Search and Biased Beliefs in Education Markets

Patrick Agte
Yale University

Claudia Allende
Stanford GSB

Adam Kapor
Princeton University

Christopher Neilson
Yale University

Fernando Ochoa
New York University *

November 29, 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 of 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, affecting their search decisions and application decisions. Most importantly, households' perceptions systematically overstate the quality ratings of schools that they know and like. Correcting misperceptions about known schools causes students to match to schools with higher quality, equal to what can be achieved under a full-information benchmark, and closes the quality gap between low-SES and high-SES applicants.

*Emails: patrick.agte@yale.edu, callende@stanford.edu, akapor@princeton.edu, christopher.neilson@yale.edu, nano.ochoa@nyu.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, María Constanza Muñoz, 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 in "school choice" markets with many 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, when schools vary along many dimensions. 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 desirability 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 asks: 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? We address this question 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. We collect novel data on school choice participants' search activity and subjective beliefs, conduct field experiments, and construct and estimate a model of search and demand for schools.

We develop and provide a "school explorer" web app with personalized information about schools. We use 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 administrative application data 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 conduct 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 applications.

[1] Hastings and Weinstein (2008)

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

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 perceive 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 objective 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 ideal forms of information interventions.

Our main finding is that households search too little, on average, and misrank the schools they know, 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. Following the literature on the importance of parental education, We explore heterogeneity by college-graduate and non-college mothers, a proxy for SES. At baseline, there is a gap between high- and low-SES households of roughly a tenth of a point on a government quality index, or 0.045 student-level of value added. Providing perfect information about all characteristics (including unobservables) of all schools, an infeasible counterfactual, would raise quality by roughly .2 points for low-SES households, and fully close SES gaps in quality and value-added measures. However, providing perfect information about "known" schools' quality and price would also achieve these gains and close these gaps.

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. We also find explorer search activity predicts knowledge of schools, and accurate beliefs about their characteristics.

Second, we analyze our experiments. We start by showing that the treatments affected households' reported beliefs on the perceived availability of schools in their neighborhoods. These effects are larger for college mothers. Consistently, we find that the first treatment intervention also increases search for college mothers, measured by the number of clicks in the explorer. In the midline surveys, these mothers also report knowing 46% more schools at least by name. These changes matter for enrollment. Relative to their control group counterparts, children from college mothers in the first treatment group enroll in schools with higher levels of value added. In contrast, the "report card" intervention, which corrects misperceptions about known schools, has large impacts on application behavor for non-college mothers.

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. The timing of information may matter: providing information "early", before search takes place, may discourage some households from searching, resulting in reductions in placement rates. A misspecifed model in which we ignore misperceptions of known schools' price and quality indices gives the wrong sign of the impacts of providing information on the quality of students' matched schools.

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 practices 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 to accomodate panel data and allow misperceptions of school characteristics.

A limitation of this paper is that we estimate demand, and conduct "single-agent" counterfactuals, holding school characteristics fixed.[3] Understanding search and demand is a needed input for further research that models equilibrium in this market.

The remainder of the paper proceeds as follows. Section 2 presents a simple rational framework to fix ideas and motivate our data collection and research design. Section 3 describes the setting. Section 4 provides descriptive analysis. Section 5 describes and evaluates the experiments. Section 6 presents the model, section 7 describes estimation, and section 8 presents results. Section 9 concludes.

2 Framework

This section presents a simple framework for search and application portfolio formation that motivates our research design and empirical analysis. We begin with a simple rational benchmark. We then consider the comparative statics of optimal search decisions. Finally, we discuss a set of possible distortions and biases in search and portfolio formation — departures from the rational benchmark — that motivate our research design and empirical analyses. Our full empirical model is given in section 6.

Portfolio Problem: Consider an agent submitting a rank-order list to a strategyproof school choice mechanism. Let $\kappa = \{(u_1, r_1), \ldots, (u_N, r_N)\}$ be a finite set of schools that the agent knows and may apply to. The agent receives a payoff u_j if he matches to the jth option, but faces uncertainty about whether this option will admit him. He is rejected by option j with probability r_j, independently across options.[4] We assume that the choice set κ contains some option with payoff and rejection chance both equal to zero.[5] Without loss, suppose $u_1 > u_2 > \ldots > u_{M_\kappa} > u_{M_\kappa+1} = 0 > \max_{j>M_\kappa+1} u_j$, for some $M_\kappa \geq 0$.

When the student-optimal stable matching mechanism is used, it is well known that truthful

[3] 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.

[4] Independence is consistent with the use of independent tiebreaking lotteries.

[5] An agent may choose to go unmatched. Setting this payoff to zero is a normalization.

reports are optimal. The value of the optimal portfolio is given by:

$$U(\kappa) = \sum_{j=1}^{M_\kappa} \left(\prod_{k=1}^{j-1} r_j \right) (1 - r_j) u_j. \tag{1}$$

The payoff $U(\kappa)$ is increasing in $\hat{u}_j$ and decreasing in $\hat{r}_j$ for all j, strictly if $j \leq M_\kappa$ and $R_j > 0$.

Returns to Search: We now consider the value of enlarging the choice set. First, consider a new school of known payoff and rejection chance, (u', r'). If $\hat{u}_{k-1} > u' > \hat{u}_k$ for some $k \in \{1, \ldots, M + 1\}$, then is optimal to add this school to the rank-order list in the kth place. In this case, the school is relevant only if the agent is rejected by schools 1 through $k - 1$. Let $\underline{k}(u', \kappa) = \min\{j : u_j < u'\}$ be the rank at which the school would be inserted. The probability of joint rejection by all j offering $u_j > u'$ is given by $\underline{R}(u'; \kappa)$ where,

$$\underline{R}(u'; \kappa) = \prod_{k=1}^{\underline{k}(u',\kappa)} r_j.$$

The agent's expected payoff under κ in the event of rejection by schools 1 through $\underline{k}(u, \kappa) - 1$ is given by:

$$\underline{U}(u'; \kappa) = \sum_{j=\underline{k}(u',\kappa)}^{M_\kappa} \left(\prod_{\ell=\underline{k}(u',\kappa)}^{j-1} \hat{r}_j \right) (1 - \hat{r}_j) u_j.$$

The value of adding (u', r') to the choice set is therefore given by:

$$U(\kappa \cup \{(u', r')\}) - U(\kappa) = R(u, \kappa)(1 - r')(u' - \underline{U}(u', \kappa)). \tag{2}$$

Expanding the term $\underline{U}(u', \kappa)$ in equation (2) gives:

$$U(\kappa \cup \{(u, r)\}) - U(\kappa) = R(u, \kappa)(1 - r) \sum_{j > \underline{k}(u,\kappa)} \left(\prod_{\ell=\underline{k}(u,\kappa)}^{j-1} r_\ell \right) (1 - r_j)(u - u_j).$$

The marginal value $U(\kappa \cup \{(u', r')\}) - U(\kappa)$ is weakly increasing in u' and $(1 - r')$, weakly decreasing in u_j and $(1 - r_j)$ for all j, and convex in u'.[6] Let $V(\kappa, F) := E_F\left(U(\kappa \cup \{(u', r')\}) - U(\kappa)\right)$ denote the value of one new draw from a distribution $F(u', r')$ when the current consideration set is κ. The value of a draw from F is:

$$V(\kappa, F) = \int R(u; \kappa)(1 - r)(u - \underline{U}(u; \kappa)) dF(u, r). \tag{3}$$

[6] From equation 2, we see that $U(\kappa \cup \{(u', r')\}) - U(\kappa)$ is continuous and piecewise linear in u', with $\frac{\partial}{\partial u'}(U(\kappa \cup \{(u', r')\}) - U(\kappa)) = R(u, \kappa)(1 - r')$ almost everywhere. By construction $R(u, \kappa)$ is nondecreasing in u.

Comparative Statics of the Returns to Search: Equation 3 has a simple interpretation. If a student draws a new school with payoff and rejection chance (u, r), it is relevant with probability $R(u; \kappa)$. In this event, with probability $1 - r$ the student is admitted and receives u. If the student were rejected, or had not added this school, she would have instead received $\underline{U}(u, r)$, the continuation value of the remaining part of the rank-order list.

The following comparative statics are immediate:

1. Consider $\kappa' \neq \kappa$. We have $V(\kappa, F) > V(\kappa', F)$ for all F if any only if $R(u, \kappa) - R(u, \kappa'))u > (R(u, \kappa)\underline{U}(u, \kappa) - R(u, \kappa')\underline{U}(u, \kappa'))$ for all u.
2. Let F and G be distributions over (u, r) satisfying independence of u and r, and with identical marginals over r, $F_r(\cdot) = G_r(\cdot)$. We have $V(\kappa, F) \geq V(\kappa, G)$ if and only if $F \succeq_{FOSD} G$.[7]

The first condition says the difference in gains $(R(u, \kappa) - R(u, \kappa'))u$ must outweigh the difference in opportunity costs. The second condition says that, for F to have a higher return than G generally, it is not enough for F to have a higher expected value of u. For example, if $\underline{U}_1$ is large it may be valuable to draw from a distribution with a small probability of a very high value of u, rather than one with a larger expected value but thinner tails.

The portfolio problem leads to challenges that that are not present when the agent is restricted to choose a single option after stopping search. For instance, because the value of drawing from a distribution F depends on the set κ, unlike in Weitzman (1979) we do not have a simple index rule characterizing the optimal policy when one may choose to draw from either F or G.

Optimal Simple Sequential Search: We now show that, when the agent may repeatedly draw from a fixed distribution at a cost nondecreasing in the number of schools drawn, the optimal stopping rule is a threshold rule.

Consider an agent who is endowed with an initial state $\kappa_0 = \{(u_1, r_1), \ldots, (u_k, r_k)\}$, and may add to this set by repeatedly drawing a new (u, r) from a distribution F, indepedently of the current state and all previous draws. After each draw, the agent may decide to stop or continue.

<u>Claim:</u> Suppose the cost of the nth draw is $c_n \geq c_{n-1} > 0$. Let κ_{n-1} denote the state after $n - 1$ draws. The optimal policy is to search if and only if $c_n > V(\kappa_{n-1}, F)$.

<u>Proof:</u> The proof is by induction on the maximum number of additional draws. Let (κ, n) be arbitrary, and suppose that an agent in state (κ, n) may take at most one additional draw. Then he should do so if and only if $c_n < V(\kappa, F)$. Now consider an agent in state (κ, n) who may take a maximum of $m + 1$ additional draws. Observe that, for any $\kappa' \supset \kappa$ we have $R(u, \kappa') \leq R(u, \kappa)$

[7] It suffices to show that $U(\kappa \cup \{(u, r)\}) - U(\kappa)$ is nondecreasing in u. We have $R(u, \kappa)$ nondecreasing in u, and $u > u_j$ for any $j > \underline{k}(u, \kappa)$ by construction. It follows that the RHS of equation 2 is nondecreasing in u.

and $\underline{U}(u, \kappa') \geq \underline{U}(u, \kappa)$. We therefore have $V(\kappa, F) \leq V(\kappa \cup (u, r), F)$ for any (u, r). If $V(\kappa, F) > c_n$ but the agent's policy prescribes stopping, the agent can improve by searching once and then stopping. If $V(\kappa) < c_n$ but the agent's policy prescribes at least one more draw, then after this draw the agent will be in some state $\kappa' \supset \kappa$ with m draws remaining. Because $V(\kappa', F) \leq V(\kappa, F)$ and $c_{n+1} \geq c_n$, the agent can improve by stopping at this point. Hence the agent must either stop at (κ, n) or take one more draw, in which case it is optimal to stop at (κ, n). ■

Finite Choice Sets and "Known" Schools: An additional complication is that, in our setting, an agent does not have access to an infinite stream of independent draws from a fixed distribution F. Rather, each household i has access to finite set of local schools, J_i.

Suppose that, at state κ, school j is chosen from the remaining options with probability Pr_j. Conditional on choosing option j, we have $(u, r) \sim F_j(\cdot)$ for some F_j capturing the distribution given the agent's information, so that F is a mixture over the F_j's, with weights Pr_j. In one limit, Pr_j selects the optimal next school with probability 1. In the other extreme, it selects schools at random. The value of search will be higher when each F_j is higher in the sense of FOSD, and also when Pr_j selects the better options with higher probability.

Distortions and Biases: Let us suppose that households search whenever the payoff of one more search exceeds that of stopping, as in equation 3. Consider an agent with perceived choice set $\hat{\kappa} = \{(\hat{u}_1, \hat{r}_1), \ldots, (\hat{u}_M, \hat{r}_M)\}$, probability Pr_j of sampling school j, and beliefs $\hat{F}_j$ over (u, r) conditional on sampling school j. We may observe low search activity because of high search costs c_n, because the search technology makes it difficult to find the good schools (as summarized by low Pr_j for good options), because of inaccurate beliefs about the known schools $\hat{\kappa} \neq \kappa$, or because of biased beliefs $\hat{F}_j$ about the value of not-yet-investigated schools.

If an agent overestimates the payoffs u_j, he will overestimate $U(u, \hat{\kappa})$, reducing the subjective returns to search. Misperceptions of known schools' payoffs may also distort the applicant's rank-order list, causing further losses conditional on the set of known schools. Systematic errors about rejection chances r_j of known schools may also affect the subjective returns to search.

If households are optimistic about schools that they believe they know, then feedback on schools that they are currently applying to may help; if agents perfectly observe all relevant characteristics of "known" schools, in contrast, this type of information should have no effect.

If households are pessimistic about unknown schools, information about the distribution of unknown schools' characteristics may affect beliefs $\hat{F}_j$ for j not currently known. In addition, improvements in search technology that raise the chances of finding the good schools (by shifting Pr_j) may also raise the probabilty that households in fact search and find those options.

After describing the choice environment, we present interventions along these lines in the following section.

3 Empirical Setting

We start by describing the institutional context and then summarize the study implementation and data collection.

3.1 Centralized School Choice in Chile

We implemented our study within the centralized school choice system in Chile (SAE). The centralized process covers 93% of primary school matriculation in the country, including almost all public schools and private schools that accept school vouchers. Our data come from the 2021 admission process.

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, alumni, and economic status priorities with lottery-based tiebreakers. Applicants may list as many schools as they want on their choice application, making the mechanism 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. 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. 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.[8]

In 2021, there were $7,979$ schools with available slots and $461,223$ students applying to grades ranging from Pre-K to 12th grade. Parents generally apply to schools within 2km of their home.[9] On average, there are approximately 15 schools available in the SAE system in a 2km radius of a student's home. Of this possible choice set, there are on average 12 schools that are free for each

[8] 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.

[9] In estimation we consider a radius of 5km. The median distance for all applications is 1.21 km. The average distance for an applicants' full application is 4.53 km. This average includes a right tail of applicants making long-distance moves. For over 75% of applicants, the distance from their home to schools in their application portfolio averages less than 2km.

student and 9 schools that are classified as medium or high quality according to the Education Quality Agency, an autonomous public service organization.[10] [11]

Despite having many options, households apply to relatively few options, potentially leading to non-placement risk (see Arteaga et al. (2022)). 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.[12] Omissions include many schools that are free and have medium-high or high quality ratings.[13]

In our analysis we focus on key entry grades for students who have not previously participated in the school choice process: pre-K, Kindergarten, and first grade. A total of NNN students applied to MMM programs in these grades, representing XXX% of the total number of applicants and YYY% of the total seats.

3.2 Study Design: Taking the Framework to the Empirical Context

With the described framework in mind, we designed and implemented an empirical study in partnership with an ed-tech NGO and the Ministry of Education of Chile. Figure 1 provides an overview of the design and timing of the implementation for our study. On the top section of the figure, and in green, we summarize the data collection instances, which will be described in the next subsection. In the middle of the figure, in blue, we present the evolution of three relevant model objects: the set of schools that the agent knows and may apply to at time t, the perception of the characteristics of those schools, and the beliefs about the characteristics of the schools that the agent does not know yet. In blue, we also described potential actions the agent may take, like searching for schools or submitting an application while the platform is open. Finally, in the bottom section of the figure, in black, we present the timing and a brief description of the two interventions we implemented.

