that provides full knowledge of "known" schools' price and quality just before applications are due, but does not affect awareness of schools, would achieve 89% of the impacts on average school quality of endowing all students with full information about all schools.[5] However, 80% of students in our search intervention sample receive a placement in some school in the data, but this share would fall to 77% if students were given information about schools' price and quality just before applications are due, and to 73% if students were given accurate information about quality, price, and admissions chances early, as this information would discourage search effort.

We believe our model estimates and simulations are important inputs for public policy, where in order to help families make informed choices, we need to understand how people search in practice. Our data are produced by a centralized mechanism, which gives us rich data on students' submitted rank-order lists. The lessons here may apply more broadly, however, to other settings in which people actively search for options under uncertainty about their availability.

We provide a unified treatment of information frictions in school choice mechanisms. Our experiments and structural model allow comparisons, in the same setting, between the impacts of providing information about admissions chances (Arteaga et al., 2022; Ajayi et al., 2020; Gurantz et al., 2021; Hoxby et al., 2013; Hoxby and Turner, 2015; Luflade, 2017) and those of providing a list of characteristics of schools (Hastings and Weinstein, 2008; Mizala and Urquiola, 2013; Corcoran et al., 2018; Cohodes et al., 2022; Andrabi et al., 2017; Allende et al., 2019; Bergman et al., 2020), from which it had been difficult to extrapolate.

Although Chile uses a SPDA mechanism with no constraint on list length, following best prac-

[3]We find impacts on the second-ranked school. Impacts on average characteristics are not significant. In future drafts we will explore heterogeneity by SES.

[4]In this draft, the model is estimated on "search sample" participants who were given early access to the explorer. Future drafts will use the universe of 2021 school choice participants in entry grades.

[5]Quality indices range from 1 to 4. At baseline, the average quality of the placed school is 3.069, while in this counterfactual it is 3.228, and in our full-information benchmark it is 3.247. The effect is thus 89% = (3.228 - 3.069)/(3.247 - 3.069).

tices in market design, and under which an optimal strategy is to report schools in true preference order (Abdulkadiroğlu and Sönmez, 2003a; Abdulkadiroglu et al., 2005; Correa et al., 2019), the presence of search costs makes beliefs about admissions chances relevant to households' decisions (Arteaga et al., 2022). We contribute to an empirical literature on consumer search (Santos et al., 2012, 2017; Dinerstein et al., 2018; Hodgson and Lewis, 2020; Moraga-González et al., 2021) by providing novel data on search activity and outcomes, as well as experimental variation in the factors affecting perceived returns to search, in a high-stakes setting. Our data tell us what a reasonable model of search looks like, and allow us to estimate it.

Our descriptive analysis of information frictions in a high-stakes setting parallels work on these topics in health economics (e.g. Handel and Kolstad (2015)). We estimate our model via a Gibbs sampler, building on prior work estimating demand (McCulloch and Rossi, 1994) in school choice settings (Agarwal and Somaini, 2018; Kapor et al., 2020) with limited availability (Kapor et al., 2022; Agarwal and Somaini, 2022). We extend these specifications by allowing misperceptions of school characteristics.

A limitation of this paper is that we estimate demand, and conduct "single-agent" counterfactuals, holding school characteristics fixed.[6] Understanding search and demand is a needed input for further research that models equilibrium in this market.

2 Empirical Setting

We implemented three surveys and two experiments in partnership with an ed-tech NGO and the Ministry of Education of Chile. We provided the school explorer software, and randomly assigned a set of "search" interventions using this software, to a sample of families with young children who agreed to participate in a pilot information project. We call this sample the "search sample". In addition, we randomly assigned "report card" interventions to those families that participated in the centralized choice system in Chile and agreed to participate in our study. We refer to this sample as the "application sample" We drew households from this sample to conduct a final "satisfaction" survey.

We start by describing the centralized school choice system in Chile. We return to the surveys and inteventions in section 2.4.

[6]Counterfactual simulations in this draft hold schools' prices, quality, unobserved match quality, and admission chances fixed. We observe that the modal first-choice school has excess capacity. Future drafts will evaluate interventions that hold capacities and the assignment mechanism fixed, allowing placement chances to adjust.

2.1 The Centralized School Choice System in Chile

Chile introduced nationwide, voucher-based school choice in 1981 (Hsieh and Urquiola 2006). Students receive vouchers they can spend at schools, and schools may charge additional fees, regulated by the Ministry of Education. For the first 35 years, school choice in Chile was decentralized. Families applied to each school separately. A series of education reforms relevant to this paper have been implemented in the last 15 years.

In 2008, the targeted voucher policy (*Subención Escolar Preferencial (SEP)*) eliminated out-of-pocket fees for 40% of the poorest students and significantly increased transfers to schools. By 2021, 93.0% of the schools that receive vouchers had a SEP agreement.

In 2011, Chile created the Education Quality Agency, an autonomous public service whose mission is to evaluate, inform, and guide the schools and families decision processes to contribute to the assurance of high educational standards. This agency implements the national standardize test (SIMCE) and their rating system substitutes previous policy efforts to inform families about schools. Since 2016 the Agency assigns a performance category for each educational institutions recognized by the State. These categories are the result of a comprehensive evaluation that classifies establishments into High, Medium, Medium-Low and Insufficient performance, with the purpose of complying with the State's duty to ensure the right of all students to have the same opportunities to receive a high quality education (*Agencia de Calidad de la Educación*).

In 2016, policy makers adopted centralized assignment with the goal of making the school choice process more transparent and equitable (*Gobierno de Chile, Ministerio de Educacion, 2017*). The centralized choice system was rolled out on a region-by-region basis, with adoption in all cities by 2019 and all grades by 2020. The centralized process includes 93% of primary school matriculation in the country, covering almost all public schools and private schools that accept school vouchers. In 2021, 450,000 applicants participated.

All cities in Chile use the same choice platform, which assigns students to schools using a student-proposing deferred acceptance algorithm (Correa et al., 2019). To ration seats in oversubscribed schools, the mechanism combines coarse sibling, school employee, and alumni priorities with lottery-based tiebreakers. Applicants may list as many schools as they want on their choice application. This means that the mechanism is strategy-proof. The approach Chile takes to centralized assignment is similar to that used in major U.S. districts such as New York and Boston (Abdulkadiroglu and Sönmez (2003b); Abdulkadiroglu et al. (2005)).

The centralized school choice platform opens in August each year. Applicants have access to the platform for roughly one month, during which time they may view, submit, and edit their applications. The application deadline falls in early September, and students are notified of their placements in late October. Applicants who receive a placement can turn down that placement if they want. Applicants who reject their placement, who are not placed, or who did not participate

in the main round can join a secondary application process in late November that lasts one week.

Between early January and the beginning of the school year in March, students who still do not have a placement and placed students who decide to decline their placements may enroll in undersubscribed schools, outside of the centralized system. See Online Appendix B for further discussion of school choice institutions and enrollment outcomes for unplaced students.

Parents within the School Admission System (SAE) are able to choose which schools to apply to, and which schools to accept offers from. The availability of schools varies by location, but many students face a rich subset of options of schools that might admit them.

In the 2021 admission process, there were 7,979 schools with available slots and 461,223 students applying to grades ranging from Pre-K to 12th grade. Survey responses indicate that parents generally maintain their applications within 2 km of their home. The median distance for all applications is 1.21 km and the average distance for an applicants' full application is 4.53 km, where over 75% apply on average to schools approximately 2 km from their home.

On average, there are approximately 15 schools available in the SAE system in a 2km radius of a student's home. When considering the Metropolitan region, this number increases to 21. Of this possible choice set, there is on average 8 schools that are considered High or Medium Quality and 11 free for each student. This number increases to 11 and 15 respectively when considering the Metropolitan region.

Despite having many options, households do not always apply to seemingly-relevant options in their choice set, and this seems to come at a cost of non-placement risk. The application process does not have a limit on how many schools parents can apply to, and parents can apply to schools in any order and change their preferences at any time before the application process closes. Nonetheless, during the main 2021 application process, parents on average applied to 2.98 different programs and 2.92 different schools, far fewer than the total available to them.[7] Applying to few schools can create non-placement risk (see Arteaga et al. (2022)), as some schools may have few available slots.

2.2 Study Partners

Heading into the pandemic, school applications to SAE was a major concern for education policy makers in Chile, in particular because of the high stakes for a relatively new centralized platform and rising concerns about the gap in the quality of the schools that high and low income families chose. Preliminary descriptive and qualitative evidence suggested that some families had inaccurate, overly optimistic beliefs about the distribution of school characteristics. Based on this

[7] The 25th-percentile application ranks 2 schools while the 75th percentile contains 4 schools. The maximum number of schools that a parent applied to is 94 schools.

evidence, we helped Mineduc design a set of information interventions designed to quantify to what extent households underinvest in search due to preferences, biased beliefs and search costs. These interventions provided families with information about the distribution of schools and incentives to search for better options. The program was called "Informarse Bien".

The key feature enabling these interventions is the ability to interact with application data, explorer usage and applicants in real time over the course of the application process, to track search behavior and beliefs. The technical and logistical demands of implementing a platform at scale led one member of the research team (Neilson) to found a nongovernmental organization (NGO), ConsiliumBots, specializing in school choice services. The NGO partnered with Mineduc to run the interventions from 2018 on. See our Disclosure Statement in the Online Appendix for details on the relationship between the research team and the implementing partner NGO.

2.3 Subject Recruitment

Between May 25 and July 2 a recruitment email was for the early information programs was distrubuted through the Ministry of Education to potentially eligible parents.[8] The eligibility condition to be part of the study was to be a parent or caregiver for children enrolled in pre-kindergarten and kindergarten who were first time applicants to an educational institution through SAE in 2021.

[8] See 8.3 for more details

TABLE 1. Descriptive Statistics for Chilean Choice Applicants and the Experiment Sample

	All	Entry Grade	Economically Vulnerable	Not Economically Vulnerable	In Sample
	(1)	(2)	(3)	(4)	(5)
N	461,223	207,578	105,198	102,379	1,633
%	1.00	0.45	0.23	0.22	0.00
Panel A: Demographics					
Economically vulnerable	0.54	0.51	1.00	0.00	0.36
Woman	0.49	0.49	0.49	0.49	0.52
Entry grade	0.45	1.00	1.00	1.00	1.00
Panel B: Application Behavior					
Length initial attempt	2.94	2.93	2.68	3.19	3.84
Length final attempt	2.98	2.97	2.71	3.24	3.94
Total attempts	1.05	1.05	1.04	1.06	1.09
Panel C: Placement					
Placed in pref.	0.80	0.88	0.91	0.85	0.93
Placed 1st pref.	0.55	0.64	0.71	0.58	0.58
Partic. in 2nd round	0.08	0.07	0.06	0.07	0.07
Panel D: Enrolled School					
Enrolled at some school	0.97	0.97	0.98	0.96	0.97
Enrolled at placed	0.70	0.71	0.74	0.68	0.71
Free Tuition	0.78	0.75	0.83	0.66	0.72
Insufficient Quality	0.04	0.03	0.04	0.02	0.02
Mid-Low Quality	0.25	0.19	0.21	0.17	0.14
Mid Quality	0.58	0.60	0.59	0.61	0.54
High Quality	0.13	0.18	0.16	0.19	0.30
Highlight Worthy	0.43	0.45	0.52	0.39	0.48

Notes N: 461,223 applicants in the principal centralized placement round. All statistics are means in the population defined by the column header. Column (3) and (4) refer to economically/ not economically vulnerable population of entry grade population. Selected row variable definitions are as follows. *Entry grade* is an indicator if the student is applying to prekindergarten, kindergarten, first grade. *Economically vulnerable* is an SES measure computed by Mineduc. *Woman* is an indicator if the student is female. *Length of initial/final attempt* is the number of programs on an applicant's first and final choice application. *Total attempts* is the number of times an applicant submitted an application to the centralized system. *Placed in pref/1st* are indicators for any placement or for the school ranked 1st. *2nd round* variables describe participation and placement outcomes in the second centralized placement round

Our sample consists of 2,279 parents who comply with all of the following conditions: (i) they have no children already enrolled in a school, (ii) they have a child applying in an entry grade to SAE in 2021, and (iii) they have at least 5 schools around them, where one of these schools is ranked high or medium quality and free. We excluded 478 parents from the sample who left the platform before the point in which intervention differ by treatment group. This leaves us with a sample of 1,801 parents

Within the applications, we find that 74.09% of schools applied are free[9] and only 2,62% of schools are over 100.000 CLP. With regards to parents, 89.04% of parents apply to at least one free

[9]This considers possible voucher status (SEP).

school in their application set. Similarly, the analysis for quality of schools applied to shows that 80.1% of schools applied are high or medium quality and only 2.57% schools applied are insufficient quality. With regards to parents, 93.16% of parents apply to at least one high or medium quality school in their application set. More in detail, 56.15% of schools applied to are free and high or medium quality. With regards to parents 80.37% of parents apply to at least one free and high or medium quality school in their application set.

Table 1 shows the comparison between the full pool of applicants (column 1), applicants applying for entry grades only (column 2), economically vulnerable and nont vulnerable entry grade applicants (columns 3 and 4), and the experiment sample (column 5). Panel A shows that in the experiment sample only 36% of the students are economically vulnerable, as opposed to 51% of entry level applicants. Panel B shows that families in the experiment sample have longer rank ordered lists (3.84 schools on average, compared to 2.93 in column 2) and submitt more applications to the centralized system. Panel C shows that hey are also more likelly to be placed in a school, but conditional on being placed, they are less likelly to be placed in their first preference, suggesting that they tengo to rank first schools that are oversubscribed. From Panel D shows that they are more likelly to be enrolled in a high quality school and less likelly to be enrolled in an insufficient or low quality school.