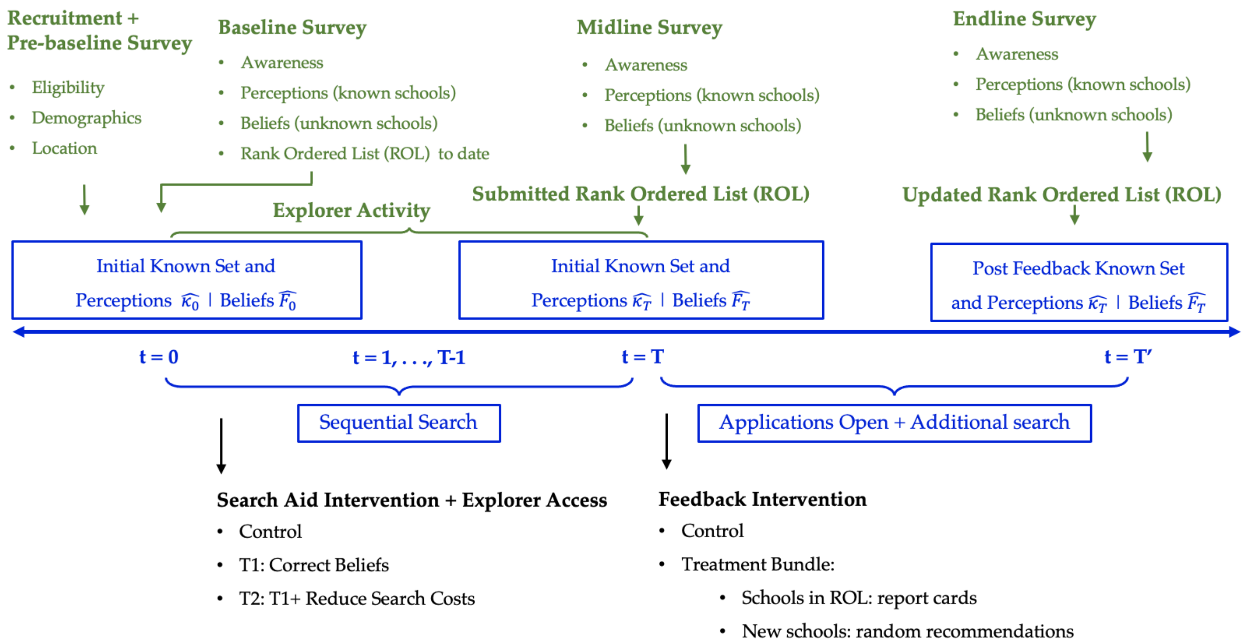
[10] The agency implements the national standardize test (SIMCE) and their rating system substitutes previous policy efforts to inform families about schools. Since 2016, it has assigned a performance category for each educational institutions recognized by the state. These categories are the result of a comprehensive evaluation that classifies establishments into high, medium, medium-low and insufficient performance

[11] When considering the Metropolitan region, the average number of schools within 2km radius increases to 21. Out of these, 16 are, on average, free and 11 have medium or high quality..

[12] 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.

[13] XXX HOW MANY HIGHLGHTED SCHOOLS DO PEOPLE NOT APPLY TO XXX

FIGURE 1. Timeline for the Model, School Choice Process, Data Collection, and Interventions



Recruitment and Collaborations: The study was embedded in a school explorer software developed by the EdTech NGO and made available to the participants through the Government website. The explorer allows parents to obtain information on schools in their neighborhood. The recruitment for the study was implemented through the Ministry of Education. Between May 25 and July 2, 2021, the government sent an email to potentially eligible parents through kindergarten principals across Chile that allowed them to sign up for the information program. [14] As described in Figure 1, we implemented a pre-baseline form to define eligibility and collect location and demographic information. To be eligible for the study, a parent needed to have a child enrolled in pre-kindergarten or kindergarten and would need to apply for the first time to an educational institution through SAE in 2021.

A total of 3,948 parents signed up for the study and completed the baseline survey. In our main analysis, we exclude 837 parents who never reached the point in the school explorer software that provides different information by treatment status, leaving us with a final sample of 3,111 parents.[15] Appendix Table A2 shows the comparison between the full pool of applicants (column 1), applicants applying for entry grades only (column 2), economically vulnerable and non vulnerable entry grade applicants (columns 3 and 4), and the experiment sample (column 5). We find that parents who selected into the experiment tend to be of higher economic status and

[14]See 9 for more details.

[15]The share of parents who did not reach the point of the platform in which the interventions differs by treatment group is 22% in the control group, 21% in treatment 1 group, and 20% in the treatment 2 group.

have longer rank-ordered lists.

3.3 Data

Figure 1 shows in green the primary sources of data that we use in this study.

Surveys: We implemented a registration form and three surveys to study participants at key points of the search and application process.

- **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 implemented months before families had to apply for schools, and we will consider it to measure objects at time $t = 0$ in our model. It was sent to eligible parents and collected a detailed list of schools that the parents know, their perceptions about the characteristics of those schools, including admission chances, and detailed elicitation of beliefs about the distribution of characteristics of schools that they do not know in the neighborhood. It also included questions about the rank-ordered list of schools that parents were planning to apply to and search behavior.
- **Midline Survey:** The midline survey was implemented once parents submitted their applications to schools were submitted. We consider this to be time $t = T$, which constitutes the final search that families made before submitting an official application. This survey was done over the phone and 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.[16]
- **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 about admission chances, and important factors in the search process.

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

[16](44% of sample parents completed the midline survey.

Administrative data: We use several sets of administrative data provided by the Government, which we are able to merge with our surveys, explorer activity and treatment assignement.

- **Application Data:** This data was provided by the Ministry of Education and included each application list submitted by parents during the application process in real-time. This data is relevant not only for the analysis but also for the design of the personalized feedback intervention that we describe below. As described in Figure 1, families may submit an application once the application process starts and can change their submitted list of schools while the platform is open at no cost. The data provided contains the rank-ordered list of schools and the application outcome.
- **Administrative and Enrollment:** This data consists of administrative data on applicants and enrollment for the 2021 school year.
- **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.

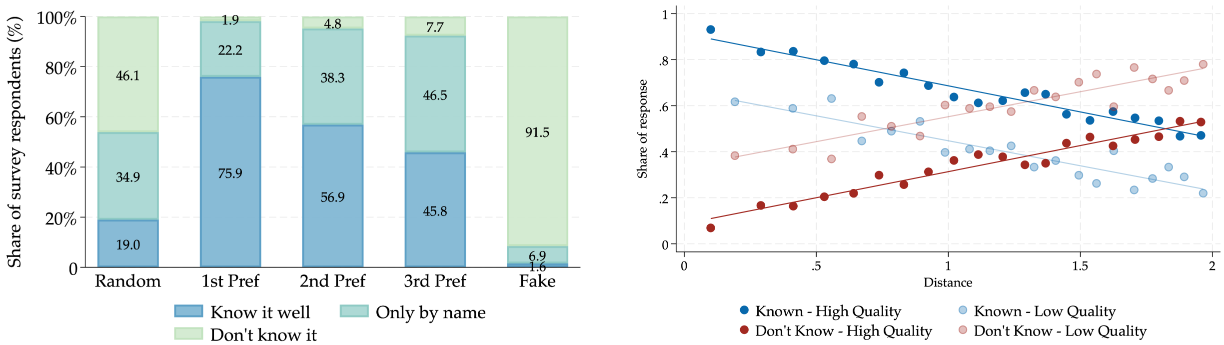
4 Descriptive Analysis

We combine our administrative and survey data to document a set of empirical patterns that are consistent with the framework described in section 2.

4.1 Knowledge, Perceptions, and Beliefs

School Knowledge: We first show that households to do not know all the schools in their neighborhood. 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.[17] Additionally, we asked during the application period the knowledge of schools on the application list. Figure 2a 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% do not 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.

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



(A) By Ranking

(B) By Distance and Quality

FIGURE 2. Knowledge of Schools

Notes: Panel (A): Stated knowledge of a random school asked in baseline, a fake school, and the first two schools on application list. 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)

Figure 2b displays the probability of knowing a school as a function of the distance from the family's house to the school. The dark blue line shows this relationship for high quality schools. Knowledge decreases sharply with distance. The light blue line shows this relationship for the low quality schools. At every distance bin, the high quality schools are considerably more likely to be known than the low quality schools.

Beliefs about the distribution of Unknown Schools: Our second result is that households hold inaccurate beliefs about the distribution of school characteristics such as quality and price in their neighborhood. We elicited parents' beliefs about the number of schools within 2 km of their homes and then asked them to allocate these schools across four quality bins and four price bins according to their beliefs. We observe from Figure 3 that households overestimate the share of high quality schools and underestimate the share of medium quality schools. Similarly, we note from Figure 3 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.

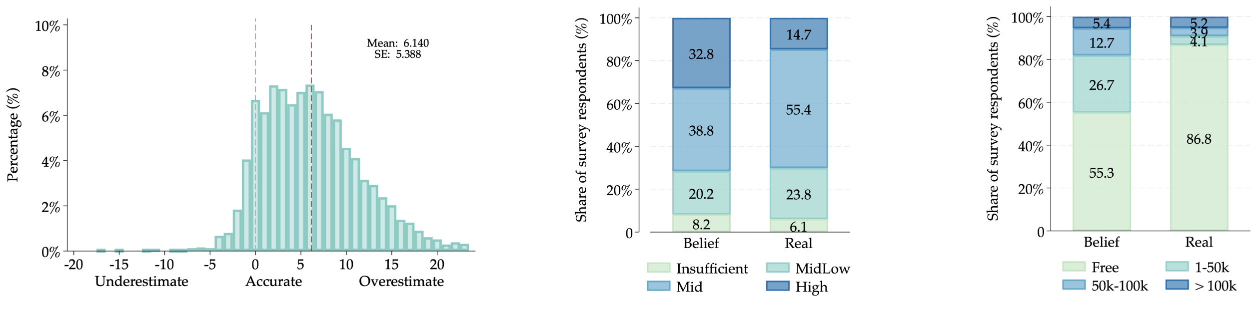
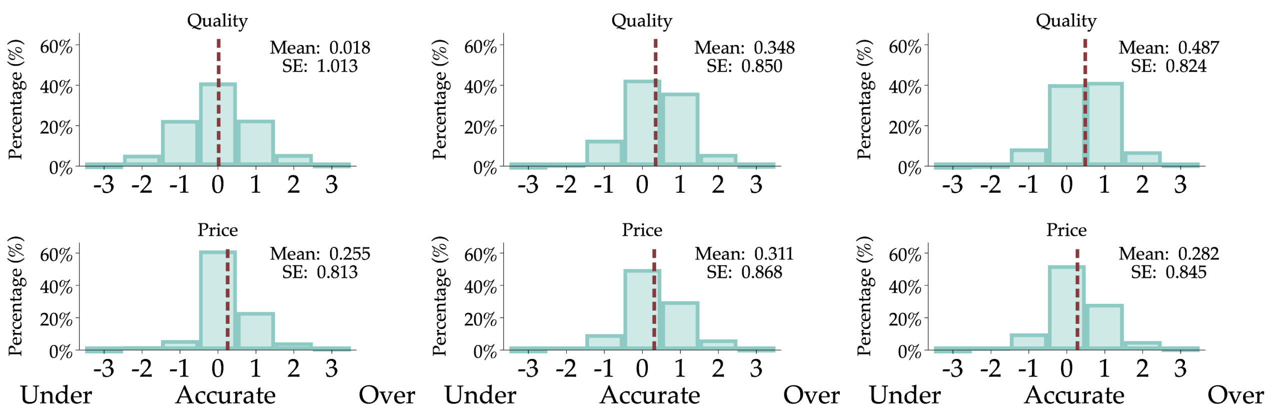


FIGURE 3. Beliefs about the Distribution of School Attributes

Notes: Panel (A) shows the bias in the beliefs for the number of schools that are in the top 2 quality categories, and are free. 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



(A) Random School Known

(B) Random School in App

(C) First Preference

FIGURE 4. Errors

Notes: Panel (A) shows the bias on perceived quality and price of a known random school asked in baseline (N = 2,523). Panel (B) shows the bias on perceived quality and price of a random school in the application list, excluding the first ranked school (N = 1,925). Panel (C) shows the bias on perceived quality and price of the first preference school at baseline (N= 3,278). 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. Source: Baseline Survey. Partial ranking based on parent's responses.

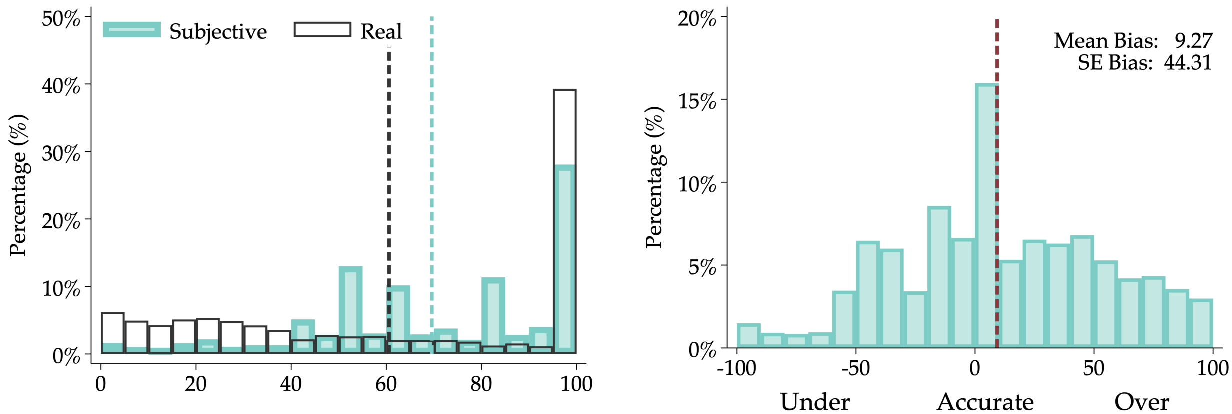
Perceptions about the characteristics of the known schools: In addition, households inaccurately perceive the characteristics of schools that they say they know. We asked parents to report the quality and price of a school they know at least by name, their first preference school, and a random school in their application list, not including the first ranked. Figure 4 shows the distribution of responses for quality and price. Households overestimate the quality of known schools. The bias is larger at schools to which households intend to apply, and especially at the most-preferred school. It is consistent with households' ranking schools with high perceived quality highly, even when these schools do not in fact have high quality scores. We find that households also overestimate the price they must pay for known schools. Because quality is a good but price is presumably a "bad", the sign of these biases' impact on search decisions will depend on the relative strength of preferences for price and quality.

TABLE 1. Effects of Perceived vs Real Characteristics 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 their perceptions as reported in our survey, not true values. Table 1 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 5. 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. Figure 5 shows that beliefs about admissions chances are upwardly biased but also exhibit compression, with households underestimating the share of schools with chances below 40% but also underestimating the share of schools at which admission is nearly certain. 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.

4.2 Search Behavior

We now use the explorer platform to analyze search behavior. Our first result is that observed effort varies with subjective 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 2 we observe that parents who believe they have a higher probability of admission to their first preference school are less likely to engage in search via the explorer. 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 2. 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	-24.943***	-4.732***	-1.590***	-4.658***	-1.587***
	(6.381)	(1.051)	(0.398)	(1.040)	(0.393)
Real Admission Chance 1st Pref	-24.662***	-4.312***	-1.370***	-4.161***	-1.339***
	(4.645)	(0.820)	(0.303)	(0.816)	(0.299)
Control Mean	47.132	9.162	3.149	7.735	2.946
Observations	2323	2323	2323	2323	2323

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 (Appendix Figure A3). Similarly, parents who report a higher probability of searching for more information on unknown schools are more likely to interact with the explorer and search for schools.

Our second result from the explorer usage data is that platform behavior affects school knowledge. Table 3 shows that when controlling for reported knowledge at baseline, parents who clicked on the school profile (i.e. they opened the details page of the school in the explorer) 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 3. 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	-0.035	-0.015		
	(0.022)	(0.053)		
Double Click	0.053**	0.108***	0.058**	0.104***
	(0.023)	(0.029)	(0.024)	(0.030)
Baseline - Knows By Name	0.414***	0.327***	0.417***	0.332***
	(0.035)	(0.038)	(0.036)	(0.039)
Baseline - Knows Well	0.489***	0.720***	0.493***	0.732***
	(0.030)	(0.030)	(0.030)	(0.030)
Control Mean	0.797	0.488	0.780	0.464
Observations	847	847	771	771

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 about the distribution of schools' characteristics. Table 4 reports 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 that people are more likely to include in their application schools they've searched in the explorer.

TABLE 4. Platform Behavior Affects Beliefs

	Perceived N of Schools (Midline)	Perceived N of Highlighted Schools (Midline)	Perceived Share of Highlighted Schools (Midline)
	(1)	(2)	(3)
Perceived Schools (Baseline)	0.241***	0.243***	0.072***
	(0.075)	(0.046)	(0.025)
Schools Clicked in Explorer	0.041*	0.050***	0.042
	(0.022)	(0.018)	(0.035)
Actual Schools	0.074***	0.035*	0.242***
	(0.018)	(0.018)	(0.067)
Control Mean	6.236	1.906	0.339
	Treat 0	Treat 0	Treat 0
Observations	547	534	516

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

Finally, we find that search history affects perceived price and quality. Table 5 presents the impacts of explorer usage and treatment on perceived price, quality, and admission chances measured at midline. We find a significant positive effect of explorer interaction the share of perceptions of schools' price and quality indices that are correct. Having investigated the school reduces the number and magnitude of errors in households' perceptions of price and quality. In contrast, it has no significant effect on the accuracy of perceptions of admission chances.

TABLE 5. Search history affects 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	-0.354***	-0.088***	0.101***	-0.268***	-0.123***	0.166***	-0.027*	0.105***	0.024
	(0.072)	(0.025)	(0.027)	(0.043)	(0.027)	(0.028)	(0.016)	(0.029)	(0.024)
Control Mean	0.789	0.306	0.609	0.617	0.364	0.520	0.350	0.425	0.196
	Treat 0	Treat 0	Treat 0	Treat 0	Treat 0	Treat 0	Treat 0	Treat 0	Treat 0
Observations	1298	1298	1298	1300	1300	1300	1085	1085	1085

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 12 and clicks and start primer.

5 Experimental Analysis

We embedded two treatment arms in the school explorer software to generate exogenous variation in beliefs about the aggregate distribution of school attributes and search costs prior to the school application. An additional treatment arm was implemented after parents submitted applications. We describe each intervention in turn and then present the treatment effects.

5.1 Intervention Design

Search Aid Intervention: The search aid intervention had two treatment arms, and was implemented through the school explorer software before parents had to submit applications. The first treatment group provided information about the availability of schools and their price and quality distribution within 2km of a respondent's home. Figure 6 shows an example. The household has 18 schools in total within 2km of the home. Seven of these schools are either free or charge a low price and there are 7 schools that are have medium or high quality. The fourth panel shows the joint distribution, indicating that there are 5 schools that are affordable and have good quality. The second treatment group received the same information but was additionally shown where these schools are located on the map (Figure 7). Both treatment groups further received a detailed table that shows the distribution of schools in each price and quality category (Figure 8). The control group also received the school explorer software but did not receive any information about the distribution of schools or their characteristics. The information was displayed when parents entered the software. After that, parents could navigate the map and click on each school to obtain additional information. For the control group and treatment 1 group, all schools were shown in the same color. By contrast, parents in treatment group 2 could directly observe the highlight - worthy schools on the map since they were shown in green with an icon indicating their price and quality (Figure 9).

Appendix Table A4 shows baseline sample characteristics for each treatment arm in the search intervention. We find no significant difference between treatment groups in any of the covariates. Appendix Tables A5 and A6 show that we are also well-balanced on covariates within the high and low SES subsamples. We also do not observe significant differences in midline survey completion rates.[18]

[18] The completion rates are 43%, 46%, and 42% for the control group, treatment 1 group, and treatment 2 group, respectively.

FIGURE 6. Search Treatment 1

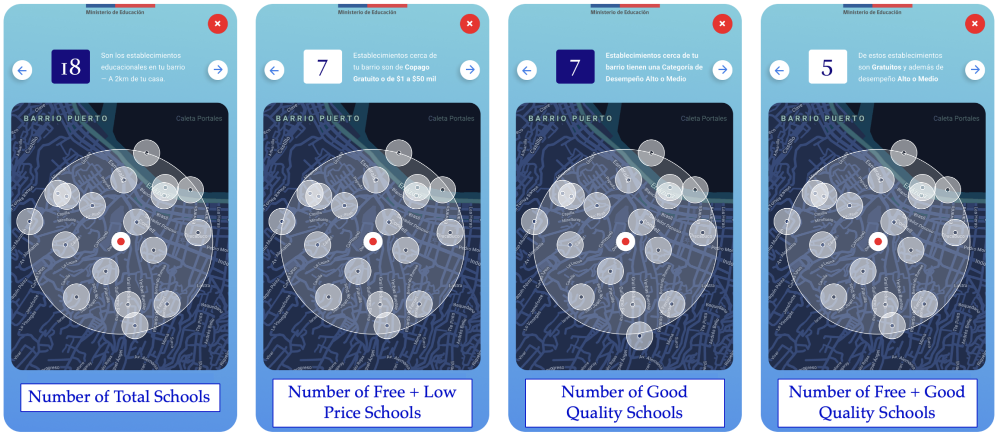


FIGURE 7. Search Treatment 2

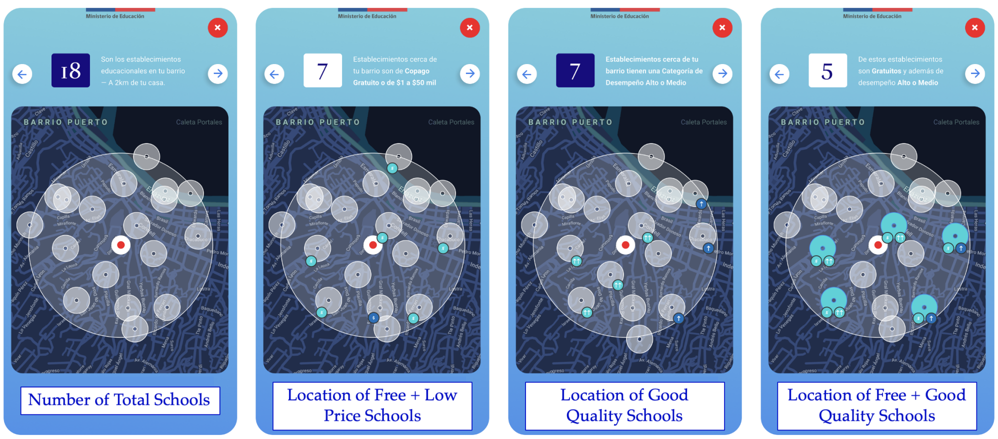


FIGURE 8. Additional Distribution Information for Treatment Groups 1 & 2

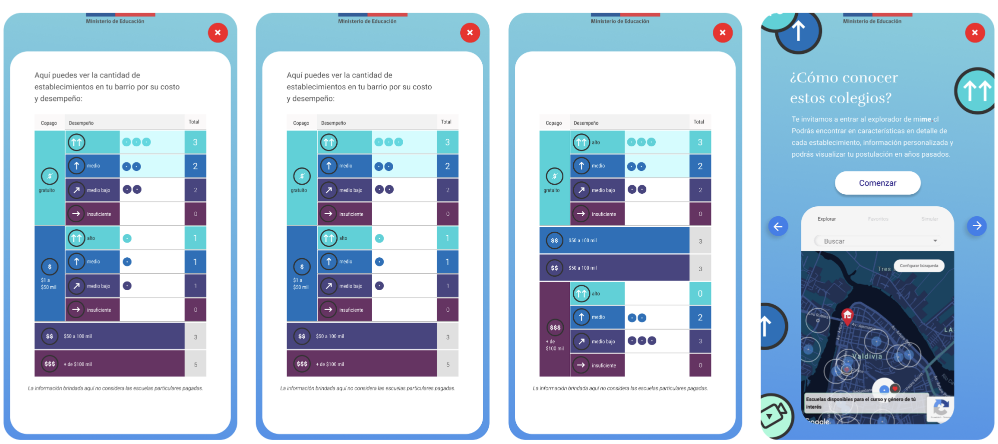
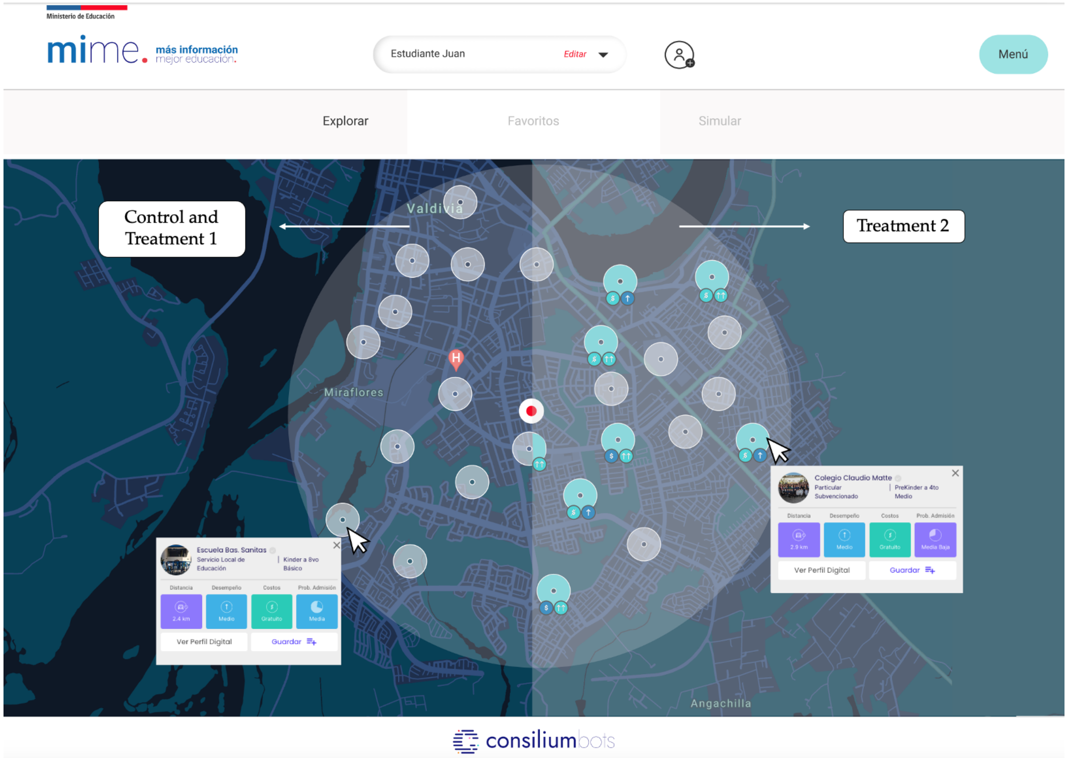
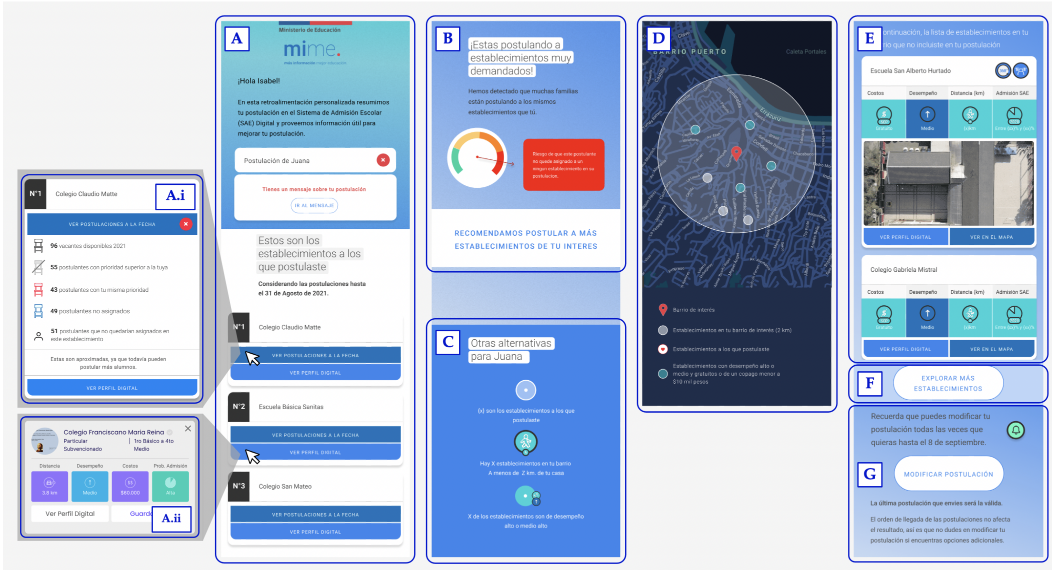


FIGURE 9. School Explorer by Treatment Status



Feedback Intervention: In addition to the search intervention, another information experiment was implemented nationwide that provided tailored feedback on the applications parents submitted (see Figure 10). Parents who are in the feedback treatment first received information on the schools that are currently included in their application (Panels A). If there was a high chance that the child would not receive any school based on the current application, a warning message was also shown to inform the parent that the application was risky (Panel B). The treatment further presented a list of alternative schools with good quality and low price in the neighborhood that were not yet included in the application. Appendix Table A7 shows the balance test corresponding to the feedback intervention. The samples are well balanced, including with respect to the previous treatment assignment in the search intervention. The only exception is that households in the feedback treatment tend to be more optimistic about their admission chances for their most preferred school at baseline.

FIGURE 10. Feedback Treatment



Notes: This figure shows the feedback treatment details. Panel (A) presents the student's current application with the option to view the applicants to date (Panel A.i) and the school characteristics (Panel A.ii). Panel (B) is a warning message if the current application is considered risky. Panel (C) and (D) presents alternatives to the current application, showing schools not included that are considered good quality and low price. Panel (E) provides a detailed view of the alternatives offered. Panel (F) invites applicant to explore more schools and Panel (G) is a link to modify the current application.

5.2 Impact of Search Treatments

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

For parent i, we estimate:

$$Y_i = \alpha + \beta_1 T1_i + \beta_2 T2_i + \theta_i + X_i + \epsilon_i. \tag{4}$$

Y_i is the outcome variable, θ_i are stratification dummies, and X_i are baseline controls selected via a double LASSO approach from Table A4 covariates. We show results for the pooled sample and separately for high and low SES households, proxied by whether the mother completed college.

Our first result is that treatment affects beliefs in the midline survey, where households assigned to treatment interventions 1 and 2 perceive a higher number of total schools as well as a a higher number of highlight-worthy schools in their neighborhood (See Table 6, Panel A, columns 1-2). Relative to their control group counterparts, households in treatment group 1 believe that there are 23% more highlight - worthy schools in their neighborhood. The treatment effects on the perceived number of total schools are concentrated among high SES households (Panel B).

TABLE 6. Treatment Effects of Search Intervention

	Perceived Number of Schools		Number of Profile Clicks		Number of Schools Known	Enrolled School		
	All	Highlighted	All	Highlighted	At Least by Name	Highlighted	Value Added	Distance
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Pooled								
Treatment 1	0.833***	0.467***	0.062	-0.006	0.240	-0.003	0.007	-0.072
	(0.313)	(0.128)	(0.161)	(0.082)	(0.211)	(0.021)	(0.020)	(0.049)
Treatment 2	0.632*	0.359***	0.018	0.051	-0.017	-0.017	-0.008	-0.056
	(0.326)	(0.131)	(0.160)	(0.084)	(0.204)	(0.021)	(0.019)	(0.049)
Control Group Mean	6.278	1.931	2.219	1.034	3.727	0.609	0.171	1.215
Observations	1671	1633	3111	3111	1076	2448	2370	2463
Panel B: Heterogeneity by Parental Education								
Treatment 1 × High SES	2.291***	0.545**	0.768**	0.358*	1.543***	0.018	0.080**	0.077
	(0.682)	(0.257)	(0.381)	(0.191)	(0.470)	(0.044)	(0.039)	(0.102)
Treatment 1 × Low SES	0.389	0.440***	-0.183	-0.127	-0.144	-0.010	-0.020	-0.122**
	(0.350)	(0.150)	(0.174)	(0.089)	(0.235)	(0.024)	(0.022)	(0.055)
Treatment 2 × High SES	1.613**	0.525*	-0.082	0.048	1.018**	0.023	0.030	-0.085
	(0.709)	(0.286)	(0.382)	(0.197)	(0.421)	(0.045)	(0.041)	(0.103)
Treatment 2 × Low SES	0.386	0.310**	0.047	0.050	-0.301	-0.033	-0.020	-0.051
	(0.369)	(0.148)	(0.174)	(0.092)	(0.234)	(0.023)	(0.022)	(0.056)
p-value: Treat 1 x High SES = Treat 1 x Low SES	0.014	0.727	0.024	0.022	0.001	0.579	0.028	0.086
p-value: Treat 2 x High SES = Treat 2 x Low SES	0.126	0.506	0.758	0.989	0.006	0.270	0.290	0.767
Control Group Mean (High SES)	6.110	1.800	2.802	1.116	3.386	0.474	0.194	1.249
Control Group Mean (Low SES)	6.324	1.967	2.043	1.010	3.838	0.652	0.165	1.206
Observations 1 (High SES)	362	357	732	732	246	587	564	568
Observations 2 (Low SES)	1308	1275	2376	2376	829	1859	1804	1893

Notes. This table presents the results of the search interventions on beliefs (Columns 1-2), search (Columns 3-4), knowledge (Column 5), and final school enrolment (Columns 6-8). In Panel A, we regress each outcome on indicator variables for both treatment arms, stratification dummies and baseline controls selected by LASSO. In Panel B, we further include the fully interacted effects of treatments and SES status. SES status is proxied by whether the mother completed college.