Table 2 describes the final sample analyzed throughout the intervention. We show baseline sample characteristics and treatment-control group balance by regressing an indicator for treatment assignment on baseline covariates. Here, we find that there is no significant difference between treatment groups, despite some attrition in take-up before the differentiated interventions take place.

TABLE 2. Balance Checks

	Control		Treatment 1		Treatment 2		
	Mean (1)	St. Dev. (2)	Coeff. (3)	St. Err. (4)	Coeff. (5)	St. Err. (6)	N (7)
Panel A: Choice Environment							
Number of schools (2km - within intervention)	18.778	[8.903]	0.154	(0.427)	0.262	(0.433)	1,801
Number of schools (2km - total school supply)	42.904	[17.580]	0.124	(0.895)	0.274	(0.894)	1,801
Number of highlighted schools (2km - total school supply)	9.414	[4.785]	0.096	(0.264)	0.254	(0.260)	1,801
Panel B: Parent/Child Characteristics							
Child is female	0.483	[0.500]	0.043	(0.029)	0.040	(0.029)	1,801
Child's birthyear	2017.076	[0.541]	0.008	(0.031)	0.029	(0.031)	1,801
Mother completed sec. school	0.942	[0.234]	0.005	(0.012)	-0.006	(0.013)	1,798
Mother completed college	0.249	[0.433]	0.000	(0.017)	-0.009	(0.017)	1,798
Number of younger siblings	1.141	[0.401]	0.023	(0.023)	0.008	(0.023)	1,801
Child has disability (Belief)	0.075	[0.263]	-0.007	(0.016)	-0.012	(0.016)	1,630
Parent works in a school	0.069	[0.255]	-0.010	(0.014)	-0.012	(0.014)	1,768
SEP Household	0.373	[0.484]	-0.015	(0.014)	-0.018	(0.014)	1,781
Panel C: Initial Knowlege and Beliefs							
Expected satisfaction with process	5.144	[1.392]	0.053	(0.083)	0.017	(0.084)	1,679
Listed any school as pref 1	0.911	[0.286]	-0.008	(0.017)	-0.008	(0.017)	1,801
Listed highlighted school as pref 1	0.490	[0.500]	0.044	(0.032)	0.026	(0.032)	1,402
Perceived admission chance of 1st preference	68.089	[25.839]	1.400	(1.538)	1.358	(1.533)	1,679
N Schools known by name	3.608	[2.745]	-0.027	(0.158)	0.068	(0.154)	1,801
N Schools known well	1.925	[2.126]	-0.038	(0.123)	-0.072	(0.120)	1,801
Perceived Number of Schools (2km) Baseline	6.803	[8.769]	-0.441	(0.420)	-0.882	(0.391)	1,801
Perceived Number of Highlighted Schools (2km) Baseline	2.086	[2.610]	0.124	(0.150)	-0.008	(0.144)	1,801
Parent believed to be SEP eligible	0.139	[0.346]	-0.021	(0.019)	-0.003	(0.020)	1,801
SEP did not know about SEP status	0.702	[0.458]	-0.009	(0.027)	0.013	(0.026)	1,801
Likelihood, add. info on known schools	66.186	[30.149]	-2.104	(1.809)	-1.461	(1.787)	1,679
Likelihood, add. info on unknown schools	55.858	[33.018]	1.071	(1.917)	0.454	(1.919)	1,679
Would add school if knew all schols	0.821	[0.384]	0.005	(0.023)	0.013	(0.022)	1,679
Likelihood, add school to pref 1 if knew all schools	56.129	[29.431]	0.633	(1.808)	-0.084	(1.746)	1,679
Likelihood, add school blw last pref if knew all schools	66.264	[27.196]	1.109	(1.624)	-0.297	(1.612)	1,679

Panel D: Treatment Summary	Control	Treatment 1	Treatment 2
Observations	604	587	610
Whatsapp Reminder + SEP Status + Explorer	X	X	X
School Distribution		X	X
Highlight School			X

Notes This table presents means and standard deviations for variables from baseline survey in control participants. Differences between treatment and control are estimated from a regression of the variable on a treatment indicator and P-values are calculated from heteroskedasticity-robust standard errors. The sample consists of 1,801 parents who opened the intervention, had no children already enrolled in a school, had children in the age of applying to schools in 2021, and had at least 5 schools in a 2km radius from their house, with one being high quality and free. *Panel A* uses the number of schools eligible in SAE and total universe of schools including private and Pre-K's. Highlight school refers to schools High or Medium Quality and Free. *Panel B* uses sample characteristics collected from Baseline Survey and *Panel C* considers responses to questions on knowledge and beliefs about the school system. *Panel D* reports the number of observations in each treatment group and the interventions assigned to each.

2.4 Intervention Design

The intervention conducted by Mineduc varied the availability of information about the schools in the neighborhood in the explorer platform. This intervention calculated the number of schools

and their characteristics within 2km of a household's home and displayed this information in the explorer platform. Families assigned to Control group received access directly to the search platform without information about the distribution of schools or their characteristics. Families assigned to Treatment 1 received access to the search platform and a slide deck with information about the distribution of schools in their neighborhood (see Figure 1), with the goal of providing accurate information about the distribution from which households might be sampling when they investigate schools they do not yet know. Families assigned to Treatment 2 received everything provided to Treatment 1, as well as a set of "highlights" indicating those schools in the neighborhood that are free and of medium or high quality (see Figure 2).

FIGURE 1. Treatment 1



FIGURE 2. Treatment 2

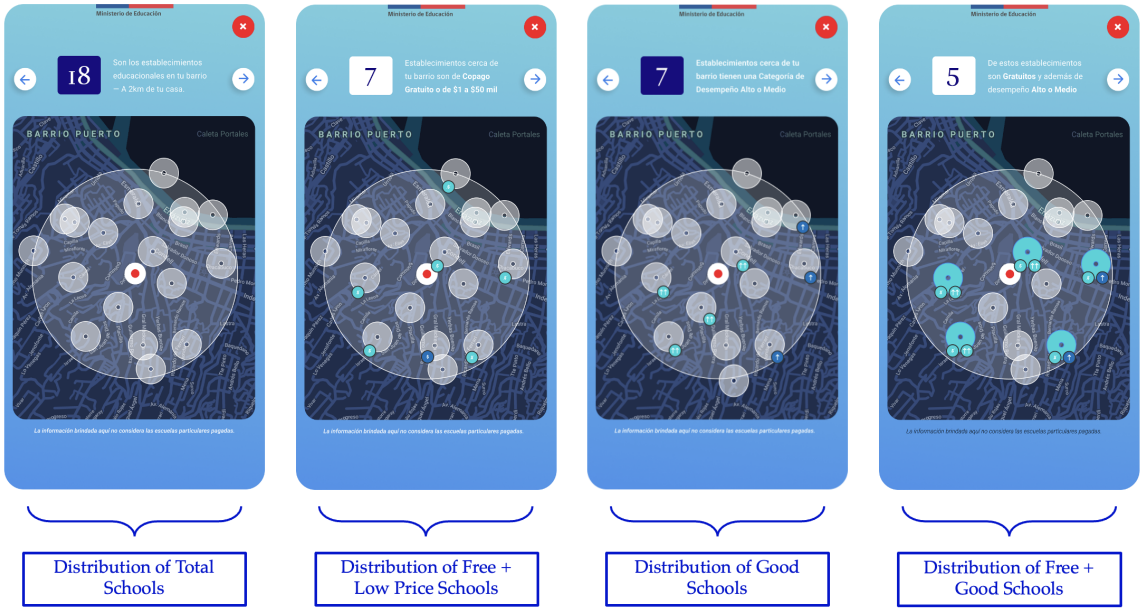
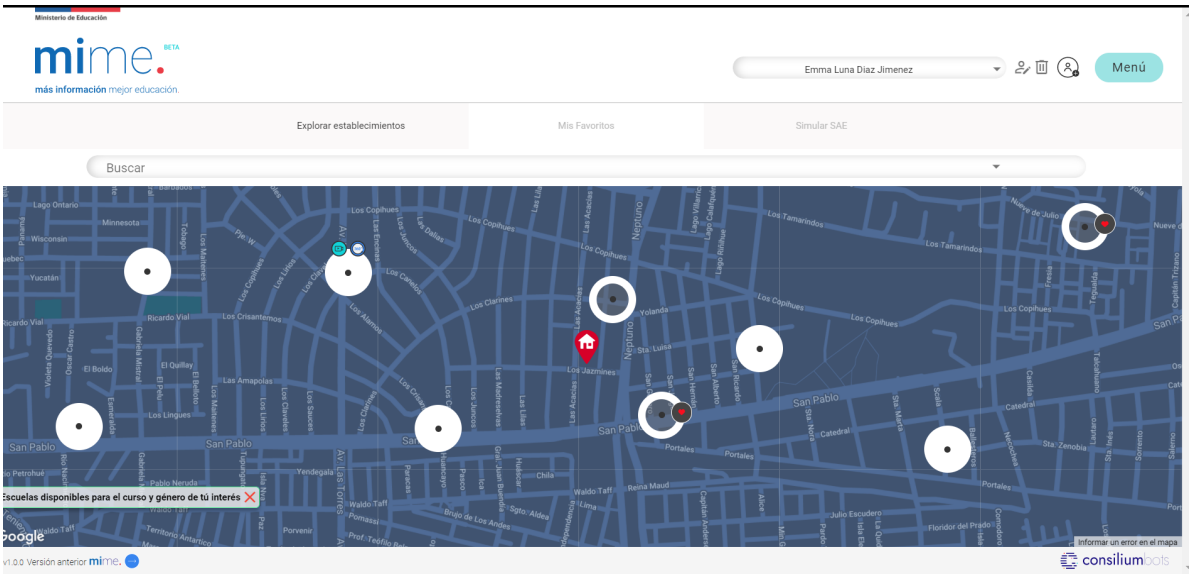


FIGURE 3. School Explorer



3 Descriptives

3.1 Data

Registration Form: The initial registration form to recruit participants for the study collected information on demographics, family structure, location, and basic beliefs about school characteristics.

Baseline Survey: This survey was sent to eligible parents and collected a more detailed set of beliefs and knowledge level about schools in the neighborhood. It also included questions about the rank ordered list of schools that parents were planning to apply to and search behavior.

Explorer Activity: From the explorer platform we are able to track the activity of each parent. This includes each unique clicks on the map, clicks on the schools and the detailed view of each click.

Midline Survey: This survey was sent to parents that opened the intervention. It collected information on beliefs and knowledge level of schools, explorer activity, and search behavior. It also included questions about the actual school application and beliefs on admission chances.

Application Data: This data was provided by Mineduc and includes each application list submitted by parents during the application process. It includes the rank ordered list of schools and the application outcome.

Administrative and Enrollment: This data was provided by Mineduc and consists of administrative data on applicants and enrollment for the 2021 school year.

Satisfaction Survey: This survey was sent to all SAE participants and collected information on the application process and knowledge of schools in the neighborhood. It included questions about application knowledge, beliefs on admission chances, and important factors in the search process.

Risk Measure: We use the 2020 application process to calculate the probability of being assigned to each school if placed in first preference for each applicant profile. We use this as a proxy for the risk of non-placement in the 2021 application process.

3.2 Survey Design

To learn more about how families engaged with the choice process, the NGO helped Mineduc conduct several surveys of choice participants in 2021 at different stages of the application process. The survey asked questions about several school choice topics. It included modules about preferences, beliefs, and search designed to provide context for the interventions we study here. The survey innovates over past surveys of choice participants (De Haan et al., 2023; Kapor et al., 2020; Wang and Zhou, 2020) by asking about search in addition to preferences and beliefs.

To start with, the NGO coordinated with Mineduc to implement a website for parents to sign up for future interventions. This website was shared with principals of kindergartens throughout Chile via email and invited parents to fill out a questionnaire about their beliefs and contact information. Following this, Mineduc sent a Baseline survey using a Whatsapp message sent to parents eligible for the study (collected from the previous questionnaire) one month before the application process started. The survey was sent to 9,062 parents and asked about parents' preferences, awareness, beliefs, and search behavior. Of these, 2,279 (25%) completed the survey, met the experiment conditions, and were randomized into the three treatment arms.

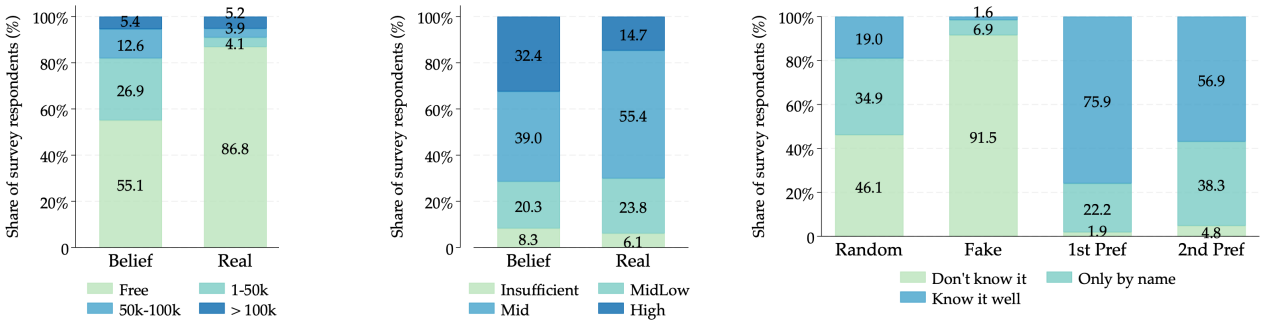
After the treatments were sent and between the start of the application process and the admission results, the NGO sent a second survey implemented through a call center to 1,810 parents that opened the intervention to measure it's effect on beliefs and search behavior. Of these, 966 (60%) completed the survey.

3.3 Descriptive and Experimental Analysis

3.3.1 Survey Findings

The main focus of our analysis is whether parents are aware of the schools in their neighborhood and how they search for schools. Our first result here is that households don't know all schools. We asked each respondent in the Baseline survey to report how well they knew 8 randomly selected nearby schools and 2 "fake" schools - that is, a school that did not exist.[10] Additionally, we asked during the application period the knowledge of schools on the application list. Figure 4c displays applicants' responses to these questions. From responses in baseline survey we can see that 19% of parents know well the schools asked about, 34.86% know them by name, and 46.13% don't know them at all. Consistent with the idea that applicants learn about schools before applying to them, respondents claim to know the schools on their applications better than they know randomly chosen nearby schools. Over 75% of families respond they know well their first preference, while only 56.87% of families respond they know well their second preference.

[10]Schools in this question were selected from the alternatives within 2 km from the residental location of the student.

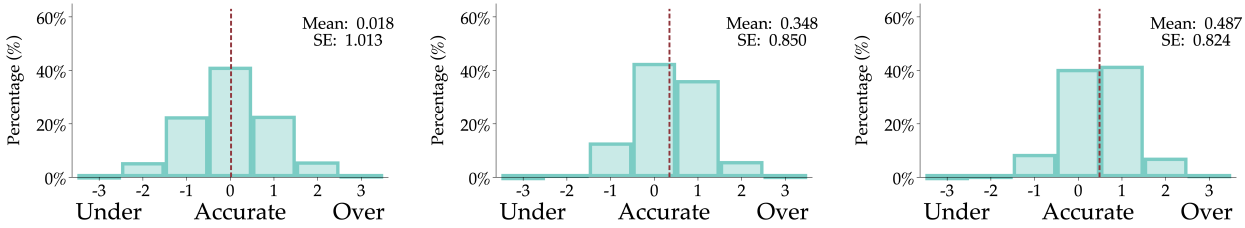


(A) (B) (C)

FIGURE 4. Knowledge and Distribution of X's

Notes: Panel (A) shows share of survey respondents stating the perceived distribution of schools in their neighborhood with monthly fee: Free, 1-50k, 50k-100k and greater than 100k versus the real distribution of schools. Panel (B) shows the share of survey respondents stating the perceived distribution of schools in their neighborhood with quality: High, Mid, Mid-Low and Insufficient versus the real distribution of schools. Panel (C): Stated knowledge of a random school asked in baseline, a fake school, and the first two schools on application list.

Our second result is that households hold inaccurate beliefs about the general distribution of school characteristics such as quality, price, and admission probability within schools in their neighborhood. We asked parents directly how many of the schools they believed were in a 2 km radius were High, Mid, Mid-Low and Insufficient performance.[11] We observe from Figure 4 that households overestimate the share of high quality schools and underestimate the share of mid quality schools. Similarly, we note from Figure 4 that households underestimate the share of free schools. This is consistent with the idea that households are not aware of the schools nearby, and therefore they do not know their characteristics.

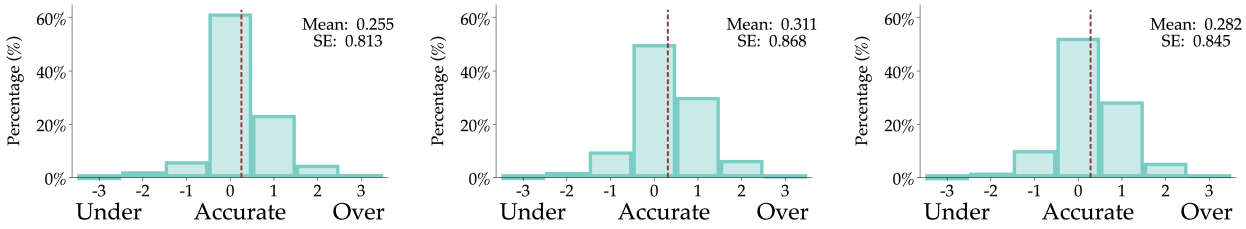


(A) Random School Known (B) Random School in App (C) First Preference

FIGURE 5. Error in Quality

Notes: Panel (A) shows the bias on perceived quality of a known random school asked in baseline. Panel (B) shows the bias on perceived quality of a random school in the application list, excluding the first ranked school. Panel (C) shows the bias on perceived quality of the first preference school at baseline. All biases are measured as perceived quality minus real quality. Positive values indicate that the parent responded a higher quality than real and negative values indicate that the parent responded a lower quality than real.

[11] We used the same performance classification of the Agencia de la Calidad de la Educación, an institution that classifies schools in four tiers using standarized test scores, after taking into account SES of the student body.



(A) Random School Known (B) Random School in App (C) First Preference

FIGURE 6. Error in Price

Notes: Panel (A) shows the bias on perceived price of a known random school asked in baseline. Panel (B) shows the bias on perceived price of a random school in the application list, excluding the first ranked school. Panel (C) shows the bias on perceived price of the first preference school at baseline. All biases are measured as perceived price minus real price. Positive values indicate that the parent responded a higher price than real and negative values indicate that the parent responded a lower price than real.

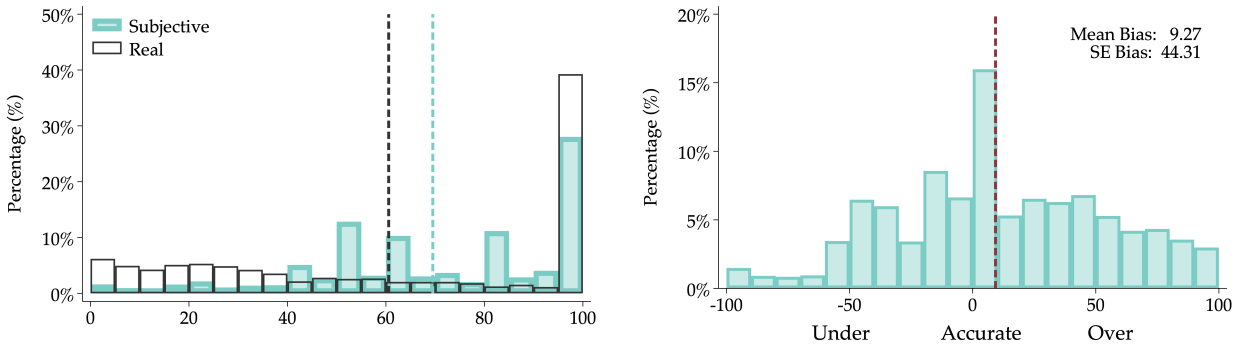
Remarkably, households also hold inaccurate beliefs about known schools. We asked parents to report the quality and price of a school they know at least by name, their first preference school and a random school in their application list, not including the first ranked. Figures 5 and 6 show the distribution of responses for quality and price respectively. We observe that in general, households overestimate the quality of known schools, yet this bias is more significant for schools they're interested in applying and especially for the most preferred school. In a similar manner, households overestimate the price they must pay for known schools, yet the bias is stronger for the perceived quality.

TABLE 3. Belief vs Real on ranking

		Partial Ranking (1)		Regular Round Ranking (2)	
Distance		-0.000	(0.003)	-0.048**	(0.020)
Perceived Price Category	1	-0.306**	(0.119)	0.280	(0.188)
	3	-0.226*	(0.133)	-0.152	(0.249)
	4	-1.269***	(0.236)	-0.329	(0.502)
Real Price Category	1	0.052	(0.118)	-0.078	(0.200)
	3	0.166	(0.134)	0.284	(0.233)
	4	0.153	(0.216)	0.295	(0.459)
Perceived Performance	1	-1.683***	(0.623)	-1.593	(1.091)
	3	1.894***	(0.176)	0.232	(0.241)
	4	3.712***	(0.202)	1.023***	(0.276)
Real Performance	1	-0.569**	(0.252)	0.020	(0.459)
	3	0.099	(0.113)	-0.113	(0.185)
	4	0.226*	(0.127)	0.028	(0.216)
Public School		-0.344***	(0.112)	-0.050	(0.172)
Observations		3568		1199	

Notes. This table presents a rank-ordered logit choice model. Column (1) refers to the partial ranking we elicited at baseline with perceived price and quality from responses to the baseline survey. Column (2) refers to the ranking from application data from SAE Regular Round, with perceived price and quality from responses to the midline survey.

Households rank schools on the basis of quality, but respond to subjective beliefs not true values. Table 3 presents the results of a rank-ordered logit choice model where the dependent variable is the rank of the school in the application, among schools for which we have belief elicitations. Column (1) illustrates how the perceived price and quality predict survey-reported ranking at baseline (i.e. before the application process), while column (2) considers the true ranking submitted to the mechanism. Conditional on surveyed beliefs, for which we find large effects, the "true" quality measures are insignificant in predicting final rankings. As households tend to overestimate quality, this result suggests that households may overestimate the value of "known" schools, reducing the perceived returns to search.



(A) Distribution of Placement Chances

(B) Bias of Placement Chances

FIGURE 7. Error in Placement Chances

Notes: Panel (a) shows the perceived and real distribution of placement chances for first preference at baseline. Placement chances are calculated according to the most common program the school has if they have more than one program in the application process. Panel (b) shows the bias on perceived placement chances of the first preference school at baseline, measured as perceived placement chances minus real placement chances. Positive values indicate that the parent responded a higher placement chance than real and negative values indicate that the parent responded a lower placement chance than real.

We also observe that households mispredict admission chances. We asked parents to report the probability of admission to their first preference school and compare respondents' reported beliefs about placement chances to our calculations of objective placement chances. We can see from Figure 7 that respondents are overconfident about their placement chances. The average perceived placement chance is 69.52% with a standard deviation of 26.53 while the real placement chance is 60.56% with a standard deviation of 37.29.

3.3.2 Search Behavior

We build on the survey findings and use the explorer platform to analyze search behavior. The focus of our analysis in this part is to describe how people search and what predictors are associated with search patterns. Our first result here is that observed effort depends on beliefs about admission chances and search probabilities. First, parents were asked at baseline to report the probability of admission to their first preference school. From Table 4 we observe that parents who believe they have a higher probability of admission to their first preference school are less likely to interact with the explorer and search for more schools. In particular, they are less likely to have any interaction with a school or with a highlight - worthy school in the explorer. Similarly, they are less likely to have clicked on any school pin or on a highlight - worthy school pin.

TABLE 4. Observed effort on beliefs about admission chances

	Explorer Interactions			Clicked School Pin	
		Unique Interactions			
	Total	Any	Highlighted	Any	Highlighted
	(1)	(2)	(3)	(4)	(5)
Belief Admission Chance 1st Pref	-27.507***	-5.058***	-1.808***	-4.986***	-1.742***
	(9.942)	(1.588)	(0.584)	(1.581)	(0.579)
Real Admission Chance 1st Pref	-18.833***	-2.922**	-0.956**	-2.727**	-0.903**
	(7.002)	(1.253)	(0.457)	(1.250)	(0.453)
Control Mean	56.899	10.790	3.700	9.351	3.501
Observations	1302	1302	1302	1302	1302

Notes. This table presents the analysis of beliefs about admission chances on search effort. Column (1) is the number of interactions. Column(2) is the unique number of schools interacted with. Column (3) is the number of highlighted schools interacted with. Column (4) is the unique number of schools - sede clicked on. Column (5) is the number of highlighted schools. We have N = 1,801 participants where 1,302 have data on the real probability of being admitted to 1st preference. The remaining have missing real probability either because we were not able to track the school or the program they're applying to does not exist. Restrictions: Sample 1 and clicks and start primer

We also included questions about search probabilities at baseline. Parents were asked about the probability of adding a school to their application if they took the time to learn of all schools, the probability of searching more for known and unknown schools in their neighborhood, the probability of adding a school to their application if they found a school with the same/worse characteristics as their first preference school, and the probability of adding a known school to their application as first preference or below their last preference. Survey responses suggest that parents who report that they would add a school to their application if they took time to learn of all their options are more likely to engage with the search platform. Similarly, parents who report a higher probability of searching for more information on unknown schools are more likely to to interact with the explorer and search for schools.