The updated beliefs affected the search behavior of parents (columns 3-4). While we find limited effects for the pooled sample, we observe substantial increases in the number of school profile clicks among high SES households in the first treatment group.[19] They click on 27% more school profiles than high SES households in the control group. We can reject that the treatment effects are the same for high and low SES households. We note that we do not have clear predictions on how the second treatment affects overall search effort. Since parents in treatment group 2 observe which schools in their neighorhood are highlight - worthy, they might only focus on these schools, which could reduce the number of total clicks.

Consistent with increased search, we also observe knowledge gains in the midline survey (column 5). High SES households in treatment group 1 report that they know 46% more schools at least by name. We also find positive but insignificant effects on the number of schools they reporting knowing well (c). By contrast, we find null to negative effects for low SES households.

Finally, we also examine the effects of the search interventions on school enrollment (columns

[19]Appendix Table A10 shows that we find similar treatment effects for school pin clicks.

6-8). We again find no effects on the pooled sample (Panel A). However, among high SES households, we find that the first treatment arm leads to a significant increase in the average value added of the enrolled school. We also find that low-SES households in treatment group 1 enroll in schools closer to their home. It is surprising that we find positive but insignificant effects for the second treatment arm, in which parents received additional information on highlight - worthy schools. A potential explanation is that the treatment mostly affected the characteristics of schools that were ranked lower in the application, making it less likely to affect final enrollment. Appendix Table A11 shows the effects of different application outcomes and we indeed find that the second treatment arm increases the likelihood that the second ranked schools is highlight - worthy. Relative to control group households, treatment 2 group households are 9% more likely to list a highlight - worthy school in rank 2. We further find that both treatments decrease the likelihood that the schools in their application are known at baseline.

5.3 Impact of Feedback Treatment

We next examine the impact of the feedback intervention. For parent i, we estimate:

$$Y_i = \alpha + \beta T_i + \lambda_i + \gamma X_i + \varepsilon_i.$$

Y_i is the outcome variable, θ_i are stratification dummies, and X_i are baseline controls. In our main specification, we use the the treatment assignment as an instrument for opening the feedback intervention.

Table 7 shows how the feedback intervention affects application behavior. We find that parents who received the feedback information are 10.2 percentage points more likely to change their application (Column 1). They are more likely to add a school, showing that the information helped to increase search. We also find that treatment group parents are more likely to delete schools, suggesting that parents previously had incorrect perceptions of the attributes of some schools in their application. Panel B shows that the effects tend to be larger for low SES households. We further find that the feedback treatment affects assignments and knowledge (Appendix Tables A13 and A14).

TABLE 7. Feedback treatment affects applications

	Change Application	Add School	Add School Recc.	Add School Highlighted	Delete School	Δ % Risk
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Pooled						
Open Feedback	0.102**	0.065**	0.047**	0.055**	0.022*	-0.002
	(0.019)	(0.017)	(0.017)	(0.015)	(0.012)	(0.005)
Control Group Mean	0.028	0.026	0.021	0.018	0.009	0.024
Observations	2116	2116	2116	2116	2116	2116
Panel B: Heterogeneity by Parental Education						
Open Feedback $\times$ High SES	0.029	0.023	0.009	0.029	0.000	-0.003
	(0.033)	(0.031)	(0.029)	(0.019)	(0.017)	(0.009)
Open Feedback $\times$ Low SES	0.124**	0.078**	0.060**	0.063**	0.029*	-0.002
	(0.024)	(0.022)	(0.020)	(0.019)	(0.014)	(0.005)
p-value: Open Feedback x High SES = Open Feedback x Low SES	0.024	0.151	0.132	0.169	0.202	0.952
Control Group Mean (High SES)	0.037	0.029	0.021	0.008	0.008	0.036
Control Group Mean (Low SES)	0.026	0.025	0.021	0.021	0.009	0.021
Observations 1 (High SES)	460	460	460	460	460	460
Observations 2 (Low SES)	1656	1656	1656	1656	1656	1656

Notes. This table presents the results of the feedback interventions on application behavior. SES status is proxied by whether the mother completed college.

6 Model

We now present 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 a subset of the schools near their house. To learn about additional schools, households may 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 features of "known" schools and the distribution of not-yet-searched schools in their neighborhood. We allow households to misperceive the rejection chances and payoff-relevant characteristics—both observed and unobserved by the econometrician —of known schools, and to hold inaccurate beliefs about the distribution of unknown schools' observed and unobserved characteristics.

6.1 Information, Preferences, and Application Portfolios

Let I denote the set of households, and J the set of schools. Household i's choice set, $J_i \subset J$, consists of all schools within five kilometers of i's house that offer a seat in i's grade level and for which i is eligible.[20] Time is discrete: $t = 0, 1, \ldots, T$. At time $t = 0$, our study begins. Applications are due at time T.

[20] Ineligibility for seats in the relevant grade level is rare. (XXX NUMBER XXX) Empirically, single-sex schools are the main source of ineligibility.

Information: For each student i and each school $j \in J_i$, i's level of knowledge of j at time t is given by $\pi_{ijt} \in \mathbb{R}$. Let u_{ij} and r_{ij} be "objective" payoffs and rejection chances faced by student i given all the information that could in principle be known at time T.[21] We consider a family of potential subjective expectations of these objects, indexed by the available amount of information. At time t, student i's subjective expected utility from a placement in j is given by $\hat{u}_{ijt} = \hat{u}_{ij}(\pi_{ijt})$, and i believes that, in the event the school receives his application, it will reject him with probability $\hat{r}_{ijt} = \hat{r}_{ij}(\pi_{ijt})$.

We impose the following parametric structure on π:

$$\pi_{ijt} = z_{ij}\alpha^z + w_{ijt}\alpha^w + w^{rc}_{ijt}\alpha^{rc}_i + \eta_j + \nu_{ijt}. \tag{5}$$

In practice, z_{ij} consists of the distance from i's residence to j's campus. The terms w_{ijt} are potentially time-varying "knowledge shifters" that will be excluded from preferences. In practice, w_{ijt} consists of three types of terms: (1) indicators for having received information treatments in current or prior periods; (2) indicators for having searched school j in the explorer software in current or prior periods; and (3) time indicators $1(t > s)$ for a set of periods s. The first set of terms capture variation induced by our experiments. The second set capture the effects of endogenous on-platform search effort. Time indicators proxy for the arrival of exogenous off-platform information over time, so that households may be systematically more knowledgable at time T than at time $t = 0$ even without our information treatments or search technology.

We place random coefficients on the time indicators: $w^{rc}_{ijt} \sim N(0, \Sigma^{\pi rc})$. The unrestricted covariance matrix allows variation in the timing of off-platform search. Households that learn a lot between times $t = 1$ and $t = 2$ may learn less between $t = 2$ and $t = 3$, for instance. The remaining terms are a school-level "discoverability" η_j and a shock ν_{ijt}. The shocks are independent across schools and households, but may be correlated over time: $\nu_{ijt} \sim N(0, \Sigma^{\nu})$.

Preferences and Rejection Chances: For the purposes of estimation, we allow three levels of knowledge with associated potential utilities.[22] The expected payoff from placing in school j, given i's information at time t, is:

$$\hat{u}_{ij}(\pi_{ijt}) = 1(\pi_{ijt} \geq 1)u^h_{ij} + 1(0 \leq \pi_{ijt} < 1)u^l_{ij} + 1(\pi_{ijt} < 0)u^0, \tag{6}$$

[21] In our setting, ties are broken by random lotteries. Because lottery realizations are not known when applications are due, not all uncertainty can be resolved.

[22] This structure is designed to match our survey questions.

where u_{ij}^h is a "high-information" subjective expected utility, u_{ij}^l is the subjective expected utility when the household has low information about j, and $u^0 < 0$.[23] If $\pi_{ijt} > 1$, the household knows j well. If $\pi_{ijt} < 0$, the household does not know j well enough to apply to it. We make the following parametric assumptions on payoffs:

$$u_{ij}^h(\hat{x}_{ij}^h) = z_{ij}\beta^z + \hat{x}_{ij}^{h,rc}\beta_i^x + \delta_j + \hat{x}_{ij}^h\gamma + \varepsilon_{ij} \tag{7}$$

$$u_{ij}^l(\hat{x}_{ij}^l) = z_{ij}\beta^z + \hat{x}_{ij}^{l,rc}\beta_i^x + \delta_j + \hat{x}_{ij}^l\gamma + \hat{E}(\varepsilon_{ij}|\tilde{\varepsilon}_{ij}) \tag{8}$$

$$\hat{x}_{ij}^l \sim \Gamma(\cdot|x_j) \;\; \hat{x}_{ij}^h = (x_j \text{ w.p. } p^h, \text{ otherwise } \hat{x}_{ij}^l). \tag{9}$$

The variable $\hat{x}_{ij}^l \in \{1,2,3,4\}^2$ denotes the household's perception of the school's price and quality category in the event $0 < \pi_{ijt} < 1$. It is drawn from a multinomial distribution, $\Gamma(\cdot|x_j)$, which depends arbitrarily on the true value x_j. When the household knows the school "well" it learns the truth with probability p^h. Otherwise, it does not update about x.

We place normal random coefficients on $\hat{x}$ and distance, with zero mean and arbitrary correlation matrix Σ^{rc}. As a normalization, we fix the mean coefficient β^z on distance to -1.

We model schools' mean utility and "discoverability" as correlated 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.

The "match value" shocks ε_{ij} are normally distributed, iid across people and schools. In the event $0 \leq \pi_{ijt} < 1$, the household observes a noisy measurement of this shock, $\tilde{\varepsilon}_{ij}$, and forms a subjective expectation of ε_{ij} given this signal. In estimating preferences, we assume

$$\begin{pmatrix} \hat{E}(\varepsilon_{ij}|\tilde{\varepsilon}_{ij}) \\ \varepsilon_{ij} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu^l \\ 0 \end{pmatrix}, \Sigma^{\epsilon}\right).$$

This is a reduced form of a model of Bayesian updating with a misspecifed prior and signal precision, which we present in the following section. It accomodates the case in which households observe ε with classical Gaussian measurement error and form the correct Bayesian posterior mean. It also allows households to be pessimistic or optimistic about schools that they do not know well, and to fail to appropriately shrink their signals toward the correct prior.

Subjective admissions chances are given by

$$\hat{r}_{ij}(\pi_{ijt}) = \min\{1, \max\{0, o_{i0} + o_{i1}(r_{ij} - o_{i0})\}\},$$

where $(o_{i0}, o_{i1}) \sim N(\mu_o, \Sigma_o)$. This specification allows for optimism or pessimism (via the random

[23] Alternatively, one may assume that j is not known when $\pi_{ijt} < 0$.

intercept o_{i0}) and compression (via o_{i1}). Consistent with our descriptive evidence, admissions beliefs do not become more accurate with higher π_{ijt} or greater search activity.

There is an outside option, not receiving a placement, which has $\hat{r}_{i0} = r_{i0} = 0$ and $\hat{u}_{i0t} = u_{i0} = 0$ for all t.

Application Portfolios: In the event student i submits a rank-ordered application at time t, then the student's choice set is

$$\hat{\kappa}_{it} = \{(\hat{u}_{ij}(\pi_{ijt}), \hat{r}_{ijt}) : j \in J_{it} \cup 0 \text{ and } \min(\hat{u}_{ij}(\pi_{ijt}), \pi_{ijt}) \geq 0\} \tag{10}$$

As in our benchmark, the optimal portfolio ranks elements of $\hat{\kappa}_{it}$ in descending order of $\hat{u}$, and the subjective expected utility of this portfolio is given by:

$$\hat{U}(\hat{\kappa}_{it}) = \sum_{j=1}^{|\hat{\kappa}_{it}|} \left(\prod_{k=1}^{j-1} \hat{r}_{ijt} \right) (1 - \hat{r}_{ijt}) \hat{u}_{ijt}. \tag{11}$$

Knowledge π_{ijt} affects the optimal portfolio and its value (equation (11)) through two channels. First, if $\pi_{ijt} < 0$ then i cannot apply to j at time t. Second, larger values of π_{ijt} indicate better knowledge of payoffs.

6.2 Timing and States

Timing: In practice, we allow four periods, $t = \{0, 1, 2, 3\}$. At time $t = 0$, our baseline survey takes place. At time $t = 1$, households are assigned to search treatments and have the opportunity to engage in on-platform search via our software. Within this period, households first receive treatments, causing their state to update. They then engage in endogenous search, potentially leading to further updates in π and other beliefs. When they choose to stop, the period ends. At time $t = 2$ ("just before feedback") household may submit initial applications to the official platform. At the beginning of period $t = T = 3$, "treated" households receive the feedback treatment. After this, final applications are submitted. Finally, the endline survey takes place.

The relative timing of the midline survey and feedback treatment varied across households. For households who were given the midline survey before they received—or would have received—the feedback treatment, the midline survey takes place at $t = 2$; otherwise it takes place at $t = 3$. Importantly, while off-platform learning (captured by changes in η and by deterministic and random coefficients on time indicators) may occur between any two periods, on-platform search takes place only at $t = 1$, consistent with descriptive evidence that most explorer use takes place at this time. In evaluating the returns to search effort, households at this time expect to submit applications at the end of period 1. Future learning and interventions including "feedback" come as a surprise.

An agent's state at the beginning of period t is given by

$$\Omega_{it} = (\{\hat{u}_{ij}(\pi_{ijt}), \hat{r}_{ij}, \pi_{ijt} : j \in J_i\}, \theta_i, w_{it}, \omega_{it}),$$

where θ_i are fixed parameters relevant for preferences—in practice, the agent's random coefficients—and ω_{it} are the parameters of the subjective distributions and models held by the agent which are relevant for making forecasts.

At time $t = 0$, agents are endowed with potential utilities and perceptions $(\hat{u}^l, \hat{u}^h, \hat{x}^l, \hat{x}^h, \hat{r}_{ij})$, as well as an initial level of information π_{ij0} for each $j \in J_i$, and beliefs ω_{i0} relevant for the search decision. Agents observe $\hat{r}_{ij}$ and $\hat{u}_{ij}(\pi_{ij0})$ for each $j \in J_i$ with $\pi_{ij0} > 0$.

The initial state ω_{i0} consists of admissions-belief parameters (o_{i0}, o_{i1}), initial x-belief parameters $(\lambda_{i0}, \Lambda_{i0})$, and search costs c_{i0}.

6.3 Search

Search decisions consist of a sequence of "detail views" of schools $j \in J_i$, i.e. the decision to double-click a pin in the explorer and read about a school. After each such view, the household chooses whether to continue or stop searching. We model the sequence of decisions via a "one-period lookahead" heuristic. That is, households form beliefs over the choice set $\hat{\kappa}'$ that they would obtain with one additional "detail view," given their current choice set $\hat{\kappa}$ as defined in equation 10 and the distribution $\hat{F}$ from which they believe they are sampling. As in the benchmark, they continue if and only if $\hat{E}(U(\hat{\kappa}')|\hat{\kappa}, \hat{F}) - U(\hat{\kappa}) > c$, where $\hat{E}$ is subjective expectations and c is the cost of the next draw. If they search, and happen to inspect school j, this causes π_{ijt} to update. We provide details below.

Search costs: If a household has done s prior "detail views", the cost of the $s + 1$th draw is given by:

$$c_{is} = c_{i0} + kt + \epsilon^c_{it},$$

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

Search technology: If a household searches, the next clicked school is j with probability $Pr_{ij} = Pr(\text{view } j|\text{continue}) \propto \exp(x_{ij}^{\text{click}}\gamma^{\text{click}})$. That is, certain schools (nearby schools; highlighted schools) may be found more easily. We do not endogenize households' choice of where to search, but we allow our information treatments to affect the chances of finding certain schools. Households may revisit already-known schools. The variables x_{ij}^{click} consist of schools' price, quality, distance, and indicators for treatment=2 and being highlighted in that treatment.

Recall that inspecting school j causes information to update: $\pi_{ijt} \rightarrow \pi'_{ijt} = \pi_{ijt} + \beta^{w}_{\text{search}}$. Let $\hat{F}_{ijt}(\cdot;\pi_{ijt},\omega_{it})$ denote the subjective distribution of payoffs $\hat{u}_{ij}(\pi'_{ijt},\hat{r}_{ij})$ conditional on double-clicking the map pin corresponding to school j. Let $\hat{F}_i$ be a mixture over the $\hat{F}_{ij}$ distributions with weights Pr_{ij}. The value of an additional search is:

$$\hat{V}(\hat{\kappa}_{it},\pi_{it},\hat{F}_{it}) = \sum j \in J_i Pr_{ij}\hat{V}_{ijt}(\hat{\kappa}_{it},\pi_{ijt},\hat{F}_{ijt}).$$

The analysis now divides into cases. Search is relevant when it causes an unknown school to become known, or when it causes a school with $\pi_{ijt} < 1$ to have $\pi'_{ijt} > 1$. In other cases (e.g. when $\pi_{ijt} > 1$) the value of search $\hat{V}_{ijt}$ is zero.

If $0 > \pi_{ijt} > 1$ but $pi_{ijt} + \beta^{w}_{\text{search}} > 1$, then search will result in knowing j well. The value of this is: $\hat{V}_{ijt}(\hat{\kappa},\hat{F}_{ijt}) = \hat{E}U(\kappa \cup \{(u^{h}_{ij},\hat{r}_ij)\} \setminus \{u^{l}_{ij},\hat{r}_ij\})$, where the subjective expectation is over the match-value shock ε_{ij} given the agent's signal $\tilde{\varepsilon}_{ij}$. Agents are naive about updates to perceived x's in this case: they assume that $\hat{x}^{h}_{ij} = \hat{x}^{l}_{ij}$.

In the event $\pi_{ijt} < 0$ but $pi'_{ijt} = \pi_{ijt} + \beta^{w}_{\text{search}} > 0$, then search results in discovering a new school. we have $\hat{V}_{ijt}(\hat{\kappa},\pi_{ijt},\hat{F}_{ijt}) = \hat{E}_{\hat{F}}U(\kappa \cup (\hat{u}_{ij}(\pi'_{ijt}),\hat{r}_{ij}))$, where the subjective expectation is over the admissions chance r_{ij}, the characteristics $\hat{x}_{ij}$, and the perceived match-value term ε_{ij} (if $\pi'_{ijt} > 1$) or belief $\hat{E}(\varepsilon_{ij}|\tilde{\varepsilon}_{ij})$ (if $0 < \pi'_{ijt} < 1$). In this case, the agent believes that $\hat{x}_{ij}$ will be drawn from some $\hat{F}^{x}_{ijt}(\cdot)$ not necessarily equal to the distribution over $\hat{x}^{l}$.

Beliefs about x's of unknown schools: Beliefs about the characteristics of unknown ($\pi_{ijt} < 0$) schools at time t are $x \sim \hat{F}^{x}_{ijt}(\cdot) = \text{Multinomial}(\lambda_{it})$, 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:

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

Beliefs about match quality: If $0 < \pi_{ijt} < 1$, household i observes ε_{ij} with classical measurement error $e_{ij} \sim N(0,\tilde{\sigma}^2)$, and forms a subjective expectation $\hat{E}(\varepsilon_{ij}|\tilde{\varepsilon}_{ij})$. Objectively,

$$\begin{pmatrix} \epsilon_{ij} + e_{ij} \\ \epsilon_{ij} \end{pmatrix} \sim N\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{\varepsilon} + \sigma^2_{e} & \sigma^2_{\varepsilon} \\ \sigma^2_{\varepsilon} & \sigma^2_{\varepsilon} \end{pmatrix}.$$

However, households believe that

$$\begin{pmatrix} \epsilon_{ij} + e_{ij} \\ \epsilon_{ij} \end{pmatrix} \sim N \begin{pmatrix} \overline{\beta} \\ \overline{\beta} \end{pmatrix}, \begin{pmatrix} \tilde{\sigma}_{\varepsilon}^2 + \tilde{\sigma}_e^2 & \tilde{\sigma}_{\varepsilon}^2 \\ \tilde{\sigma}_{\varepsilon}^2 & \tilde{\sigma}_{\varepsilon}^2 \end{pmatrix}.$$

That is, $\overline{\beta}$ represents the subjective mean unobserved match-value shock at unknown schools. If $\overline{\beta} < 0$, households are pessimistic about their payoffs at unknown schools. Households may also hold inaccurate beliefs about the prior variance or the informativeness of the signal.

A household forms a subjective mean,

$$\hat{E}(\varepsilon_{ij}|\varepsilon_{ij} + e_{ij}) = \overline{\beta} + \frac{\tilde{\sigma}_{\varepsilon}^2}{\tilde{\sigma}_{\varepsilon}^2 + \tilde{\sigma}_e^2}(\varepsilon_{ij} + e_{ij} - \overline{\beta}). \tag{12}$$

Plugging in the subjective distribution $\varepsilon_{ij} + e_{ij} \sim N(\overline{\beta}, \tilde{\sigma}_{\varepsilon}^2 + \tilde{\sigma}_e^2)$, the value of an unknown school that will be known "by name" (i.e. with $0 < \pi_{ijt}^l < 1$) is distributed, from the household's perspective, according to a $N(\overline{\beta}, \frac{\tilde{\sigma}_{\varepsilon}^2}{\sqrt{\tilde{\sigma}_{\varepsilon}^2 + \tilde{\sigma}_e^2}})$ distribution. This is the distribution from which the household believes it is drawing when finding a school with $\pi_{ijt} < 0$.

In the event $0 < \pi_{ijt} < 1$, and when evaluating u_{ij}^l, the subjective distribution of $\varepsilon_{ij}|\varepsilon_{ij} + e_{ij}$ is given by equation 12. From the analyst's point of view, plugging in the objective joint distribution, we have

$$\begin{pmatrix} E(\varepsilon_{ij}|\varepsilon_{ij} + e_{ij} \\ \varepsilon_{ij} \end{pmatrix} \sim N \left(\begin{pmatrix} \overline{\beta} \frac{\sigma_e^2}{\sigma_{\varepsilon}^2 + \sigma_e^2} \\ 0 \end{pmatrix}, \Sigma^{\varepsilon} \right).$$

These objects relate to the data as follows. A Bayesian agent will shrink the noisy measures of match value that obtained when $0 < \pi_{ijt} < 1$ toward their mean, resulting in compression and in fewer very high values of u^l. This will result in a penalty for schools known "by name" in applicants' rankings, provided that the mean is negative, and in less dispersion in utilities, conditional on observed characteristics, for schools known "by name" than for those schools that are known well. Thus we should predict the rankings of schools known "by name" to be lower than those of schools known "well", but to be better predicted by observables.

If $\overline{\beta} < 0$, the household will systematically penalize schools with $\pi_{ijt} < 1$ by a greater amount than would be expected. The penalty depends on both the mean $\overline{\beta}$ and on the subjective correlation parameters. If the household overestimates the precision of its noisy signal, then it will "shrink" less, leading to a smaller reduction in the variance of utilities than would be expected.

Effects of information treatments: We model treatments 1 and 2 as having direct impacts on belief parameters Λ. In addition, treatment 2 affects the search technology via Pr_{ij}, raising the probability of finding “highlighted” schools.

7 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 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.

8 Results

Table 8 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 perceived 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 13 and Tables 9 and 10 provide a summary of the counterfactual simulations in which we provide information at scale. For the tables, we estimate the model separately by mother's education. We look at the following counterfactuals:

I. Baseline: as in the data.

II. Learn x Late: Provide accurate info about price and quality (i.e. $\hat{x} := x$) of known schools—those with $\pi_{ijT} > 0$—at time T, taking awareness (π_{ijT}) as given.

IV. Learn (x,r) Early: Provide accurate info to everyone about the distribution of x's, x of known schools, and admissions chances, at the beginning of period t = 1. When households discover a new school j, they observe the observable characteristics x_j without error. This is an ideal version of our "search" interventions.

V. Full Info Benchmark: $\hat{x} = x$ and $\pi_{ijt} > 1$ for all $j \in J_i$.

VI. Baseline (x^ = x): we estimate the model assuming $\hat{x}_{ijt} = x_{ij}$ for all (i, j, t).

VII. Full Info ($\hat{x} = x$:) we simulate counterfactual V. within the misspecified model that assumes $\hat{x} = x$.

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:).[24]

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 13 and Tables 9 and 10 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

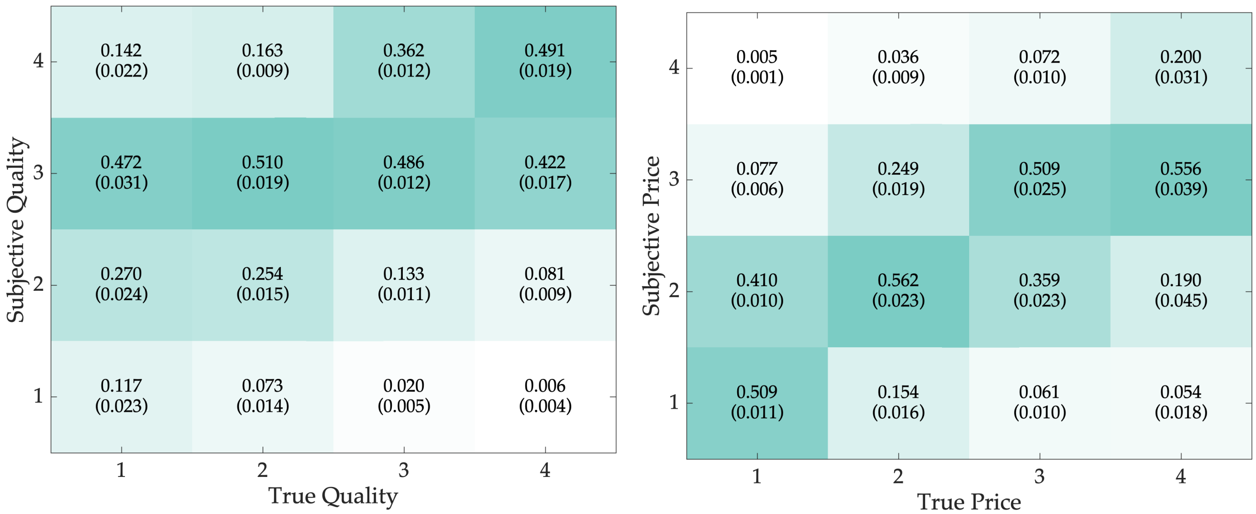
[24] Quality indices range from 1 to 4. At baseline, the average quality of the placed school is 3.069, while in this counterfactual it is 3.228, and in our full-information benchmark it is 3.247. The effect is thus 89% = (3.228 - 3.069)/(3.247 - 3.069).

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.

TABLE 8. Model Estimates

		Low SES		High SES	
	Parameter	Coeff	Std Err.	Coeff	Std Err.
Panel A: Preferences and Information					
Perceived Price	γ_p	-0.439	(0.040)	-0.369	(0.052)
Perceived Quality	γ_q	0.631	(0.039)	0.687	(0.085)
Know by Name Penalty	β_z	-0.380	(0.040)	-0.577	(0.092)
Constant	$\bar{\beta}$	-2.346	(0.225)	-3.884	(0.401)
Price	β_p	0.151	(0.026)	0.657	(0.071)
Quality	β_q	0.355	(0.031)	0.440	(0.052)
Constant	$\bar{\alpha}$	-0.289	(0.071)	-0.659	(0.092)
Price	α_p	0.119	(0.035)	0.282	(0.035)
Quality	α_q	0.142	(0.029)	0.137	(0.032)
Measurement Err Baseline Survey Utility	σ^2_ϵ	0.121	(0.011)	0.026	(0.002)
Measurement Err Survey Awareness	$\sigma^2_{\eta_s}$	0.298	(0.031)	0.017	(0.003)
Pr(learn true x's | know well)	p^h	0.036	(0.005)	0.278	(0.032)
Pr(answer some garbage | surveyed about x's)	p^s	0.221	(0.014)	0.213	(0.027)
Panel B: Variance Covariance of Random Coefficient					
Variances					
Constant	$\sigma_{constant}$	16.791	(1.306)	15.805	(1.402)
Distance	σ_d	0.524	(0.028)	0.653	(0.045)
Price	σ_p	1.701	(0.192)	1.285	(0.191)
Quality	σ_q	0.697	(0.059)	0.693	(0.105)
Covariances					
Constant - Distance	$\sigma_{constant,d}$	-1.378	(0.097)	-2.044	(0.190)
Constant - Price	$\sigma_{constant,p}$	-3.517	(0.339)	-2.550	(0.480)
Constant - Quality	$\sigma_{constant,q}$	-2.733	(0.230)	-2.507	(0.359)
Distance - Price	$\sigma_{d,p}$	0.168	(0.025)	0.079	(0.058)
Distance - Quality	$\sigma_{d,q}$	0.097	(0.025)	0.248	(0.071)
Price - Quality	$\sigma_{p,q}$	0.193	(0.045)	0.057	(0.097)
Panel C: Variance Covariance of δ and η					
Variances					
Mean Utility	σ_δ	0.507	(0.057)	1.031	(0.123)
Discoverability	σ_η	0.508	(0.025)	0.435	(0.038)
Covariances					
Mean Utility - Discoverability	$\sigma_{\delta,\eta}$	0.259	(0.015)	0.289	(0.030)

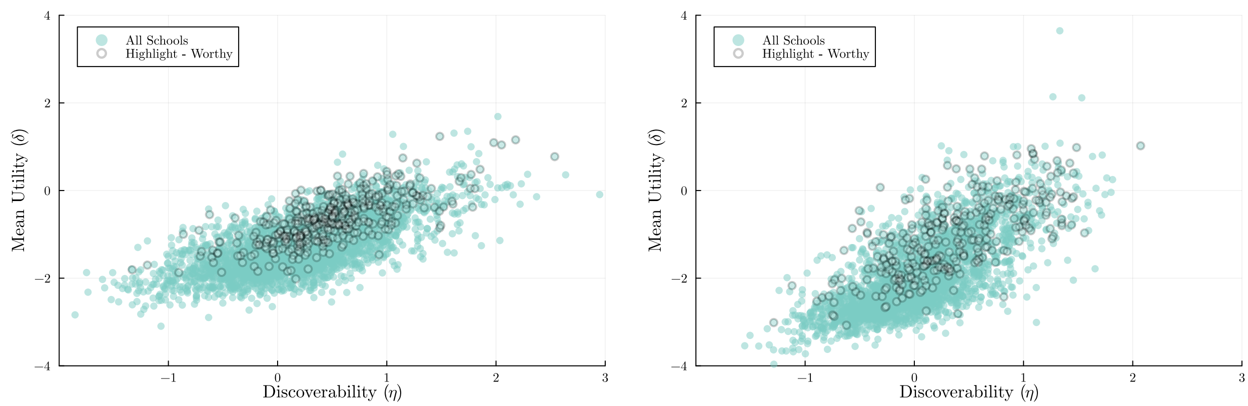
Notes. This table presents presents results from the model estimation.