TABLE 5. Observed effort on beliefs about search probabilities

	Explorer Interactions			Clicked School Pin		Clicked Profile	
		Unique Interactions					
	Total	Any	Highlighted	Any	Highlighted	Any	Highlighted
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Add School if Know All	15.937***	2.880***	0.958***	2.925***	0.959***	0.580**	0.251**
	(4.829)	(0.828)	(0.327)	(0.820)	(0.324)	(0.237)	(0.126)
Prob Search More							
Unknown Sch.	19.167**	3.492***	1.311***	3.400***	1.246**	1.531***	0.843***
	(8.711)	(1.274)	(0.502)	(1.262)	(0.497)	(0.456)	(0.226)
Known Sch.	12.216	1.298	0.246	1.302	0.292	0.281	0.076
	(9.434)	(1.445)	(0.563)	(1.444)	(0.561)	(0.470)	(0.232)
Prob Add School							
Same X's	0.604	0.544	0.572	0.541	0.514	0.219	0.203
	(9.121)	(1.696)	(0.627)	(1.684)	(0.620)	(0.478)	(0.230)
Worse X's	-2.039	-1.968	-0.674	-1.894	-0.662	0.028	-0.109
	(9.350)	(1.517)	(0.566)	(1.513)	(0.560)	(0.508)	(0.245)
Prob Add Sch. if Knew							
To Pref 1	-35.452**	-6.991**	-2.041**	-7.114**	-2.098**	-0.921	-0.028
	(17.685)	(2.822)	(1.022)	(2.822)	(1.022)	(0.811)	(0.379)
blw Last Pref	35.930**	7.013**	1.476	7.197**	1.619	0.855	-0.040
	(18.166)	(2.873)	(1.071)	(2.875)	(1.072)	(0.783)	(0.390)
Control Mean	59.813	11.182	3.972	9.765	3.777	2.773	1.322
Observations	1679	1679	1679	1679	1679	1679	1679

Notes. This table presents the regression of perceived search effort measured from Baseline on real search effort from the explorer platform controlling for number of total and highlighted schools within 2km. The dependent variable includes different measures of explorer usage and the regressors are reported responses on beliefs about search. Column (1) is the number of interactions. Column(2) is the unique number of schools interacted with. Column (3) is the number of highlighted schools interacted with. Column (4) is the unique number of schools - sede clicked on. Column (5) is the number of highlighted schools. Column (6) is the unique number of schools where the participant opened the profile and Column (7) is the number of highlighted schools. We have N = 1679 participants in the explorer data. Restrictions: Sample 1 and clicks and start primer

Our second result from the explorer usage data is that platform behavior affects school knowledge. Table 6 shows that when controlling for reported knowledge at Baseline, parents who clicked on the school profile (Double Click) are more likely to respond in Midline Survey that they know the school. This is consistent with the theory that parents learn from search about schools and update their beliefs. Consistently, when we condition on having clicked on the school pin, we find that seeing the school profile (a more detailed view of the school) increases the probability of knowing the school at midline.

TABLE 6. Platform Behavior Affects School Knowledge

	Unconditional		Conditional on Pin Click	
	Knows At Least by Name	Knows Well	Knows At Least by Name	Knows Well
	(1)	(2)	(3)	(4)
Single Click	-0.056*	-0.063		
	(0.033)	(0.070)		
Double Click	0.099***	0.149***	0.112***	0.151***
	(0.030)	(0.040)	(0.032)	(0.042)
Baseline - Knows By Name	0.341***	0.311***	0.351***	0.325***
	(0.046)	(0.051)	(0.047)	(0.052)
Baseline - Knows Well	0.414***	0.615***	0.420***	0.626***
	(0.041)	(0.046)	(0.041)	(0.046)
Control Mean	0.821	0.512	0.804	0.488
Observations	496	496	449	449

Notes. This table presents the regression of search on knowledge level at midline for the control group. Column (1) and (2) is unconditional on pin click and column (3) and (4) are conditional on pin click. For Column (1) and (2), we have N = 316 control participants with response to knowledge level at Midline, Baseline and explorer usage, with 496 total participant-school observations. For Column (3) and (4) we have 496 participant-school observations conditional on having any interaction with the explorer. Controls include knowledge level at Baseline. Restrictions: Sample 1 and Start Primer and clicks

Further results indicate that platform behavior affects beliefs on the general distribution of schools. Table 7 reports the regressions of explorer usage on perceived number of schools. Column (1) indicates that the higher number of schools clicked in the explorer, the more schools perceived by parents at midline, controlling for the perceived number of schools reported at baseline and the actual number of schools within 2 km of the household. Column (2) is similar in structure but reports the coefficients for the number of highlighted schools. These results suggest that a higher number of highlight - worthy schools clicked in the explorer, the more highlight - worthy schools perceived by parents at midline. Column(3) reports the same results as Column (1) and (2) but for the share of highlight - worthy schools in the neighborhood. We also find the people are more likely to include in their application schools they've searched in the explorer.

TABLE 7. Platform Behavior Affects Beliefs

	Perceived N of Schools (Midline) (1)	Perceived N of Highlighted Schools (Midline) (2)	Perceived Share of Highlighted Schools (Midline) (3)
Perceived Schools (Baseline)	0.161	0.287***	0.090**
	(0.104)	(0.069)	(0.040)
Schools Clicked in Explorer	0.054*	0.064***	0.048
	(0.033)	(0.023)	(0.049)
Actual Schools	0.055**	0.038	0.310***
	(0.026)	(0.026)	(0.098)
Control Mean	6.777	2.032	0.332
	Treat 0	Treat 0	Treat 0
Observations	318	311	300

Notes. This table presents the regression of explorer usage on perceived number of schools in Midline Survey. Column (1) refers to the perceived number of schools within 2km of the participant, with controls for perceived number of schools at baseline, N schools clicked in explorer, and actual N of schools within 2km. Column (2) refers to the perceived number of highlighted schools within 2km of the participant, with controls for perceived number of highlighted schools at baseline, N schools clicked in explorer, and actual N of highlighted schools within 2km. Column (3) refers to the perceived share of highlighted schools of the total schools, with controls for perceived share of highlighted schools at baseline, share of highlighted schools clicked on the explorer and actual share of highlighted schools within 2km. We have N = 326 control participants with consent to the Midline Survey, where 318 and 311 have answers for perceived N schools and perceived N highlighted schools respectively. Restrictions: Sample 1 and Start Primer and clicks

3.3.3 Information Treatments

We use the treatment randomization to study how search behavior and application outcomes vary in the intensive margin when households are provided with information about the distribution of schools in their neighborhood.

Our first result is that treatment and search history affects perceived price and quality. Table 8 presents the results of explorer usage and treatment on perceived price, quality, and admission chances measured at midline. Consistent with our hypothesis that parents update their beliefs after learning about schools, we detect a positive significant effect of explorer interaction on correct measure of school price and quality. We find that having interacted with the school updates households' beliefs in the correct direction for price and quality, but no significant effect for admission chances.

TABLE 8. Treatment and history affect perceived price and quality

	Price			Quality			Pr Admission Chance		
	Absolute Value	Over	Correct	Absolute Value	Over	Correct	Absolute Value	Over	Correct
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Any Click	-0.349***	-0.073**	0.111***	-0.275***	-0.136***	0.166***	0.005	0.101***	0.001
	(0.094)	(0.033)	(0.035)	(0.057)	(0.034)	(0.036)	(0.020)	(0.039)	(0.031)
Control Mean	0.802	0.307	0.596	0.581	0.336	0.552	0.352	0.454	0.185
	Treat 0	Treat 0	Treat 0	Treat 0	Treat 0	Treat 0	Treat 0	Treat 0	Treat 0
Observations	762	762	762	763	763	763	632	632	632

Notes. This table presents the results of pdslasso regressions of search history on perception of price (Columns 1-3), quality (Columns 4-6), and probability of being admitted (Columns 7-9) of schools asked at midline. Columns (1), (4) and (7) represent the absolute difference between the believed price/quality/prob admit and the real. Columns (2), (5) and (8) are indicators if the parent stated higher price/quality/prob admit than real. Columns (3), (6) and (9) are indicators if the parent answered correctly, considering correct admission chance within the interval of +- 10 from the real. The regressors *Any Click* considers any click pin or profile pin. We have N = 839 participants with responses to price and quality, with 2252 participant - school observations. Restrictions: Sample 1 and clicks and start primer.

Furthermore, we find that treatment affects beliefs and search, where households assigned to treatment perceive a higher number of highlighted schools in their neighborhood (See Table 9). Parents assigned to treatment 2 search less in comparison to control and treatment 1, consistent with the notion that if we tell parents the exact distribution and where these schools are placed, they search less than if we just give them the distribution of school characteristics in their neighborhood.

Table 13 presents the effects of the RCT on various application outcomes at the school level. Panel A runs a pdslasso regression of treatment on school characteristics, clustering at the parent level. We test for highlight - worthy schools, free, high quality, value added[12], distance and an indicator variable if the parent responded they knew the school from baseline. We find that both treatments decrease the likelihood that the schools in their application are known at baseline. Panel B follows the same structure as Panel A but only considering the first ranked school in the application list. We find less effects of treatment which is indicative that households maintain a firm preference and perception for their first ranked school. Panel C runs the same regression but with the second choice school in the application list. We find that there's significant treatment effects on school characteristics, in particular for treatment 2. These parents are more likely to have a higlight - worthy, high quality, higher value added new school as their second preference. At the same time, parents assigned to treatment 1 are more likely to have a new closer school as their second preference. Panel D and Panel E run the same specification as the others but for assigned school and enrolled school. We find no significant treatment effects on the majority of these outcomes, yet we find that parents are more likely to enroll in schools closer to their home.

[12]We obtain value added data from Neilson (2021).

TABLE 9. Treatment affects beliefs

	Perceived Number of Schools		Number of Schools Know Well		Recommends Search
	All	Highlight	Direct	Detailed	
	(1)	(2)	(3)	(4)	(5)
Treatment 1	0.798	0.389**	-0.343*	-0.088	-0.006
	(0.503)	(0.191)	(0.205)	(0.132)	(0.037)
Treatment 2	0.377	0.421**	-0.438**	-0.121	-0.052
	(0.496)	(0.195)	(0.219)	(0.134)	(0.037)
pval: T1 = T2	0.419	0.881	0.616	0.813	0.214
Control Mean	6.868	2.048	2.132	1.859	0.330
Observations	934	915	594	819	929

Notes. This table presents the responses from Midline Survey following the intervention. The total number of parents with consent is 958. Column (1) refers to the question: *How many primary schools do you believe exist 2 km from your house?* where the total number of responses is 934. Column (2) refers to the question: *How many of these schools do you believe are High or Medium Quality and Free?* with a total number of responses of 915. Column (3) refers to the question: *Of the N schools 2km from your house that you know by name, how many do you know well?* with 594 responses since it was a question added later. Column (4) is a measure constructed from eliciting knowledge for specific schools and Column(5) refers to the question: *Would recommend neighbor to keep searching* with 929 total responses. Restrictions: Sample 1 and Start Primer and clicks

Table 14 presents the effects of the RCT on application measures at the household level and expected school level. Similar to our previous analysis, Panel A runs a pdslasso regression of treatment on application and expected characteristics. Column (1) - (2) uses as outcomes an indicator variable if the parent submitted an application and the application length. Columns (3) to (6) present the expected school characteristics when integrating over the admission chances. We find that treatment 2 parents are less likely to report knowing the school at baseline on expected. Panel B runs the same regression of treatment on several indicators of school assignment. We find that households assigned to treatment 1 are less likely to end up not assigned in the school process. Panel C follows a similar stucture but uses indicators of enrollment. We find that both treatments have a positive significant effect on being enrolled in the school they were assigned to after both school choice process. Furthermore, parents assigned to treatment 2 are more likely to be enrolled in a SAE school (refers to schools that belonged to the centralized school choice platform), more likely to be enrolled in a school that they listed in their application, and more likely to be enrolled in their first or second preference school.

4 Model

We now construct a model of households' preferences, beliefs, search efforts, and application decisions that is consistent with our descriptive evidence. Each household is endowed with knowledge of subset of the schools near their house, but does not necessarily know all of the options

that may be relevant. In order to learn about additional schools, households must engage in costly search. In doing so, a household trades off the costs of further search effort against the expected benefits given their beliefs about the distribution of school characteristics in their neighborhood.

When a household decides to stop searching, it submits an application to the centralized school choice mechanism. In this application, it ranks all schools in its consideration set which it prefers to its outside option, in descending order of preferences, as is optimal under the Chilean school choice mechanism.

We allow the following forms of misperceptions. Households may be misinformed about admissions chances, about the observables (price, quality) of "known" schools, about the distribution of observables at not-yet-discovered schools, and about unobserve match quality and its distribution. Each of these channels may distort search and application decisions. A household may wrongly conclude that it is likely to get into known schools, that the known schools are of high quality, that schools it does not know are unaffordable, and that schools it has not yet discovered must have something wrong with them on average, even if they have high quality ratings.

We now describe the model in detail.

4.1 Preferences and Awareness

Let $i = 1, \ldots, I$ denote households, and $j = 1, \ldots, J$ denote schools. Household i's choice set, $J_i \subset \{1, \ldots, J\}$, consists of all schools within 5km of i's house. For each i and each $j \in J_i$, at time t household i has awareness $\pi_{ijt} \in \mathbb{R}$ of j, and expected payoff $\hat{u}_{ijt} \in \mathbb{R}$ from placing in school j, given by:

$$\pi_{ij}^t = (z_{ij}, w_{ijt})\alpha^{zw} + \eta_j + \nu_{ij}^0 + (\nu_{ij}^T - \nu_{ij}^0)1(t \geq T_i - 1) \tag{1}$$

$$\hat{u}_{ijt} = 1(\pi_{ij}^t \geq 1)u_{ij}^h(\hat{x}_{ij}^h) + 1(0 \leq \pi_{ij}^t < 1)u_{ij}^l(\hat{x}_{ij}^l), \tag{2}$$

where

$$u_{ij}^h(\hat{x}_{ij}^h) = z_{ij}\beta^z + \hat{x}_{ij}^{h,rc}\beta_i^x + \delta_j + \hat{x}_{ij}^h\gamma + \varepsilon_{ij} \tag{3}$$

$$u_{ij}^l(\hat{x}_{ij}^l) = z_{ij}\beta^z + \hat{x}_{ij}^{l,rc}\beta_i^x + \delta_j + \hat{x}_{ij}^l\gamma + \hat{E}(\varepsilon_{ij}|\tilde{\varepsilon}_{ij}) \tag{4}$$

$$\hat{x}_{ij}^l \sim \Gamma(\cdot|x_j), \;\; \hat{x}_{ij}^h = (x_j \text{ w.p. } p^h, \text{ otherwise } \hat{x}_{ij}^l). \tag{5}$$

If $\pi_{ijt} \geq 1$ we say that i knows j well at time t. If $0 < \pi_{ijt} < 1$ we say that i knows j somewhat. Otherwise, i is unaware of j. The random variables u_{ij}^h and u_{ij}^l are potential subjective expected utilities from placing in j. The level of awareness determines which potential outcome is relevant for the household. If j is known well, the expected payoff given the household's information is u_{ij}^h. If j is known by name, the expected payoff given the household's information is u_{ij}^l.