(A) Quality (B) Price

FIGURE 11. Distortion Function

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



(A) Low SES (B) High SES

FIGURE 12. Estimates School Unobservables

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

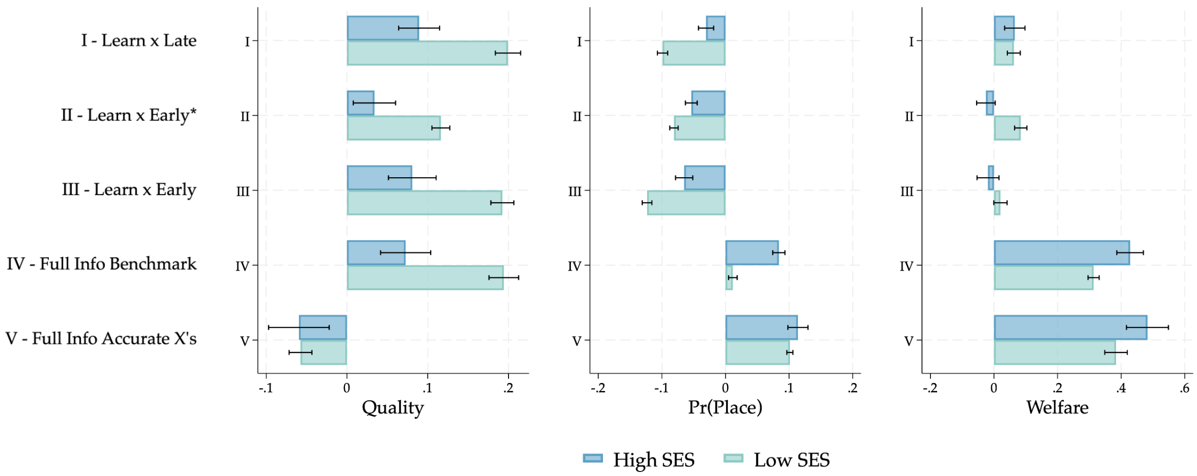


FIGURE 13. Counterfactuals

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 9. Main Results: Non-College Mothers

	EU	Place	E(rank)	Distance	Price	Quality	VA
baseline	0.877 (0.01)	0.818 (0.0)	1.447 (0.0)	1.567 (0.0)	1.248 (0.0)	2.978 (0.0)	0.139 (0.0)
learn x late	0.939 (0.011)	0.718 (0.006)	1.454 (0.008)	1.568 (0.021)	1.351 (0.007)	3.178 (0.008)	0.22 (0.004)
learn (x,r) early	0.896 (0.011)	0.693 (0.006)	1.44 (0.008)	1.577 (0.02)	1.35 (0.007)	3.171 (0.007)	0.217 (0.004)
full info benchmark	1.191 (0.013)	0.828 (0.006)	1.54 (0.009)	1.604 (0.021)	1.375 (0.007)	3.173 (0.009)	0.21 (0.004)
baseline ($\hat{x} = x$)	1.115 (0.035)						
full info ($\hat{x} = x$)	1.475 (0.061)	0.917 (0.004)	1.477 (0.006)	1.588 (0.014)	1.24 (0.006)	2.922 (0.008)	0.099 (0.005)

Notes: This table displays main estimates for students with mothers without a four-year college degree. Columns: "EU": EU according to fully informed payoffs. "Place": probability of placement. ("E(rank)","Distance","Price","Quality", "VA"): avg. (rank of placed school within ROL,distance,price,quality,school value added (in student-level SD)), 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. "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 10. Main Results: College-Graduate Mothers

	EU	Place	E(rank)	Distance	Price	Quality	VA
baseline	1.134 (0.028)	0.772 (0.0)	1.703 (0.0)	1.81 (0.0)	1.544 (0.0)	3.105 (0.0)	0.184 (0.0)
learn x late	1.193 (0.037)	0.74 (0.005)	1.684 (0.01)	1.817 (0.031)	1.553 (0.012)	3.192 (0.011)	0.216 (0.007)
learn (x,r) early	1.111 (0.033)	0.706 (0.007)	1.64 (0.012)	1.82 (0.037)	1.554 (0.014)	3.181 (0.014)	0.211 (0.009)
full info benchmark	1.555 (0.028)	0.858 (0.005)	1.792 (0.015)	1.814 (0.031)	1.523 (0.014)	3.178 (0.017)	0.204 (0.01)
baseline ($\hat{x} = x$)	1.171 (0.046)						
full info ($\hat{x} = x$)	1.64 (0.061)	0.889 (0.005)	1.742 (0.021)	1.798 (0.03)	1.492 (0.015)	3.04 (0.019)	0.145 (0.01)

Notes: This table displays main estimates for students with mothers with at least a four-year college degree. Columns: "EU": EU according to fully informed payoffs. "Place": probability of placement. ("E(rank)","Distance","Price","Quality", "VA"): avg. (rank of placed school within ROL,distance,price,quality,school value added (in student-level SD)), 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. "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.

9 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.

Detailed data analysis and field experiments provide evidence about 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 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, leading to lower placement rates. Providing information to households who have already searched and have submitted applications, as in our feedback intervention and our "learn x late" counterfactual, may enhance demand for quality schools.

We provide a unified framework for analyzing 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.

Weitzman, Martin L., "Optimal Search for the Best Alternative," *Econometrica*, 1979, *47* (3), 641–654.

Appendix

A.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. [25] 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)). [26]

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%). [27]

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%).[28]

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

[26] 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.

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

[28] 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.[29] 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.[30]

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 A1. 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 **?**.

[29]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.

[30]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*. [31]

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.

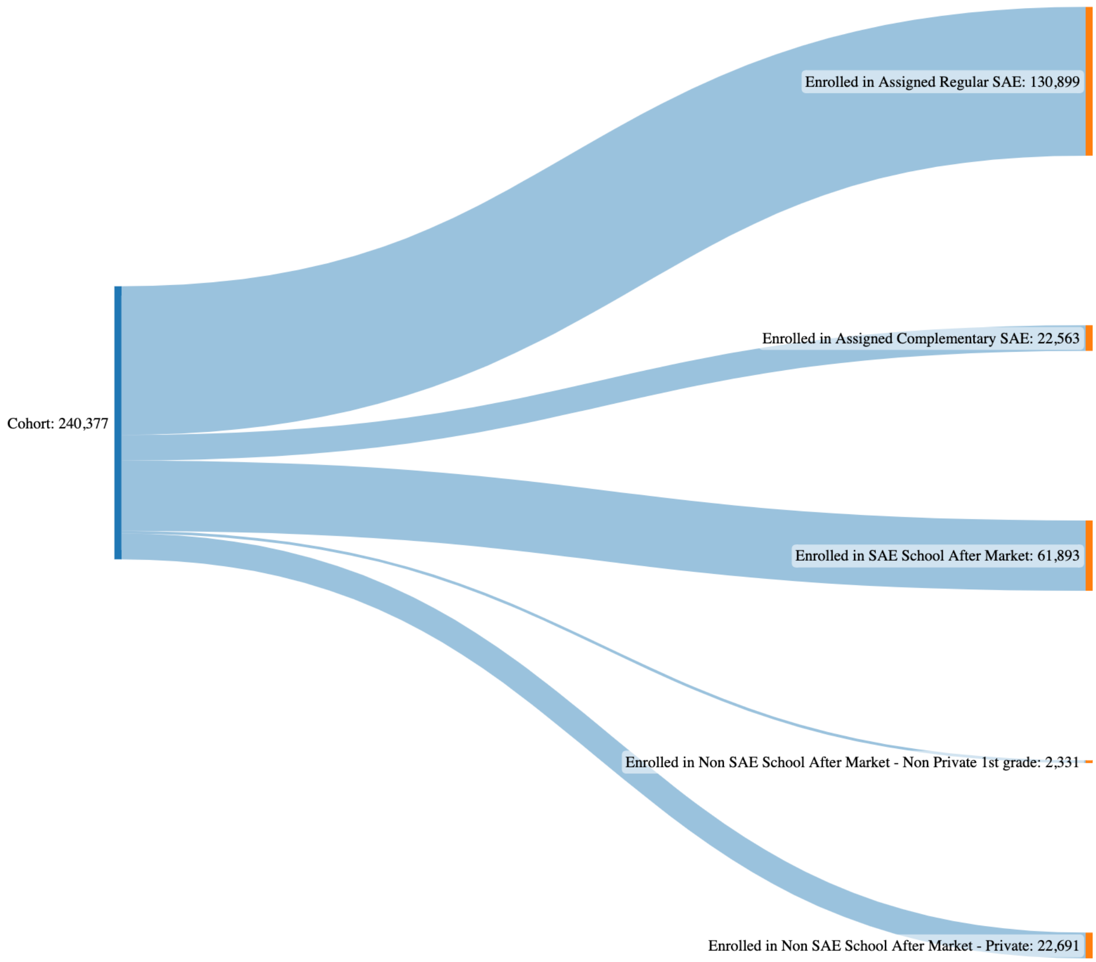
Looking at the 2021 regular stage, 56.98% of the families applied in the first five days of the regular stage, while 10.9% applied in the last five. As parents are allowed to change their application during each stage, 26.93% changed theirs during the regular stage.

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 A1). 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.

[31] 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 A1. 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 A2, 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 A2. Regular Round SAE

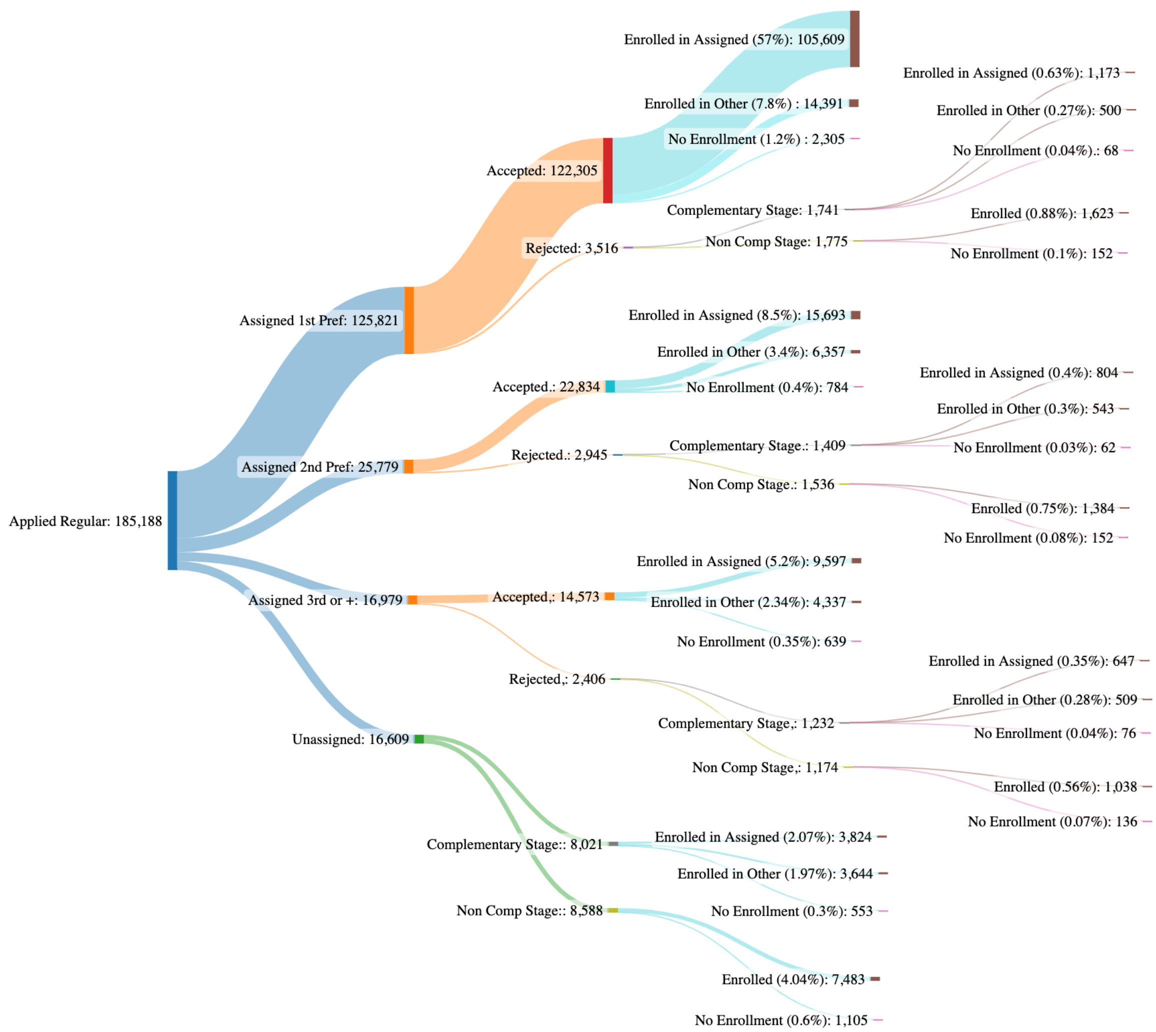
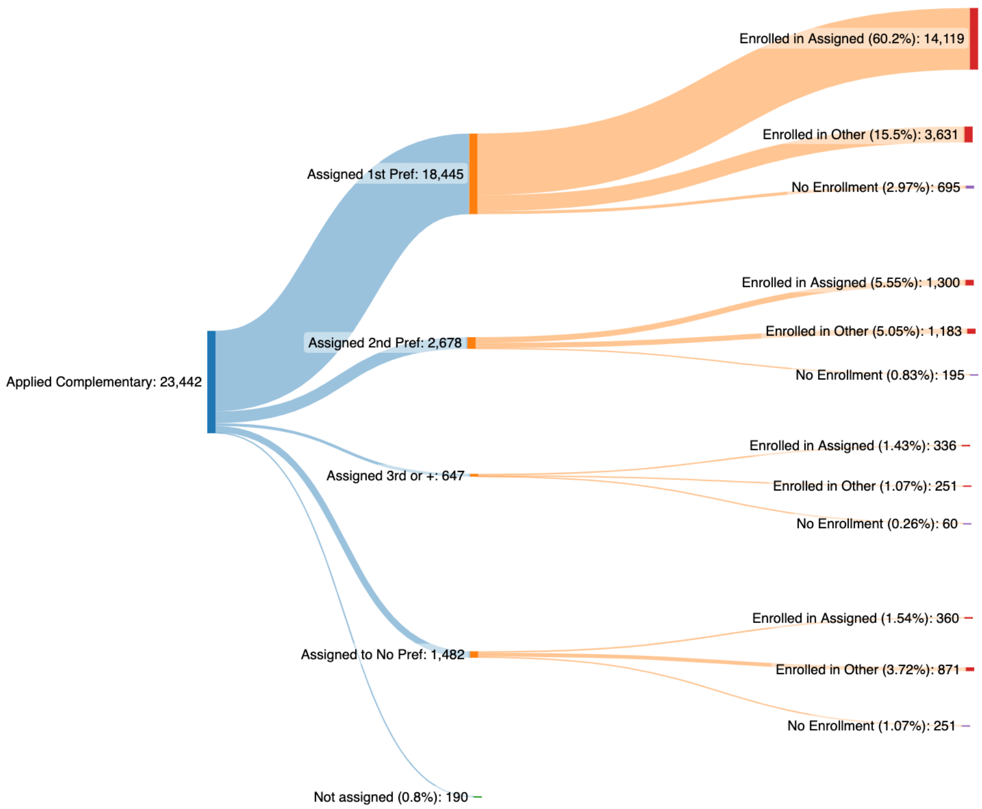


FIGURE A3. Complementary Round SAE



A.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. [32]. 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.

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

A.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.