The variable $\hat{x}_{ij} \in \{1,2,3,4\}^2$ denotes the household's perception of the school's price and quality category. It is drawn from a function $\Gamma(\cdot|x_j)$ where x_j is the truth, in the event the household knows the school somewhat. Consistent with descriptive evidence, when the household knows the school "well" it learns the truth with probability p^h. With the complementary probability, it remains misinformed.

z, which enters both awareness and preferences, is distance from home to school. w, which are excluded from preferences, consist of indicators for treatment assignments, and indicators for having viewed any details of school j by time t in the explorer.

We assume that the submitted raking is a truthful ranking, in order of $\hat{u}_{ijT_i}$, of those schools with $\hat{u}_{ijT_i} > 0$ and $\pi_{ijT_i} > 0$.

We model schools' mean utility and "discoverability" as random effects: $(\delta_j, \eta_j)' \sim N\left((x_j\overline{\beta}, x_j\overline{\alpha})', \Sigma^{\delta\eta}\right)$. As δ and η may be correlated, our model allows more popular schools to be more famous. We assume $\beta_i^{rc} \sim MVN(0, \Sigma^{rc})$. Post-search (exogenous) off-platform learning is given by $\left(\nu_{ij}^0, \nu_{ij}^T\right)' \sim N\left((0,\overline{\eta})', \Sigma^{\pi}\right)$. Match value is given by $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$. If households have low information, they observe this shock with classical measurement error $\eta_{ij} \sim N(0, \tilde{\sigma}^2)$, and form a subjective expectation $\hat{E}(\varepsilon_{ij}|\tilde{\varepsilon}_{ij})$. In fact, we have

$$\begin{pmatrix} \epsilon_{ij} + \eta_{ij} \\ \eta_{ij} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1+\tilde{\sigma}^2 & 1 \\ 1 & 1 \end{pmatrix}\right).$$

However, households believe that

$$\begin{pmatrix} \epsilon_{ij} + \eta_{ij} \\ \eta_{ij} \end{pmatrix} \sim N\left(\begin{pmatrix} \overline{\beta}_i \\ \overline{\beta}_i \end{pmatrix}, \begin{pmatrix} \sigma_{i0}^2 + \sigma_i^2 & \sigma_{i0}^2 \\ \sigma_{i0}^2 & \sigma_{i0}^2 \end{pmatrix}\right).$$

That is, $\overline{\beta}_i$ represents the subjective mean unobserved utility component of unknown schools. A household may be pessimistic ($\overline{\beta}_i < 0$) or optimistic ($\overline{\beta}_i > 0$) about these unobservables. In addition, households may hold inaccurate subjective beliefs about the precision of their signal (σ_i^2) and/or the variance of the unobservable (σ_{i0}^2).

4.2 Timing and search

Time runs for T_i periods. At $t = 0$, households are endowed with preferences $(\beta_i^{rc}, \hat{\epsilon}, \epsilon)$, awareness ($\pi_i^0$), beliefs about unknown schools ($(p,q) \sim$ Categorical(λ_i^0)), beliefs about known schools ($\{\hat{x}_{ij}^0\}$), and admissions optimism (opt_i). Times $t = 1, \ldots, T_i - 1$ consist of a sequence of decisions to continue searching. At time T_i, additional information shocks $(\nu^T - \nu^0)$ realize exogenously, after which households submit rank-order lists and the midline survey takes place.

Search decisions consist of a sequence of "detail views", i.e. the decision to double-click a pin in the explorer and read about a school. After each view, households choose whether to continue or stop searching. Huuseholds compute the subjective EU of optimal portfolio given current info, $\hat{U}_{it}$, and the expected value of stopping after viewing one more school, $\hat{E}(\hat{U}_{i,t+1}|\mathcal{I}_{it})$. They stop if

$$c_{it} = c_{i0} + \kappa t + \epsilon^c_{it} > \hat{E}(\hat{U}_{i,t+1}|\mathcal{I}_{it}) - \hat{U}_{it},$$

where $\kappa > 0$, $c_{i0} \sim N(y_i\beta^c, \sigma_c^2)$, $\epsilon^c \sim$ logit. These costs depend flexibly on baseline characteristics, including baseline knowledge π^0.

We parameterize the first search decision separately, as the household has not yet seen any detail pages.

Conditional on searching, the identity of the next clicked school is given by$Pr(\text{view } j|\text{continue}) \propto \exp(x_{ij}^{\text{click}}\gamma^{\text{click}})$. That is, we model the decision to stop or continue. Conditional on continuing, certain schools (e.g. nearby schools) may be found more easily, but we do not model the choice of where to search. The variables x_{ij}^{click} consist of schools' price, quality, distance, and indicators for treatment=2 and being highlighted in that treatment.[13]

Beliefs about unknown schools at time t are given by $x \sim \text{Multinomial}(\lambda_{it})$, with $\lambda \sim Dir(\Lambda_{it})$. Dirichlet params Λ_i are functions of the truth, treatment status, and the number of "known" schools $\hat{N}$ with (perceived) chars $\hat{x}$. We have:

$$\Lambda_{it} = \Lambda_{i0} + \sum_s \left(c_{1s} * 1(\text{treat}_{it} = s) * N_{it}^{\text{true}} + c_{2s} * 1(\text{treat}_{it} = s) * \hat{N}_{it}^{\text{viewed}} \right)$$

Let the true chance of being accepted by j be be $1 - r_{ij}$. A household believes its chances are $1 - \hat{r}_i = \min(\max(1 - o_i(1 - r_{ij}), 0), 1)$, with $o_i \sim N(\mu_o, \sigma_o^2)$.

5 Estimation

We estimate the model in two steps. First, we estimate the distribution of $(u, \pi, \hat{x})$, and the parameters relevant for these objects, via a Gibbs sampler.

We use the submitted ROL, explorer "detail views", Baseline and midline survey responses to questions on "how well do you know" school j, baseline survey rank-order lists, and perceived x's at baseline and midline. We place random coefficients on $(1, \text{dist}, \hat{x})$. As a scale normalization, we fix the mean coefficient on distance to $= -1$. We allow measurement error on all survey variables. Reported awareness π^{survey}, and the payoffs that enter the baseline rank-order lists, have additive Normal measurement error, whose variance we estimate. Perceived $\hat{x}$ are also misreported (drawn

[13] Treatment 2 may make highlighted schools more salient, and other schools less salient.

from some distribution Γ^{surv}, independently of the beliefs that enter the household's decisions) with a probability that we estimate.

Identification comes from repeated measurements of rank-order lists, awareness π, and subjective beliefs $\hat{x}$, as well as variation in treatment assignments.

In a second step, we impose optimality of the search decision, and estimate search costs, admissions optimism, and the parameters of belief Λ_{it} via MLE, taking the means from the first stage as point estimates of the relevant parameters.

6 Results

Table 10 presents results for the estimated parameters for our model of households' preferences, beliefs, search efforts, and application decisions. Panel A shows the parameters that describe family's preferences and information. Preferences for preceived school attributes present the expected signs: parents value school quality and do not like price and schools that are far from their home. There is a relevant penalty in the expected payoffs of schools known only by name. There is significant measurement error in the Baseline Survey's Utility reports and in the awareness questions. When the household knows the school "well" it learns the truth with probability 22.5%. Panel B describes the variance-coariance matrix for the estimated random coefficients. For all attributes, there is heterogeneity in preferences that is not explained by observable characteristics. Unobserved price and distance sensitivity are positively correlated, which implies that a price sensitive families are also less willing to travel. The covariance between price and quality, and distance and quality is also positive, which implies that parents that value school quality tend to be less price and distance sensitive. Panel C shows the variance-covariance matrix for school discoverability and mean utility. The positive correlation parameter implies that more desirable schools are more likely to be known at baseline, and that the schools that one may find via search have lower mean utilities, on average, than those that are already known.

Figure 10 and Table 11 provide a summary of the counterfactual simulations in which we provide information at scale. We look at the following counterfactuals:

I. Baseline: as in the data.

II. Learn x Late: Provide accurate info about price, quality (i.e. $\hat{x} := x$) at time T_i , taking awareness as given.

III. Learn (x,r) Early: Provide accurate info to to those people who didn't do any online search on their own about distribution of x's, x of known schools, and admissions chance, at time t = 1.

IV. Learn (x,r) Early*: Provide accurate info to everyone about distribution of x's, x of known schools, and admissions chance, at time t = 1.

V. Full Info Benchmark: $\hat{x} = x$ and $\pi_{ij}^T > 1$ for all (i, j) with $\text{dist}_{ij} < 5km$.

VI. Baseline (x^ = x): we assume $\hat{x} = x$.

VII. Full Info ($\hat{x} = x$:) we assume $\hat{x} = x$.

Overall, we find that in order to sort students into schools with high quality indices, the policymaker must provide students with information about schools' quality, but the timing of this information matters. A hypothetical intervention that provides full knowledge of "known" schools' price and quality just before applications are due, but does not affect awareness of schools (II. Learn x Late), would achieve 89% of the impacts on average school quality of endowing all students with full information about all schools (V. Full Info Benchmark:).[14]

However, 80% of students in our search intervention sample receive a placement in some school in the data, but this share would fall to 77% if students were given information about schools' price and quality just before applications are due (II. Learn x Late), and to 73% if students were given accurate information about quality, price, and admissions chances early, as this information would discourage search effort (III. Learn (x,r) Early).

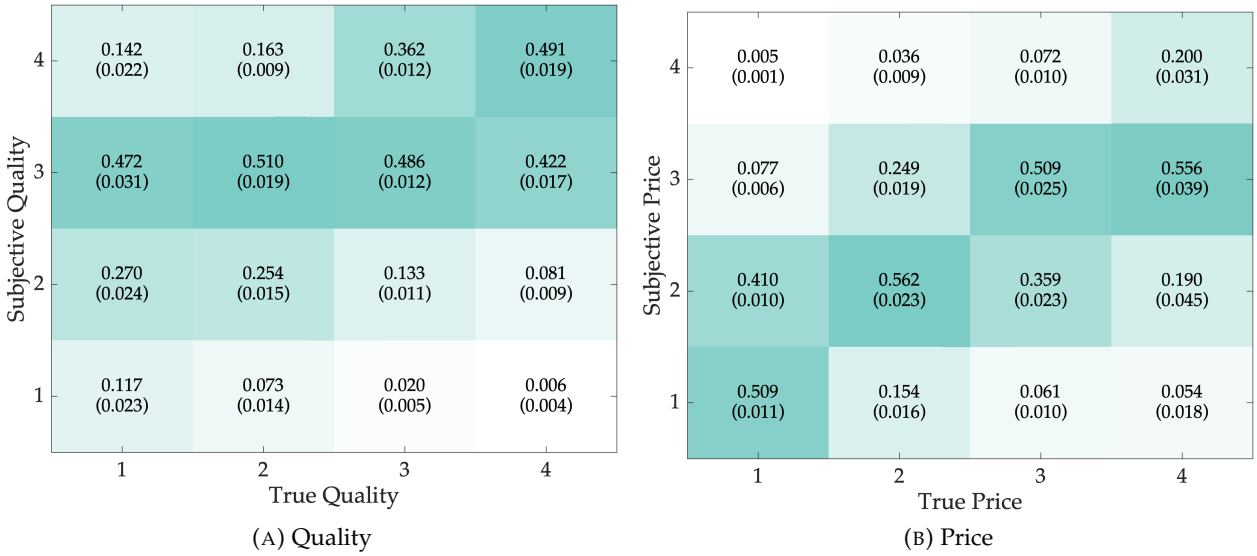
The final two rows of Figure 10 and Table 11 show results from a simpler model in which we ignore survey questions about perceptions of x, and assume that all households observe known schools' quality and price without error. We compare a full-information counterfactual, simulated under these assumptions, to baseline estimates. We find that, had we ignored misperceptions about schools' price and quality, we would have reversed the sign of estimated impacts on school quality. providing full information about all schools would have led households to sort into schools with quality 2.97 on average, lower than at baseline.

[14]Quality indices range from 1 to 4. At baseline, the average quality of the placed school is 3.069, while in this counterfactual it is 3.228, and in our full-information benchmark it is 3.247. The effect is thus 89% = (3.228 - 3.069)/(3.247 - 3.069).

TABLE 10. Model Estimates

	Parameter	Coeff	Std Err.
Panel A: Preferences and Information			
Perceived Price	γ_p	-0.261	(0.042)
Perceived Quality	γ_q	0.671	(0.059)
Know by Name Penalty	β_z	-1.371	(0.054)
Constant	$\bar{\beta}$	-2.471	(0.166)
Price	β_p	-0.147	(0.034)
Quality	β_q	0.446	(0.032)
Constant	$\bar{\alpha}$	-0.383	(0.057)
Price	α_p	-0.076	(0.023)
Quality	α_q	0.222	(0.023)
Measurement Err Baseline Survey Utility	σ^2_ϵ	0.540	(0.062)
Measurement Err Survey Awareness	$\sigma^2_{\eta_s}$	0.017	(0.002)
Pr(learn true x's | know well)	p^h	0.225	(0.011)
Pr(answer some garbage | surveyed about x's)	p^s	0.026	(0.002)
Panel B: Variance Covariance of Random Coefficient			
Variances			
Constant	$\sigma_{constant}$	7.162	(0.867)
Distance	σ_d	0.645	(0.034)
Price	σ_p	0.638	(0.052)
Quality	σ_q	0.088	(0.033)
Covariances			
Constant - Distance	$\sigma_{constant,d}$	-1.315	(0.112)
Constant - Price	$\sigma_{constant,p}$	-1.698	(0.170)
Constant - Quality	$\sigma_{constant,q}$	-0.502	(0.167)
Distance - Price	$\sigma_{d,p}$	0.106	(0.026)
Distance - Quality	$\sigma_{d,q}$	0.016	(0.032)
Price - Quality	$\sigma_{p,q}$	0.113	(0.031)
Panel C: Variance Covariance of δ and η			
Variances			
Mean Utility	σ_δ	1.102	(0.081)
Discoverability	σ_η	0.343	(0.022)
Covariances			
Mean Utility - Discoverability	$\sigma_{\delta,\eta}$	0.116	(0.020)

Notes. This table presents presents results from the model estimation.