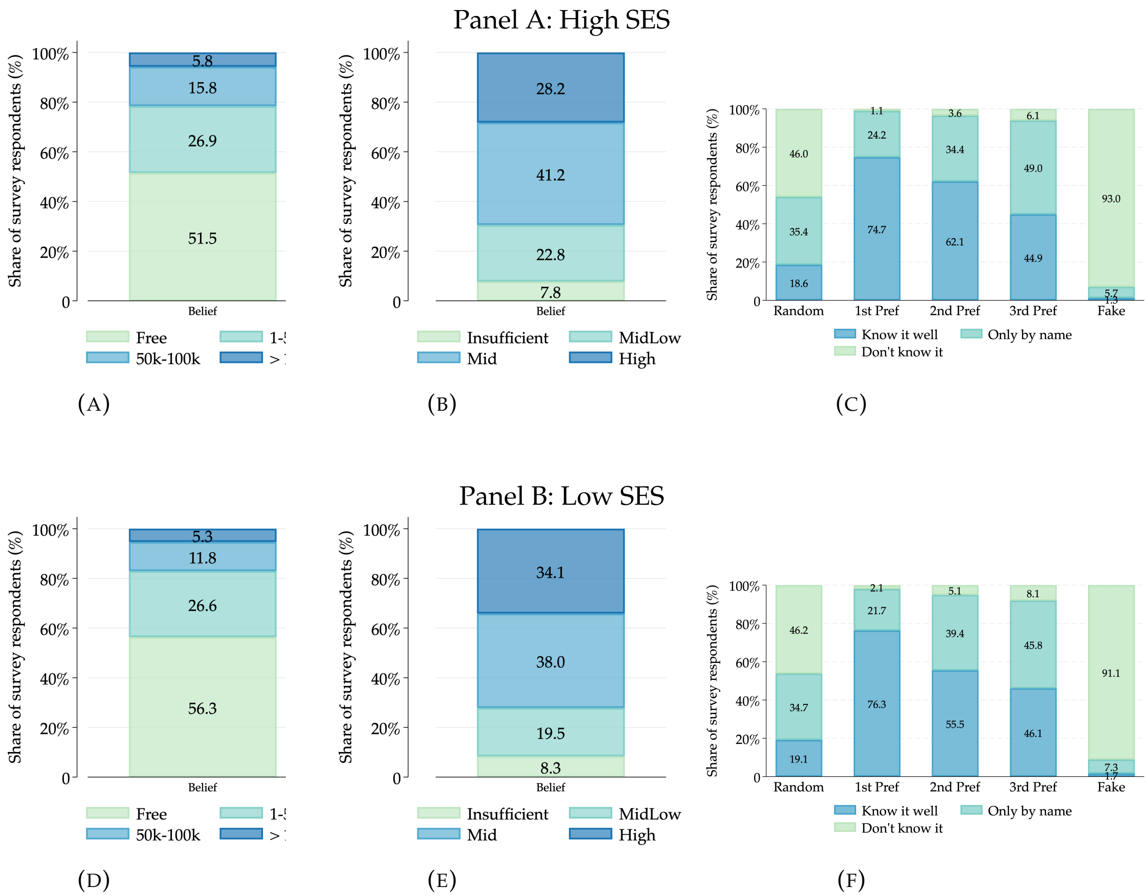
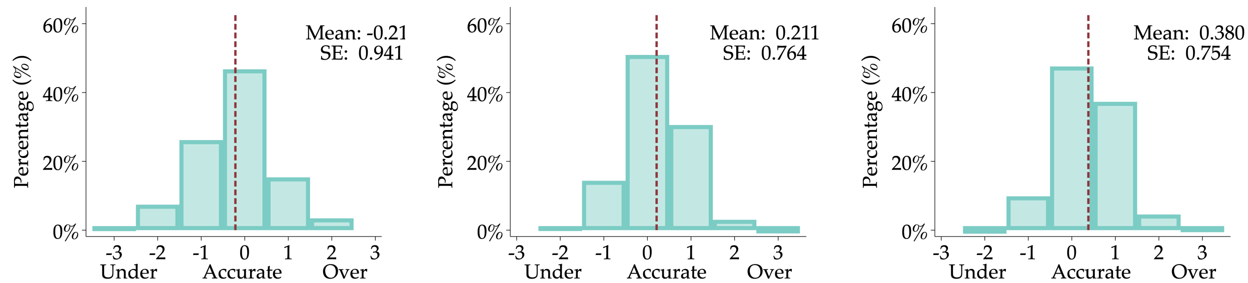


FIGURE A4. Knowledge and Distribution of X's by SES Status

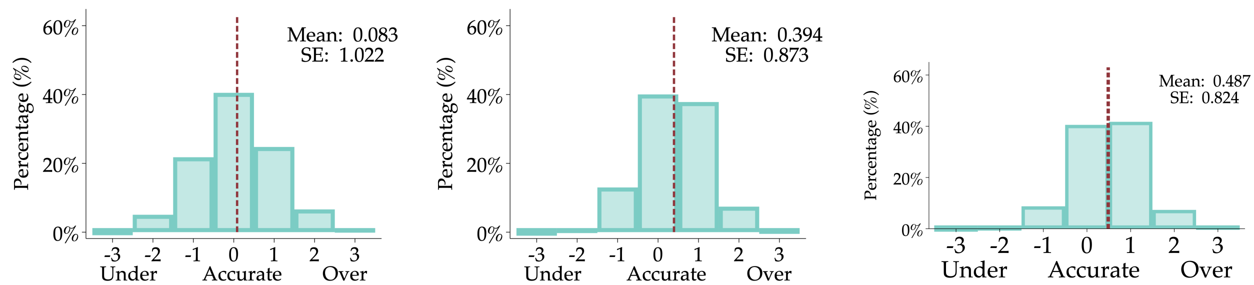
Notes: Panel (A) shows share of survey respondents stating the perceived distribution of schools in their neighborhood with monthly fee: Free, 1-50k, 50k-100k and greater than 100k versus the real distribution of schools. Panel (B) shows the share of survey respondents stating the perceived distribution 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.

Panel A: High SES



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

Panel B: Low SES

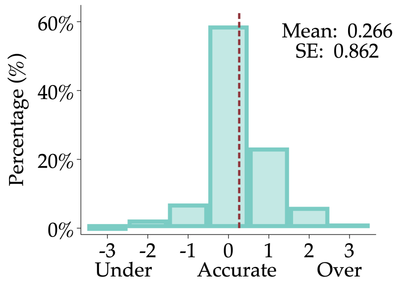


(D) Random School Known (E) Random School in App (F) First Preference

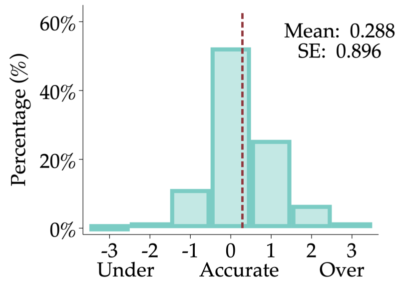
FIGURE A5. Error in Quality by SES Status

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.

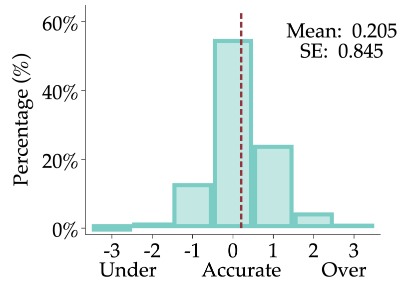
Panel A: High SES



(A) Random School Known

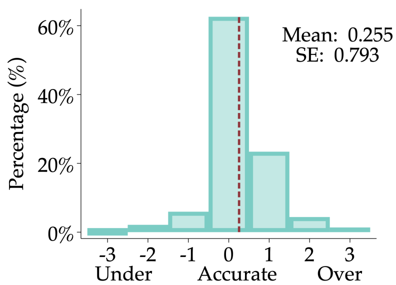


(B) Random School in App

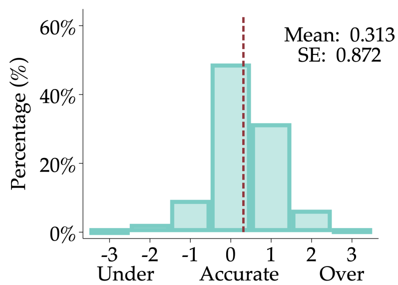


(C) First Preference

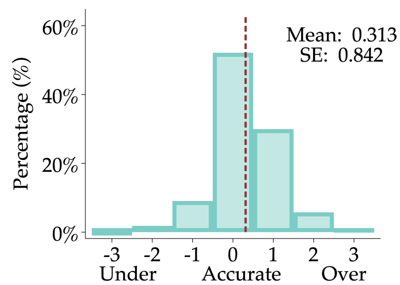
Panel B: Low SES



(D) Random School Known



(E) Random School in App

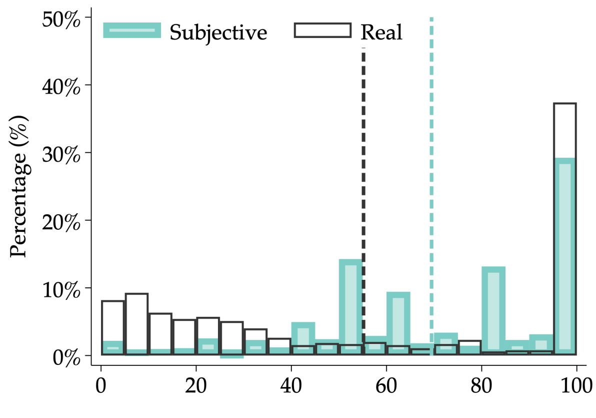


(F) First Preference

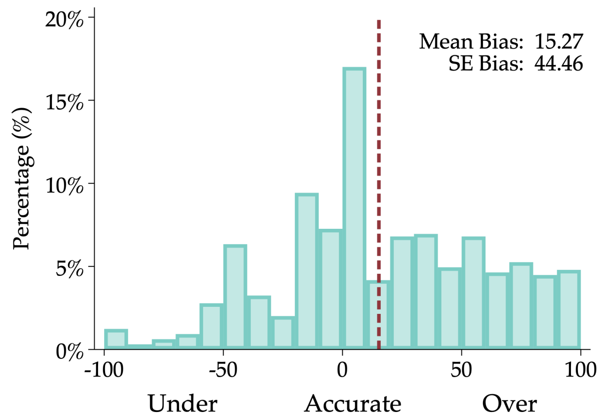
FIGURE A6. Error in Price by SES Status

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.

Panel A: High SES

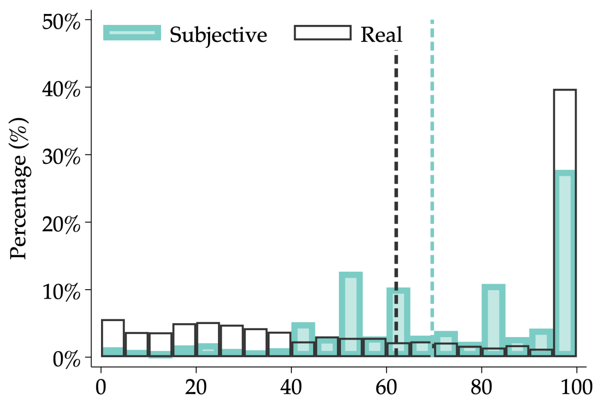


(A) Distribution of Placement Chances

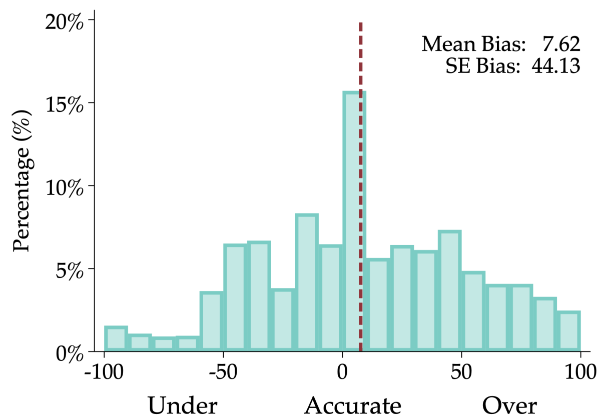


(B) Bias of Placement Chances

Panel B: Low SES



(C) Distribution of Placement Chances



(D) Bias of Placement Chances

FIGURE A7. Error in Placement Chances by SES Status

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.

TABLE A2. 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	2,805
%	1.00	0.45	0.23	0.22	0.01
Panel A: Demographics					
Economically vulnerable	0.54	0.51	1.00	0.00	0.42
Woman	0.49	0.49	0.49	0.49	0.51
Entry grade	0.45	1.00	1.00	1.00	1.00
Mother Completed College	0.20	0.23	0.07	0.36	0.24
Panel B: Application Behavior					
Length initial attempt	2.94	2.93	2.68	3.19	3.59
Length final attempt	2.98	2.97	2.71	3.24	3.68
Total attempts	1.05	1.05	1.04	1.06	1.08
Panel C: Placement					
Placed in pref.	0.80	0.88	0.91	0.85	0.94
Placed 1st pref.	0.55	0.64	0.71	0.58	0.63
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	0.98
Enrolled at placed	0.70	0.71	0.74	0.68	0.73
Free Tuition	0.78	0.75	0.83	0.66	0.75
Insufficient Quality	0.04	0.03	0.04	0.02	0.03
Mid-Low Quality	0.25	0.19	0.21	0.17	0.15
Mid Quality	0.58	0.60	0.59	0.61	0.57
High Quality	0.13	0.18	0.16	0.19	0.25
Highlight Worthy	0.43	0.45	0.52	0.39	0.49

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

TABLE A3. 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	14.828***	2.764***	1.014***	2.788***	0.999***	0.604***	0.286***
	(3.257)	(0.572)	(0.217)	(0.569)	(0.216)	(0.154)	(0.080)
Prob Search More							
Unknown Sch.	19.409***	4.077***	1.429***	3.911***	1.345***	1.291***	0.681***
	(5.829)	(0.928)	(0.353)	(0.919)	(0.347)	(0.303)	(0.150)
Known Sch.	15.291**	1.755*	0.513	1.787*	0.545	0.421	0.138
	(5.967)	(0.995)	(0.374)	(0.994)	(0.371)	(0.300)	(0.149)
Prob Add School							
Same X's	-0.046	0.188	0.385	0.129	0.328	0.143	0.138
	(5.650)	(1.051)	(0.394)	(1.044)	(0.389)	(0.296)	(0.147)
Worse X's	-1.229	-0.715	-0.416	-0.630	-0.371	0.102	-0.038
	(6.265)	(1.039)	(0.388)	(1.031)	(0.381)	(0.331)	(0.162)
Prob Add Sch. if Knew							
To Pref 1	-25.875**	-5.594***	-1.667**	-5.640***	-1.694**	-0.626	-0.077
	(11.712)	(1.942)	(0.679)	(1.943)	(0.680)	(0.559)	(0.267)
blw Last Pref	23.576*	5.205***	1.190*	5.292***	1.281*	0.665	0.041
	(12.112)	(1.993)	(0.710)	(1.998)	(0.713)	(0.537)	(0.271)
Control Mean	50.253	9.846	3.462	8.403	3.244	2.227	1.059
Observations	2900	2900	2900	2900	2900	2900	2900

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

TABLE A4. Balance Checks for Search Interventions

	Control		Treatment 1		Treatment 2		
	Mean	St. Dev.	Coeff.	St. Err.	Coeff.	St. Err.	N
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Choice Environment							
Number of schools (2km - within intervention)	16.616	[9.410]	-0.096	(0.324)	0.059	(0.328)	3,111
Number of schools (2km - total school supply)	34.546	[17.829]	0.031	(0.637)	0.283	(0.646)	3,111
Number of highlighted schools (2km - total school supply)	8.316	[5.079]	-0.011	(0.194)	0.049	(0.192)	3,111
Panel B: Parent/Child Characteristics							
Child is female	0.503	[0.500]	0.003	(0.022)	0.013	(0.022)	3,111
Child's birthyear	2017.090	[0.539]	0.003	(0.023)	0.013	(0.023)	3,111
Mother completed college	0.236	[0.425]	0.017	(0.015)	-0.015	(0.015)	3,108
Number of younger siblings	1.137	[0.382]	0.023	(0.017)	0.006	(0.017)	3,111
Child has disability (Belief)	0.068	[0.252]	0.003	(0.012)	-0.003	(0.012)	2,799
Parent works in a school	0.072	[0.259]	-0.009	(0.011)	-0.014	(0.011)	3,067
SEP Household	0.435	[0.496]	-0.020	(0.016)	-0.009	(0.016)	3,087
Panel C: Initial Knowlege and Beliefs							
Expected satisfaction with process	5.250	[1.393]	0.017	(0.063)	-0.017	(0.064)	2,916
Listed any school as pref 1	0.910	[0.286]	0.001	(0.012)	0.011	(0.012)	3,111
Listed highlighted school as pref 1	0.522	[0.500]	0.039	(0.024)	0.023	(0.024)	2,517
Perceived admission chance of 1st preference	68.880	[26.447]	0.634	(1.199)	1.880	(1.195)	2,916
N Schools known by name	3.353	[2.658]	-0.055	(0.117)	0.080	(0.114)	3,111
N Schools known well	1.866	[2.029]	0.004	(0.087)	0.027	(0.088)	3,111
Perceived Number of Schools (2km) Baseline	6.075	[7.237]	0.044	(0.274)	-0.401	(0.258)	3,111
Perceived Number of Highlighted Schools (2km) Baseline	2.125	[2.475]	0.106	(0.108)	-0.024	(0.105)	3,111
Parent believed to be SEP eligible	0.167	[0.373]	-0.012	(0.016)	-0.007	(0.016)	3,111
SEP did not know about SEP status	0.671	[0.470]	0.003	(0.021)	0.028	(0.020)	3,111

Panel D: Treatment Summary	Control	Treatment 1	Treatment 2
Observations	1027	1036	1048
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 who opened the intervention. *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.