(A) Quality (B) Price

FIGURE 8. Distortion Function

Notes: Panel (A) shows the quality distortion function from model estimates. Panel (B) shows the price distortion function from model estimates.

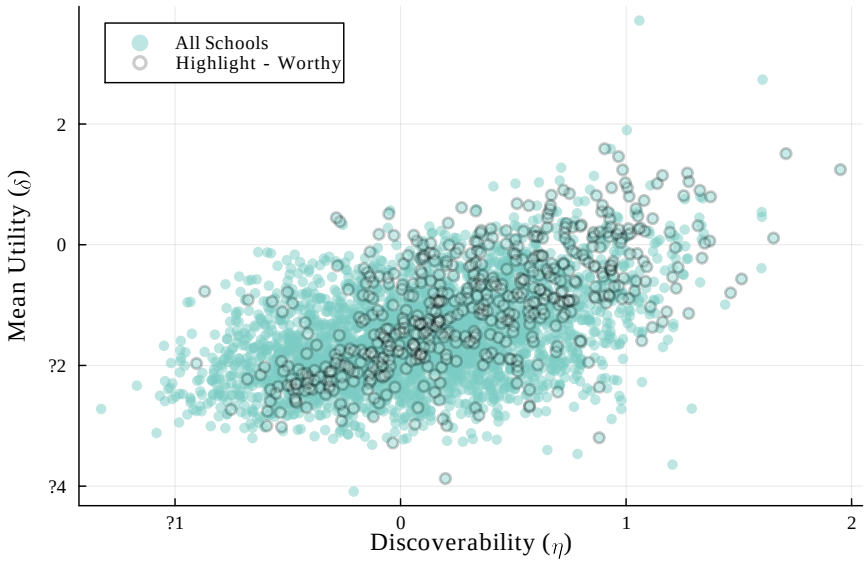
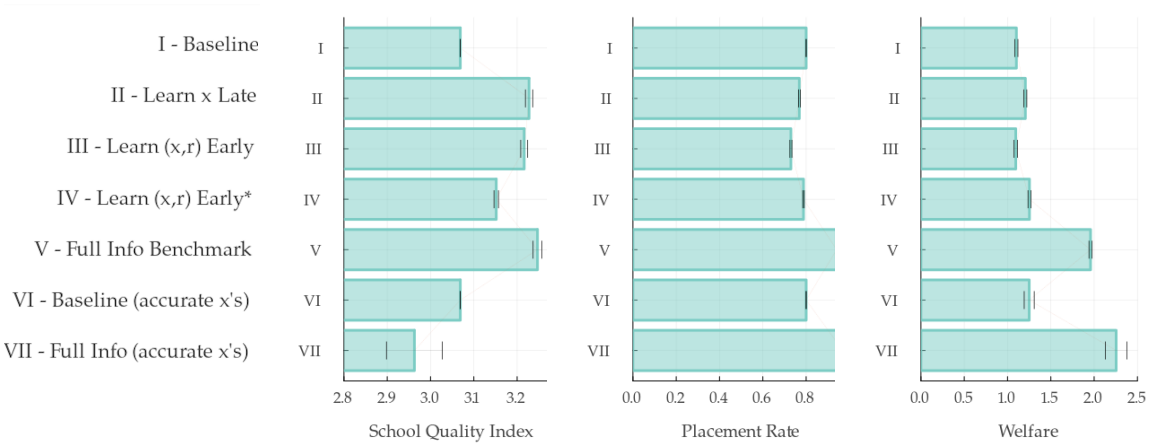


FIGURE 9. Estimates School Unobservables

Notes: This figure shows the relationship between the estimated discoverability and the estimated mean utility for each school.



(A) (B) (C)

FIGURE 10. Counterfactuals

Notes: This figure displays estimated impacts on average school quality, share placed, and welfare according to fully informed preferences. "Baseline": as estimated. "Learn x late": provide full information about price and quality of known schools (i.e. $\hat{x}_{ij} = x_j$ for all schools with $\pi_{ijT} > 0$) just before applications are due. "Learn (x, r) early": full information about x, r of known schools, prior to search. "Learn (x, r) early*": same as "Learn (x, r) early," but only to households who do not search in the data. "Full info benchmark": $\pi_{ijT} > 1$ for all (i, j), and $\hat{x} = x$. "accurate x's": estimate assuming $\hat{x}_{ij} = x_j$ for all i, j.

TABLE 11. Main Results

	EU (perceived)	EU	Place	E(rank)	Distance	Price	Quality
I - Baseline	1.304 (0.016)	1.1 (0.019)	0.8 (0.0)	1.581 (0.0)	1.641 (0.0)	1.362 (0.0)	3.069 (0.0)
II - Learn x Late	1.252 (0.016)	1.204 (0.019)	0.768 (0.004)	1.575 (0.007)	1.629 (0.016)	1.378 (0.006)	3.228 (0.009)
III - Learn (x,r) Early	1.128 (0.016)	1.093 (0.02)	0.73 (0.005)	1.53 (0.008)	1.662 (0.019)	1.372 (0.007)	3.217 (0.008)
IV - Learn (x,r) Early*	1.279 (0.016)	1.252 (0.018)	0.788 (0.003)	1.579 (0.005)	1.642 (0.013)	1.372 (0.004)	3.152 (0.006)
V - Full Info Benchmark	1.958 (0.017)	1.958 (0.017)	0.941 (0.004)	1.999 (0.017)	1.639 (0.018)	1.373 (0.008)	3.247 (0.011)
VI - Baseline ($\hat{x} = x$)	1.318 (0.027)	1.259 (0.033)					
VII - Full Info ($\hat{x} = x$)	2.242 (0.05)	2.242 (0.05)	0.969 (0.003)	1.857 (0.015)	1.651 (0.017)	1.306 (0.011)	2.968 (0.019)

Notes: This table displays main estimates. Columns: "EU (perceived)": expected utility according to households' beliefs at time T_i. "EU": EU according to fully informed payoffs. "Place": probability of placement. ("E(rank)","Distance","Price","Quality"): avg. (rank of placed school on ROL,distance,price,quality), conditional on placement. Rows are as follows. "Baseline": as estimated. "Learn x late": provide full information about price and quality of known schools (i.e. $\hat{x}_{ij} = x_j$ for all schools with $\pi_{ijT} > 0$) just before applications are due. "Learn (x, r) early": full information about x, r of known schools, prior to search. "Learn (x, r) early*": same as "Learn (x, r) early," but only to households who do not search in the data. "Full info benchmark": $\pi_{ijT} > 1$ for all (i, j), and $\hat{x} = x$. "accurate x's": estimate assuming $\hat{x}_{ij} = x_j$ for all i, j.

7 Conclusion

This paper investigates the nature, extent, and impact of families' limited awareness and imperfect information about schools in a large "school choice" market, and the consequences of information frictions for families' search efforts, application decisions, and school assignments.

Through detailed data analysis and innovative field experiments, we provide insights into households' preferences, awareness of schools, search activity, and beliefs about school attributes. We find that, on average, households tend to undertake insufficient search due to excessive optimism about the quality of known and preferred schools. Information interventions, particularly those correcting misperceptions about familiar schools, demonstrate the potential to considerably improve school assignment outcomes.

The results highlight the trade-offs faced by policymakers when providing information to families. Offering information broadly and early has the potential to reach households while they have the opportunity to investigate more schools, but it may discourage some households from engaging in search. Targeted information provision to households who have already submitted applications, and/or to those who appear to have searched little, can encourage more search and enhance demand for quality schools.

We provide a unified framework for analysing information interventions in school choice markets, including information about admissions chances and strategic behavior, and information about schools' characteristics. This framework allows us to understand when interventions providing information about schools may have the most impact, and through what channels they operate. By providing novel data on search activity and outcomes, this paper contributes to our understanding of consumer search in a high-stakes setting and lays the groundwork for further research on equilibrium modeling in school choice markets.

References

Abdulkadiroğlu, Atila and Tayfun Sönmez, "School choice: A mechanism design approach," *American Economic Review*, 2003, *93* (3), 729–747.

Abdulkadiroglu, Atila and Tayfun Sönmez, "School choice: A mechanism design approach," *The American Economic Review*, 2003, *93* (3), 729–747.

_ , Parag A. Pathak, Alvin E. Roth, and Tayfun Sonmez, "The Boston Public School Match," *American Economic Review, Papers and Proceedings*, 2005, *95*, 368–371.

Agarwal, Nikhil and Paulo J Somaini, "Demand analysis under latent choice constraints," Technical Report, National Bureau of Economic Research 2022.

_ and Paulo Somaini, "Demand Analysis using Strategic Reports: An application to a school choice mechanism," *Econometrica*, 2018, *86* (2), 391–444.

Ajayi, Kehinde F, Willa H Friedman, and Adrienne M Lucas, "When information is not enough: Evidence from a centralized school choice system," Technical Report, National Bureau of Economic Research 2020.

Allende, Claudia, Francisco Gallego, and Christopher Neilson, "Approximating the equilibrium effects of informed school choice," Technical Report 2019.

Andrabi, Tahir, Jishnu Das, and Asim Ijaz Khwaja, "Report cards: The impact of providing school and child test scores on educational markets," *American Economic Review*, 2017, *107* (6), 1535–1563.

Arteaga, Felipe, Adam J Kapor, Christopher A Neilson, and Seth D Zimmerman, "Smart matching platforms and heterogeneous beliefs in centralized school choice," *The Quarterly Journal of Economics*, 2022, *137* (3), 1791–1848.

Bergman, Peter, Eric W Chan, and Adam Kapor, "Housing search frictions: Evidence from detailed search data and a field experiment," Technical Report, National Bureau of Economic Research 2020.

Cohodes, Sarah, Sean Corcoran, Jennifer Jennings, and Carolyn Sattin-Bajaj, "When do informational interventions work? Experimental evidence from New York City high school choice," Technical Report, National Bureau of Economic Research 2022.

Corcoran, Sean P, Jennifer L Jennings, Sarah R Cohodes, and Carolyn Sattin-Bajaj, "Leveling the playing field for high school choice: Results from a field experiment of informational interventions," Technical Report, National Bureau of Economic Research 2018.

Correa, Jose, Rafael Epstein, Juan Escobar, Ignacio Rios, Bastian Bahamondes, Carlos Bonet, Natalie Epstein, Nicolas Aramayo, Martin Castillo, Andres Cristi et al., "School choice in Chile," in "Proceedings of the 2019 ACM Conference on Economics and Computation" 2019, pp. 325–343.

Dinerstein, Michael, Liran Einav, Jonathan Levin, and Neel Sundaresan, "Consumer price search and platform design in internet commerce," *American Economic Review*, 2018, *108* (7), 1820–1859.

Gurantz, Oded, Jessica Howell, Michael Hurwitz, Cassandra Larson, Matea Pender, and Brooke White, "A national-level informational experiment to promote enrollment in selective colleges," *Journal of Policy Analysis and Management*, 2021, *40* (2), 453–479.

Haan, Monique De, Pieter A Gautier, Hessel Oosterbeek, and Bas Van der Klaauw, "The performance of school assignment mechanisms in practice," *Journal of Political Economy*, 2023, *131* (2), 388–455.

Handel, Benjamin R and Jonathan T Kolstad, "Health insurance for "humans": Information frictions, plan choice, and consumer welfare," *American Economic Review*, 2015, *105* (8), 2449–2500.

Hastings, Justine S and Jeffrey M Weinstein, "Information, school choice, and academic achievement: Evidence from two experiments," *The Quarterly journal of economics*, 2008, *123* (4), 1373–1414.

Hodgson, Charles and Gregory Lewis, "You can lead a horse to water: Spatial learning and path dependence in consumer search," 2020.

Hoxby, Caroline M and Sarah Turner, "What high-achieving low-income students know about college," *American Economic Review*, 2015, *105* (5), 514–517.

Hoxby, Caroline, Sarah Turner et al., "Expanding college opportunities for high-achieving, low income students," *Stanford Institute for Economic Policy Research Discussion Paper*, 2013, *12* (014), 7.

Kapor, Adam J, Christopher A Neilson, and Seth D Zimmerman, "Heterogeneous beliefs and school choice mechanisms," *American Economic Review*, 2020, *110* (5), 1274–1315.

Kapor, Adam, Mohit Karnani, and Christopher Neilson, "Aftermarket frictions and the cost of off-platform options in centralized assignment mechanisms," Technical Report, National Bureau of Economic Research 2022.

los Santos, Babur De, Ali Hortaçsu, and Matthijs R Wildenbeest, "Testing models of consumer search using data on web browsing and purchasing behavior," *American economic review*, 2012, *102* (6), 2955–2980.

Luflade, Margaux, "The value of information in centralized school choice systems," *Job Market Paper, Duke University*, 2017.

McCulloch, Robert and Peter E Rossi, "An exact likelihood analysis of the multinomial probit model," *Journal of Econometrics*, 1994, *64* (1-2), 207–240.

Mizala, Alejandra and Miguel Urquiola, "School markets: The impact of information approximating schools' effectiveness," *Journal of Development Economics*, 2013, *103*, 313–335.

Moraga-González, José L, Zsolt Sándor, and Matthijs R Wildenbeest, "Consumer search and prices in the automobile market," 2021.

Neilson, Christopher, "Targeted Vouchers, Competition Among Schools, and the Academic Achievement of Poor Students," *Working Paper*, 2021.

Santos, Babur De Los, Ali Hortaçsu, and Matthijs R Wildenbeest, "Search with learning for differentiated products: Evidence from e-commerce," *Journal of Business & Economic Statistics*, 2017, *35* (4), 626–641.

Wang, Tong and Congyi Zhou, "High school admission reform in China: a welfare analysis," *Review of Economic Design*, 2020, *24* (3-4), 215–269.

8 Appendix

8.1 The Centralized School Choice System in Chile

We conduct our intervention within the Chilean School Admission System (SAE), which includes students applying to schools that receive public funds. The SAE is a centralized system that allows students to apply to multiple schools, and rank them in order of preference. The SAE is a deferred acceptance mechanism, where students are assigned to schools in order of priority. The SAE is administered by the Ministry of Education (Mineduc), and is the main mechanism for assigning students to schools in Chile.

There are three types of educational provisions in Chile: public schools owned and managed by the state mainly through municipalities, privately owned and managed schools subsidized by the state (voucher schools), and private schools owned and managed by the private sector. Voucher schools account for 55.58% of the total enrollment, and can charge out-of-pocket fees while receiving flat subsidies for each student, depending on the grade. [15] If a voucher school holds a *Subvención Escolar Preferencial* (SEP) agreement, students from low socioeconomic status do not pay any fee (for more details, see Neilson (2021)). [16]

Chile holds a student-proposed deferred acceptance system for centralized assignment. On a single nationwide online platform, parents with children from all levels, from Pre-K to 12th grade, apply to all public and voucher schools. Thus, the off-platform schools are those fully private, gathering 9% of the total enrollment in 2019. For a detailed description of how the school admission system is implemented, see Correa et al. (2019).

In detail, 454,415 students participated in the school admission system in 2020, facing 1,035,133 slots offered by 8,014 schools. SAE applicants concentrate on entry levels: Pre-K (25.43%), 1st grade (13.54%) and 9th grade (24.38%). [17]

In order to secure admission equity and non-discrimination, the system allocates students following publicly known criteria, based on ordered priorities and quotas. The established priorities along with their 2020 shares among students are:

a) Students with secured enrollment (29.93%).[18]

[15] In 2015, the School Inclusion Law froze the co-payment, which will gradually fade out while subsidized funds increase.

[16] There are several other vouchers based on geographic location or student's special educational needs, but with significantly less participation on the school budget compared to the flat voucher and SEP subsidy.

[17] Some schools start their instruction on 7th grade, particularly the high achieving ones, but still congregate a small fraction of the applicants (6.02%).

[18] Students with secured enrollment are those who, by being already enrolled in one school and applying to transfer to another, are guaranteed a slot at their original school if they are not assigned to one in their ranked ordered list.

b) Students with a sibling already enrolled in the school (21.77%).
c) Student with parents that work at the school (1.59%).
d) Former students which have not been expelled (4.67%).

SAE quotas work in an orderly manner as described:

1. Applicants with permanent special educacional needs (PSEN): There are, at most, two seats reserved by classroom in schools with a certified agreement (68.27% of schools). In 2020, 2.7% of the total slots were reserved for this purpose, while 1.4% of all applicants were certified with SEN.

2. High achieving students: A restricted set of schools can select between 30% to 85% of their seats in 7th and 9th grade, based on an entrance test or, alternatively, student class ranking from the previous year. In 2020, only 66 schools (0.8% of all schools with 7th and 9th grade) selected, accounting for 8,148 slots (2.4% of the total offered in 7th and 9th grade).

3. Students from low socioeconomic status (low SES): 15% of the seats within each grade in a school must be held by students from low SES, regardless of whether or not the school has a SEP agreement.[19] To meet this requirement, 7.1% of the vacancies offered in the 2020 process were reserved for low SES students, whereas 46.2% of applicants were categorized as low SES students.[20]

4. Regular students: When all three quotas mentioned above have been filled, the remaining slots are assigned in order of priority. 89.2% of the vacancies available in the process were allocated to the regular quota.

TABLE 12. Priorities, Quotas and Slots in the School Admission System in Chile

		Quotas		
Priorities	Special Needs	High-achieving	Disadvantaged	Regular
1	Secured enrollment	Secured enrollment	Secured enrollment	Secured enrollment
2	Special Needs	High-achieving	Siblings	Siblings
3	Siblings	Siblings	Disadvantaged	Working parent
4	Working parent	Working parent	Working Parent	Former students
5	Former students	Former students	Former students	No priority
6	No priority	No priority	No priority	
Slots	*At most 2 seats per classroom*	*Between 30% to 85% of total seats*	*15% of seats by level*	*Remaining seats*

Table based on **?**.

[19]The *Subención Escolar Preferencial* (SEP) agreement aims to assist low SES students. One of the central axes of the SEP agreement is that students with low SES who attend a school with such agreement do not have to pay their tuition, while the school receives a state subsidy for each student enrolled in return. 93.0% of the schools who participated in the SAE in 2020 had a SEP agreement.

[20]Whether a student is categorized as low SES or not, depends mainly on their score in the social register of households.

The system uses a multiple-tie-breaking lottery whenever applicants exceed slots in a school. Sibling applicants get a single ballot to increase their chances of being assigned to the same school. In addition, families can increase their chances of being accepted in the same school by choosing a *family application*. [21]

Information about slots, academic programs, government programs in which each school participates, addresses, fees, and quality index is provided openly through the platform. Families have around one month to apply to as many schools -and as many times- as they like in the regular stage of the process, starting in mid-August. Then, after first round of results, they accept or reject their offers. In case of rejection or if someone did not participate in the first stage, a complementary stage takes place, with the same rules, offering the remaining slots. Collectively, the admission process takes about four months, including enrollment. The default option in both stages whenever a student is not assigned to any seat in their listed schools is to be assigned to a nearby school with vacant seats.

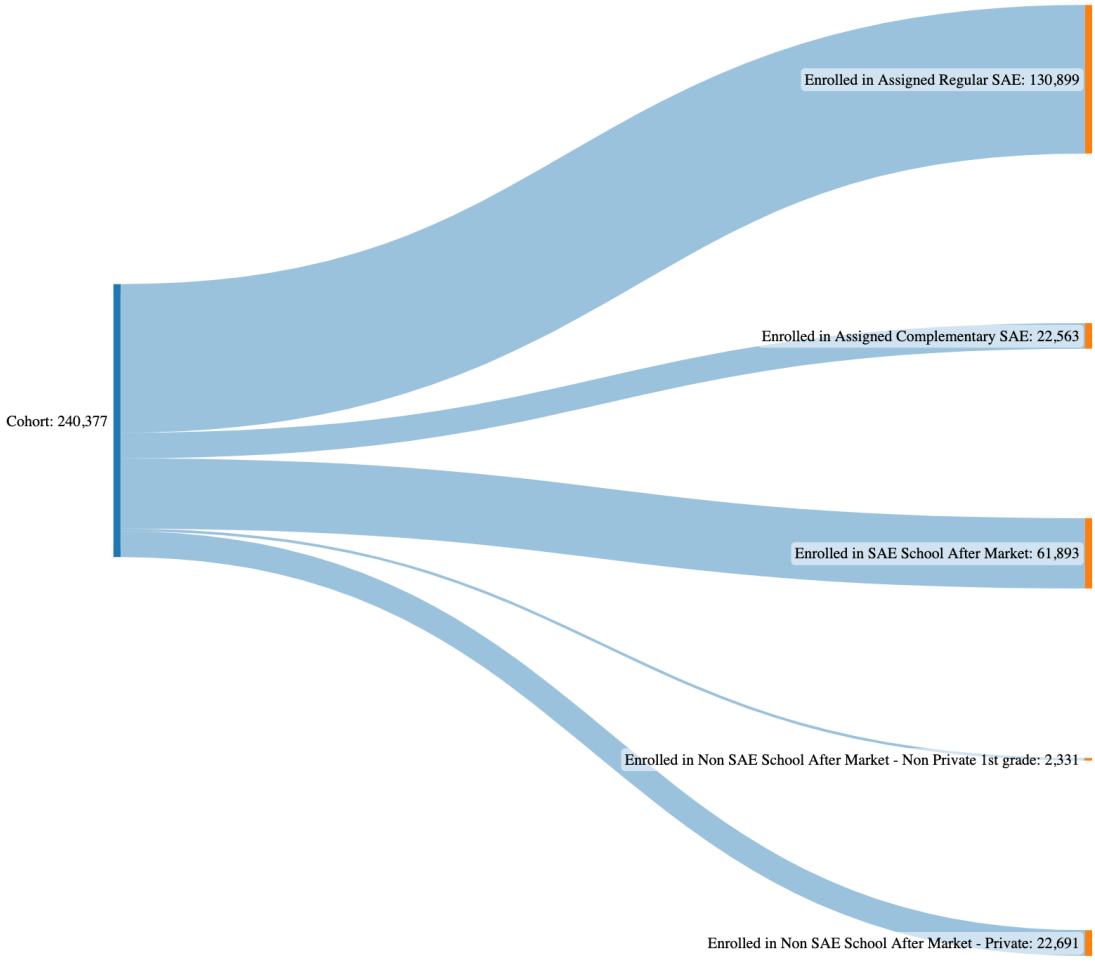
Looking at the 2021 regular stage, 56.98% of the families applied in the first five days of the regular stage, while 10.9% applied in the last five. As parents are allowed to change their application during each stage, 26.93% changed theirs during the regular stage.

In the regular stage, 91% of the applicants were assigned to a school. Among them, 82% were assigned within their first two options and 68% to their most preferred one. 8,867 of the 185,188 (4.9%) families rejected their assigned seat, which means they either participate in the complementary stage, enroll in a school through the aftermarket or apply to a private school out of the centralized platform.

In this particular context, we're interested in recreating a particular cohort in the regular education system. In 2021, the total number of students enrolled in 1st grade was 252,368, where approximately 10% of students are enrolled in a private school (25,362). We can assume that all students appear in the system in 1st grade and either applied by SAE or aftermarket. In this manner, we can re create how a cohort of students would have been assigned to a school in 2022. We separate the students in entry grade (Pre-Kinder, Kindergarten and 1st grade) in enrollment by the regular, complementary, and aftermarket process (See Figure 11). From this snapshot of how a cohort is assigned and enrolled in each school, we have that 55% comes from enrollment in assigned school in regular process, 9% comes from assignment in complementary process, and 26% comes from enrollment in the aftermarket for schools with SAE slots. The remaining 10% comes from enrollment in private schools or schools that do not participate in SAE.

[21] If *family application* is chosen, being assigned to the same school as an older sibling participating in the same process takes precedence over the particular preferences of younger siblings.

FIGURE 11. Cohort by Enrollment



In the same manner, it's neccesary to understand the flow of applications and how students are assigned once the apply. From Figure 12, we can see that in total, 185,188 students applied to SAE in an entry grade level in the regular round. In particular, we have that 68% of students that apply in the regular round are assigned to their first preference and 9% remained unassigned. Furthermore, 71% of students that apply in the regular round enroll in the school assigned by the system. Similar analysis can be done for the 23,442 students who applied only in the complementary period, where 77% of students are assigned to their first preference and 69% of students enroll in the school assigned by the system.

FIGURE 12. Regular Round SAE

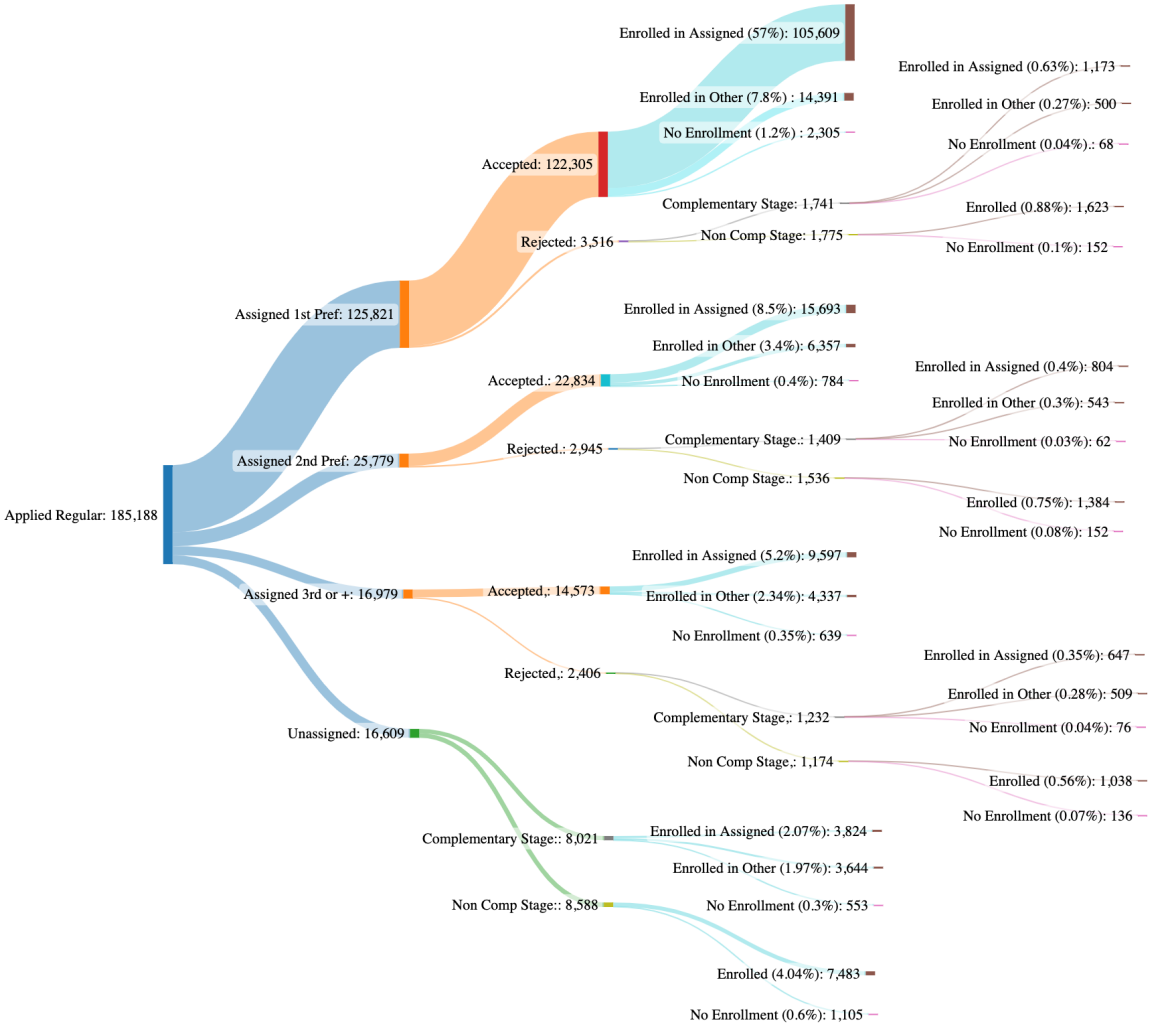
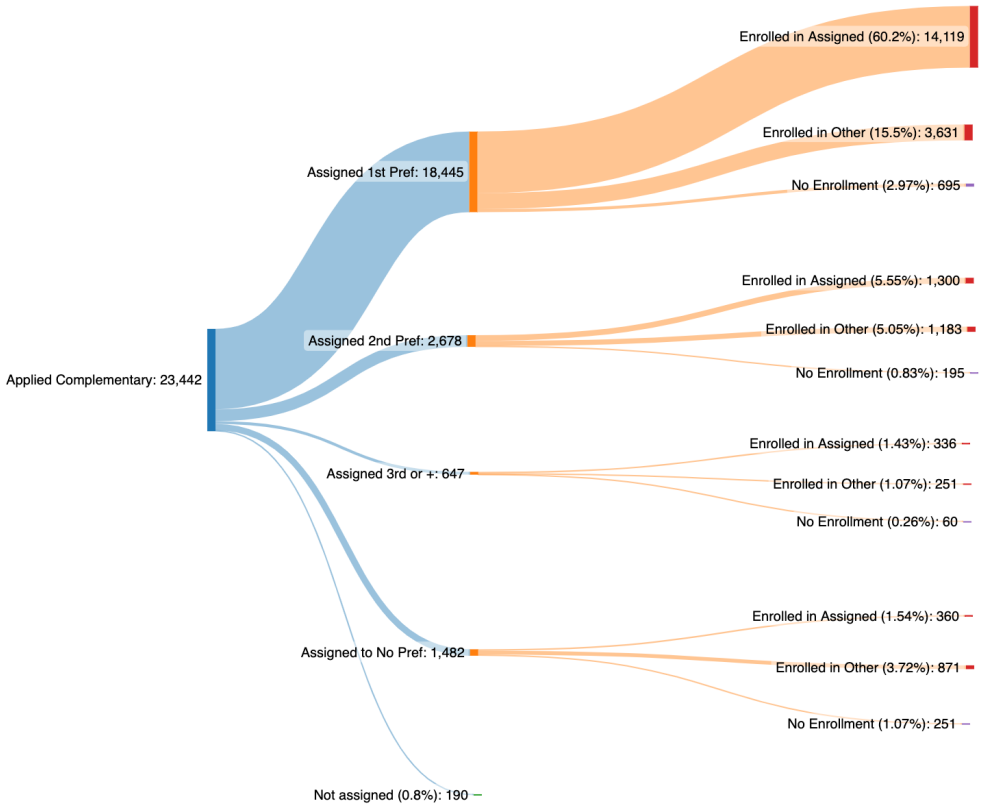


FIGURE 13. Complementary Round SAE



8.2 Extra

The Chilean SAE had a gradual implementation, reaching full coverage in 2020. Magallanes Region, the least numerous out of sixteen, acted as a pilot in 2016 with its entry levels, continuing with the remaining levels the following year. [22]. Other regions proceeded that way until 2020, when Santiago Metropolitan Region, containing Chile's capital and largest city, entered last and fully to the SAE.

[22]Entry levels consider Pre-K, 1st grade, 7th grade, 9th grade.

8.3 Experiment Sample Recruitment

Between May 25 and July 2, with the support of the Subsecretaría de Educación Parvularia, parents potentially eligible to participate in the study were recruited through an e-mail campaign sent to kindergarten directors. In this mailing we asked them to share an invitation with the parents of their institution to inform them of this initiative and in turn to share with them the personalized enrollment link so that interested parents could fill in their contact information and fill in a short form that would allow us to verify their eligibility.

The eligibility condition to be part of the study was to be a parent or caregiver of children enrolled in kindergarten education at the middle school, pre-kindergarten and kindergarten levels and who were interested in applying to an educational institution through the SAE in 2021. As a result of this recruitment, around 33 thousand parents from approximately 2,700 kindergartens were enrolled, of which 48% (15,854 parents) completed the eligibility form.

TABLE 13. Application Outcomes + Enrollment

	Highlighted		Free		High Quality		Value Added		Distance		Knew School (Base)	
Panel A: All Choices	Coef	StdErr	Coef	StdErr	Coef	StdErr	Coef	StdErr	Coef	StdErr	Coef	StdErr
Treatment 1	-0.004	(0.017)	0.005	(0.019)	-0.011	(0.014)	0.015	(0.016)	-0.052	(0.047)	-0.038***	(0.013)
Treatment 2	0.025	(0.017)	0.021	(0.018)	0.006	(0.013)	0.010	(0.015)	0.005	(0.046)	-0.025**	(0.012)
pval: T1 = T2	0.089		0.404		0.229		0.771		0.248		0.319	
Outcome Mean	0.565		0.658		0.848		0.207		1.388		0.259	
Observations	5927		5987		5974		5843		5521		6038	
Panel B: First Choice												
Treatment 1	-0.023	(0.025)	0.013	(0.025)	-0.024	(0.019)	-0.007	(0.023)	-0.073	(0.053)	-0.041*	(0.025)
Treatment 2	-0.023	(0.024)	0.019	(0.025)	-0.013	(0.018)	-0.021	(0.023)	0.027	(0.056)	-0.039	(0.024)
pval: T1 = T2	0.998		0.798		0.583		0.552		0.076		0.945	
Outcome Mean	0.597		0.641		0.887		0.296		1.095		0.450	
Observations	1608		1621		1625		1594		1542		1637	
Panel C: Second Choice												
Treatment 1	0.026	(0.029)	-0.003	(0.027)	0.016	(0.022)	0.034	(0.025)	-0.127**	(0.059)	-0.068***	(0.022)
Treatment 2	0.079***	(0.027)	0.022	(0.026)	0.046**	(0.021)	0.057**	(0.025)	-0.046	(0.060)	-0.042*	(0.022)
pval: T1 = T2	0.057		0.355		0.155		0.331		0.155		0.223	
Outcome Mean	0.558		0.659		0.827		0.180		1.355		0.284	
Observations	1596		1621		1610		1572		1517		1634	
Panel D: Assigned School in SAE												
Treatment 1	-0.036	(0.028)	-0.017	(0.025)	-0.023	(0.023)	0.018	(0.025)	-0.071	(0.059)	0.001	(0.025)
Treatment 2	-0.042	(0.027)	-0.037	(0.025)	-0.008	(0.022)	-0.017	(0.024)	-0.001	(0.059)	-0.012	(0.024)
pval: T1 = T2	0.816		0.439		0.519		0.153		0.250		0.590	
Outcome Mean	0.628		0.753		0.824		0.168		1.224		0.325	
Observations	1563		1578		1574		1536		1477		1592	
Panel E: Enrolled School 2022												
Treatment 1	-0.040	(0.029)	-0.010	(0.025)	-0.039*	(0.023)	0.001	(0.026)	-0.127**	(0.062)	-0.001	(0.025)
Treatment 2	-0.032	(0.028)	-0.021	(0.025)	-0.026	(0.022)	-0.027	(0.025)	-0.113*	(0.061)	-0.021	(0.025)
pval: T1 = T2	0.773		0.643		0.580		0.271		0.824		0.414	
Outcome Mean	0.615		0.741		0.854		0.203		1.286		0.327	
Observations	1414		1546		1425		1387		1443		1588	

Notes. This table presents pdslasso regression on school level characteristics for application, assignment and enrollment outcomes. Within the school attributes, we include indicator variables for characteristics such as High Quality, Free, Highlighted (high quality + free), Value Added, Distance to the school, and if the parent knew the school before the intervention. *Panel A* aggregates all schools in the application, *Panel B* considers only the first preference, and *Panel C* considers only the second preference. *Panel D* refers to the school assigned after the whole SAE process (including complementary) and *Panel E* refers to the school enrolled in 2022. We have N = 6,038 distinct participant - school observations, N = 1,637 observations for first preference, N = 1,634 observations for second preference, N = 1,592 observations for being assigned a school after whole SAE process (42 weren't assigned in the regular process and did not go to complementary), and N = 1,588 observations for enrolled school. Missings in administrative records account for differences within characteristics. Restrictions: Sample 1 and Start Primer and Clicks

TABLE 14. Application Outcomes Continued

	Application		Expected Characteristics			
	Submitted Application	Application Length	Highlighted	Free	High Quality	Knew at Baseline
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Application						
Treatment 1	0.011	0.044	-0.024	0.005	-0.031	-0.028
	(0.016)	(0.150)	(0.022)	(0.022)	(0.020)	(0.019)
Treatment 2	0.015	0.023	-0.014	0.006	-0.005	-0.034*
	(0.016)	(0.113)	(0.022)	(0.021)	(0.020)	(0.019)
pval: T1 = T2	0.827	0.887	0.637	0.973	0.199	0.741
Outcome Mean	0.901	3.824	0.561	0.656	0.756	0.319
Observations	1801	1637	1637	1637	1637	1637

	Reject Assignment	Not Assigned	Complementary Period	Has + 3	Has + 4	Assigned to 1st
Panel B: Assignment						
Treatment 1	0.010	-0.030**	-0.018	-0.033	-0.031	-0.002
	(0.013)	(0.015)	(0.014)	(0.029)	(0.025)	(0.030)
Treatment 2	0.015	0.004	0.005	0.007	-0.003	0.022
	(0.013)	(0.016)	(0.015)	(0.029)	(0.026)	(0.030)
pval: T1 = T2	0.738	0.023	0.110	0.167	0.269	0.416
Outcome Mean	0.044	0.081	0.071	0.456	0.265	0.644
Observations	1633	1633	1801	1634	1634	1513

	Enroll in Assigned Final SAE	Enroll in Assigned Regular SAE	Enroll in SAE School	Enroll in School in App	Enroll in 1st Choice	Enroll in 1st or 2nd Choice
Panel C: Enrollment						
Treatment 1	0.055**	0.035	0.029	0.027	0.026	0.028
	(0.026)	(0.026)	(0.020)	(0.026)	(0.030)	(0.028)
Treatment 2	0.053**	0.035	0.040**	0.043*	0.037	0.047*
	(0.026)	(0.026)	(0.020)	(0.026)	(0.030)	(0.028)
pval: T1 = T2	0.963	0.996	0.554	0.546	0.703	0.496
Outcome Mean	0.737	0.769	0.859	0.687	0.519	0.673
Observations	1588	1471	1588	1801	1588	1588

Notes. This table presents pdslasso regressions on treatment effects on further application results. *Panel A* refers to the application, where columns (1) and (2) are if the participant submitted an application and the length of the application (i.e how many programs they applied to). We have application data for N = 1,637 participants. Columns (3) to (6) refer to the expected characteristics of the assigned school, integrating over the probability of being assigned to each program in the application. *Panel B* refers to the assignment, where Columns (1) and (2) are if the participant either rejected the assigned school or was not assigned. Column (3) is an indicator if the participant applied in the complementary period. Column (4) and (5) refers to how many distinct programs the participant has in their application. Column (6) is an indicator if the participant was assigned to their first preference in the regular round, conditional on being assigned. We have assignment data for N = 1,633 participants (3 participants are from application 2022 so we include them in application analysis but not in assignment analysis and 1 is not found in the assignment data). *Panel C* refers to the enrollment, where Column (1) indicates if the participant enrolled in the assigned school after the whole SAE process (including complementary) and Column (2) is if enrolled in the assigned school after the Regular Round of SAE. Columns (3) and (4) is if the participant is enrolled in a school that was available in SAE and if the participant is enrolled in a school that was in their application. Columns (5) and (6) are if the participant is enrolled in their first choice or in their first or second choice, conditional on being enrolled. We have enrollment data for N = 1,588 participants (we don't observe enrollment in private day cares). Restrictions: Sample 1 and Start Primer and Clicks