TABLE A5. Balance Checks for High SES Households

	Control		Treatment 1		Treatment 2		
	Mean	St. Dev.	Coeff.	St. Err.	Coeff.	St. Err.	N
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Choice Environment							
Number of schools (2km - within intervention)	16.050	[8.065]	0.044	(0.632)	-0.949	(0.655)	732
Number of schools (2km - total school supply)	34.913	[16.212]	-0.090	(1.241)	-2.122	(1.327)	732
Number of highlighted schools (2km - total school supply)	7.802	[4.341]	0.511	(0.377)	-0.042	(0.386)	732
Panel B: Parent/Child Characteristics							
Child is female	0.442	[0.498]	0.050	(0.046)	0.043	(0.048)	732
Child's birthyear	2017.091	[0.523]	-0.035	(0.050)	0.035	(0.051)	732
Number of younger siblings	1.165	[0.404]	0.008	(0.036)	0.005	(0.040)	732
Child has disability (Belief)	0.084	[0.278]	-0.025	(0.025)	-0.018	(0.027)	678
Parent works in a school	0.219	[0.414]	-0.079**	(0.035)	-0.103**	(0.035)	729
SEP Household	0.155	[0.362]	0.013	(0.026)	-0.021	(0.025)	727
Panel C: Initial Knowlege and Beliefs							
Expected satisfaction with process	5.053	[1.372]	0.033	(0.139)	0.224	(0.138)	685
Listed any school as pref 1	0.917	[0.276]	-0.006	(0.027)	0.010	(0.027)	732
Listed highlighted school as pref 1	0.391	[0.489]	0.023	(0.051)	0.105	(0.053)	572
Perceived admission chance of 1st preference	68.739	[27.948]	0.287	(2.658)	3.081	(2.627)	685
N Schools known by name	3.293	[2.672]	-0.000	(0.240)	0.253	(0.253)	732
N Schools known well	2.025	[2.142]	-0.094	(0.197)	-0.147	(0.204)	732
Perceived Number of Schools (2km) Baseline	7.355	[12.034]	-0.960	(0.900)	-1.660	(0.882)	732
Perceived Number of Highlighted Schools (2km) Baseline	1.950	[2.311]	0.401	(0.249)	0.191	(0.220)	732
Parent believed to be SEP eligible	0.107	[0.310]	-0.011	(0.029)	0.011	(0.031)	732
SEP did not know about SEP status	0.657	[0.476]	-0.046	(0.043)	0.030	(0.044)	732

Panel D: Treatment Summary	Control	Treatment 1	Treatment 2
Observations	242	260	230
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 college mothers who opened the intervention. *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.

TABLE A6. Balance Checks for Low SES Households

	Control		Treatment 1		Treatment 2		
	Mean	St. Dev.	Coeff.	St. Err.	Coeff.	St. Err.	N
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Choice Environment							
Number of schools (2km - within intervention)	16.783	[9.775]	-0.052	(0.383)	0.391	(0.385)	2,376
Number of schools (2km - total school supply)	34.424	[18.292]	0.110	(0.747)	1.096	(0.746)	2,376
Number of highlighted schools (2km - total school supply)	8.479	[5.281]	-0.119	(0.230)	0.094	(0.224)	2,376
Panel B: Parent/Child Characteristics							
Child is female	0.521	[0.500]	-0.006	(0.026)	0.007	(0.025)	2,376
Child's birthyear	2017.088	[0.543]	0.015	(0.026)	0.013	(0.027)	2,376
Number of younger siblings	1.128	[0.374]	0.023	(0.019)	0.008	(0.019)	2,376
Child has disability (Belief)	0.063	[0.243]	0.009	(0.013)	0.000	(0.013)	2,118
Parent works in a school	0.026	[0.159]	0.010	(0.009)	0.013	(0.009)	2,335
SEP Household	0.522	[0.500]	-0.016	(0.018)	-0.009	(0.018)	2,357
Panel C: Initial Knowlege and Beliefs							
Expected satisfaction with process	5.316	[1.393]	0.015	(0.072)	-0.080	(0.073)	2,228
Listed any school as pref 1	0.908	[0.289]	0.004	(0.014)	0.011	(0.014)	2,376
Listed highlighted school as pref 1	0.561	[0.497]	0.052*	(0.027)	0.004*	(0.027)	1,942
Perceived admission chance of 1st preference	68.940	[26.018]	0.644	(1.359)	1.496	(1.366)	2,228
N Schools known by name	3.372	[2.658]	-0.097	(0.135)	0.029	(0.129)	2,376
N Schools known well	1.815	[1.994]	0.039	(0.100)	0.079	(0.099)	2,376
Perceived Number of Schools (2km) Baseline	5.683	[4.838]	0.273	(0.249)	0.003	(0.230)	2,376
Perceived Number of Highlighted Schools (2km) Baseline	2.171	[2.521]	0.053	(0.120)	-0.042	(0.122)	2,376
Parent believed to be SEP eligible	0.185	[0.389]	-0.011	(0.019)	-0.012	(0.019)	2,376
SEP did not know about SEP status	0.676	[0.468]	0.011	(0.024)	0.020	(0.023)	2,376

Panel D: Treatment Summary	Control	Treatment 1	Treatment 2
Observations	783	775	818
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 non-college mothers who opened the intervention. *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.

TABLE A7. Balance Check for Feedback Intervention

	Control		Feedback Treatment		
	Mean	St. Dev.	Coeff.	St. Err.	N
	(1)	(2)	(3)	(4)	(5)
Panel A: Choice Environment					
Number of schools (2km - within intervention)	17.641	[9.278]	-1.161	(0.929)	2,116
Number of schools (2km - total school supply)	36.400	[17.498]	-0.391	(1.490)	2,116
Number of highlighted schools (2km - total school supply)	8.695	[5.141]	0.057	(0.614)	2,116
Panel B: Parent/Child Characteristics					
Child is female	0.521	[0.500]	-0.006	(0.025)	2,116
Child's birthyear	2017.090	[0.533]	-0.019	(0.028)	2,116
Mother completed college	0.230	[0.421]	0.033	(0.031)	2,114
Number of younger siblings	1.124	[0.359]	0.028	(0.021)	2,116
Child has disability (Belief)	0.055	[0.229]	0.016	(0.012)	1,895
Parent works in a school	0.069	[0.254]	0.005	(0.010)	2,085
SEP Household	0.434	[0.496]	-0.020	(0.030)	2,116
Panel C: Initial Knowlege and Beliefs					
Expected satisfaction with process	5.201	[1.400]	0.007	(0.071)	1,997
Listed any school as pref 1	0.916	[0.278]	-0.010	(0.022)	2,116
Listed highlighted school as pref 1	0.542	[0.499]	0.050	(0.036)	1,741
Perceived admission chance of 1st preference	68.898	[26.717]	3.051**	(1.263)	1,997
N Schools known by name	3.318	[2.575]	0.026	(0.177)	2,116
N Schools known well	1.865	[1.994]	-0.084	(0.106)	2,116
Perceived Number of Schools (2km) Baseline	6.091	[4.935]	0.306	(0.232)	2,116
Perceived Number of Highlighted Schools (2km) Baseline	2.077	[2.257]	0.075	(0.134)	2,116
Parent believed to be SEP eligible	0.160	[0.366]	-0.014	(0.017)	2,116
SEP did not know about SEP status	0.682	[0.466]	0.031	(0.024)	2,116
Panel D: Search Treatments					
Search Treatment 1	0.000	[0.000]	0.001	(0.026)	2,116
Search Treatment 2	0.000	[0.000]	0.005	(0.021)	2,116
Observations	1168		948		

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 clustered standard errors. The sample consists of 2,116 parents. *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, *Panel C* considers responses to questions on knowledge and beliefs about the school system, and *Panel D* considers the treatment arms in the search intervention. *Panel E* reports the number of observations in each treatment group.

TABLE A8. Balance Check for Feedback Intervention for Low SES Households

	Control		Feedback Treatment		
	Mean	St. Dev.	Coeff.	St. Err.	N
	(1)	(2)	(3)	(4)	(5)
Panel A: Choice Environment					
Number of schools (2km - within intervention)	18.052	[9.642]	-0.781	(0.984)	1,646
Number of schools (2km - total school supply)	36.625	[17.848]	0.312	(1.559)	1,646
Number of highlighted schools (2km - total school supply)	8.978	[5.367]	0.318	(0.669)	1,646
Panel B: Parent/Child Characteristics					
Child is female	0.542	[0.499]	0.005	(0.029)	1,646
Child's birthyear	2017.089	[0.541]	-0.005	(0.031)	1,646
Number of younger siblings	1.113	[0.340]	0.025	(0.024)	1,646
Child has disability (Belief)	0.048	[0.214]	0.009	(0.014)	1,463
Parent works in a school	0.023	[0.149]	-0.010	(0.008)	1,617
SEP Household	0.518	[0.500]	-0.002	(0.030)	1,646
Panel C: Initial Knowlege and Beliefs					
Expected satisfaction with process	5.282	[1.375]	-0.031	(0.076)	1,553
Listed any school as pref 1	0.918	[0.274]	-0.014	(0.024)	1,646
Listed highlighted school as pref 1	0.584	[0.493]	0.055	(0.038)	1,361
Perceived admission chance of 1st preference	68.572	[26.120]	3.233**	(1.457)	1,553
N Schools known by name	3.390	[2.618]	0.038	(0.190)	1,646
N Schools known well	1.800	[1.951]	-0.137	(0.130)	1,646
Perceived Number of Schools (2km) Baseline	5.909	[4.830]	0.258	(0.264)	1,646
Perceived Number of Highlighted Schools (2km) Baseline	2.134	[2.247]	0.059	(0.139)	1,646
Parent believed to be SEP eligible	0.176	[0.381]	-0.009	(0.019)	1,646
SEP did not know about SEP status	0.688	[0.464]	0.031	(0.027)	1,646
Panel D: Search Treatments					
Search Treatment 1	0.000	[0.000]	0.001	(0.031)	1,646
Search Treatment 2	0.000	[0.000]	0.012	(0.027)	1,646
Observations	917		729		

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 clustered standard errors. *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, *Panel C* considers responses to questions on knowledge and beliefs about the school system, and *Panel D* considers the treatment arms in the search intervention. *Panel E* reports the number of observations in each treatment group.

TABLE A9. Balance Check for Feedback Intervention for High SES Households

	Control		Feedback Treatment		
	Mean	St. Dev.	Coeff.	St. Err.	N
	(1)	(2)	(3)	(4)	(5)
Panel A: Choice Environment					
Number of schools (2km - within intervention)	16.236	[7.732]	-2.238**	(1.100)	438
Number of schools (2km - total school supply)	35.627	[16.213]	-2.489	(2.151)	438
Number of highlighted schools (2km - total school supply)	7.770	[4.196]	-0.602	(0.746)	438
Panel B: Parent/Child Characteristics					
Child is female	0.447	[0.499]	-0.067	(0.044)	438
Child's birthyear	2017.087	[0.505]	-0.053	(0.049)	438
Number of younger siblings	1.155	[0.412]	0.026	(0.045)	438
Child has disability (Belief)	0.081	[0.273]	0.040	(0.028)	401
Parent works in a school	0.224	[0.418]	0.055	(0.039)	436
SEP Household	0.161	[0.369]	-0.040	(0.040)	438
Panel C: Initial Knowlege and Beliefs					
Expected satisfaction with process	4.946	[1.451]	0.188	(0.180)	412
Listed any school as pref 1	0.907	[0.292]	0.006	(0.038)	438
Listed highlighted school as pref 1	0.392	[0.490]	0.077	(0.072)	346
Perceived admission chance of 1st preference	70.087	[28.878]	1.693	(2.679)	412
N Schools known by name	3.075	[2.430]	0.193	(0.262)	438
N Schools known well	2.075	[2.135]	-0.001	(0.203)	438
Perceived Number of Schools (2km) Baseline	6.720	[5.260]	0.524	(0.730)	438
Perceived Number of Highlighted Schools (2km) Baseline	1.851	[2.270]	0.229	(0.289)	438
Parent believed to be SEP eligible	0.106	[0.308]	-0.022	(0.037)	438
SEP did not know about SEP status	0.665	[0.474]	0.039	(0.051)	438
Panel D: Search Treatments					
Search Treatment 1	0.000	[0.000]	-0.008	(0.046)	438
Search Treatment 2	0.000	[0.000]	-0.013	(0.043)	438
Observations	227		211		

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 clustered standard errors. *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, *Panel C* considers responses to questions on knowledge and beliefs about the school system, and *Panel D* considers the treatment arms in the search intervention. *Panel E* reports the number of observations in each treatment group.

TABLE A10. Additional Treatment Effects of Search Intervention

	Number of Pin Clicks		Number of Schools Known
	All	Highlighted	Well
	(1)	(2)	(3)
Panel A: Pooled			
Treatment 1	0.658	0.175	-0.156
	(0.550)	(0.199)	(0.134)
Treatment 2	-0.489	0.188	-0.322**
	(0.495)	(0.191)	(0.129)
Control Group Mean	8.171	3.066	1.967
Observations	3111	3111	1075
Panel B: Heterogeneity by Parental Education			
Treatment 1 × High SES	3.694***	1.080**	0.443
	(1.354)	(0.455)	(0.295)
Treatment 1 × Low SES	-0.344	-0.152	-0.340**
	(0.578)	(0.218)	(0.151)
Treatment 2 × High SES	0.147	0.434	0.139
	(1.214)	(0.440)	(0.271)
Treatment 2 × Low SES	-0.660	0.102	-0.460***
	(0.532)	(0.212)	(0.149)
p-value: Treat 1 x High SES = Treat 1 x Low SES	0.006	0.015	0.019
p-value: Treat 2 x High SES = Treat 2 x Low SES	0.542	0.496	0.055
Control Group Mean (High SES)	10.136	3.483	1.670
Control Group Mean (Low SES)	7.581	2.944	2.063
Observations 1 (High SES)	732	732	246
Observations 2 (Low SES)	2376	2376	828

TABLE A11. 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	0.007	(0.013)	0.009	(0.014)	-0.006	(0.010)	0.012	(0.012)	-0.002	(0.037)	-0.024**	(0.010)
Treatment 2	0.020	(0.012)	0.019	(0.014)	-0.004	(0.010)	0.004	(0.012)	0.034	(0.036)	-0.009	(0.010)
pval: T1 = T2	0.302		0.490		0.813		0.477		0.334		0.137	
Outcome Mean	0.574		0.672		0.840		0.188		1.380		0.280	
Observations	9536		9648		9626		9310		8821		9734	
Panel B: First Choice												
Treatment 1	0.000	(0.018)	0.010	(0.019)	-0.016	(0.015)	0.011	(0.018)	-0.050	(0.043)	-0.036*	(0.019)
Treatment 2	-0.014	(0.018)	0.018	(0.019)	-0.020	(0.015)	0.005	(0.018)	0.012	(0.044)	-0.007	(0.018)
pval: T1 = T2	0.447		0.651		0.758		0.769		0.151		0.109	
Outcome Mean	0.604		0.665		0.870		0.247		1.068		0.491	
Observations	2745		2775		2781		2692		2635		2806	
Panel C: Second Choice												
Treatment 1	0.018	(0.022)	0.017	(0.020)	-0.007	(0.017)	0.011	(0.019)	-0.013	(0.048)	-0.035**	(0.018)
Treatment 2	0.051**	(0.021)	0.035*	(0.020)	0.011	(0.017)	0.021	(0.019)	-0.005	(0.048)	-0.030*	(0.017)
pval: T1 = T2	0.121		0.354		0.303		0.560		0.868		0.756	
Outcome Mean	0.581		0.672		0.834		0.172		1.354		0.282	
Observations	2710		2757		2731		2632		2554		2777	
Panel D: Assigned School in SAE												
Treatment 1	-0.020	(0.020)	-0.030	(0.019)	-0.005	(0.018)	0.030	(0.019)	-0.056	(0.047)	-0.008	(0.019)
Treatment 2	-0.039**	(0.020)	-0.025	(0.018)	-0.018	(0.017)	0.006	(0.018)	-0.016	(0.047)	0.012	(0.019)
pval: T1 = T2	0.329		0.793		0.436		0.194		0.393		0.310	
Outcome Mean	0.631		0.767		0.812		0.138		1.198		0.382	
Observations	2678		2711		2705		2606		2520		2738	
Panel E: Enrolled School 2022												
Treatment 1	-0.003	(0.021)	-0.004	(0.019)	-0.012	(0.018)	0.008	(0.020)	-0.072	(0.049)	-0.015	(0.020)
Treatment 2	-0.017	(0.021)	-0.015	(0.019)	-0.018	(0.018)	-0.006	(0.019)	-0.055	(0.049)	0.004	(0.019)
pval: T1 = T2	0.516		0.567		0.765		0.479		0.718		0.315	
Outcome Mean	0.608		0.754		0.825		0.169		1.214		0.376	
Observations	2436		2661		2464		2359		2453		2736	

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 A12. 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	0.006	0.032	-0.011	-0.005	-0.015	-0.020
	(0.012)	(0.101)	(0.014)	(0.014)	(0.014)	(0.013)
Treatment 2	0.026*	0.007	-0.012	0.016	-0.012	0.002
	(0.012)	(0.080)	(0.014)	(0.014)	(0.014)	(0.013)
pval: T1 = T2	0.077	0.797	0.955	0.113	0.819	0.091
Outcome Mean	0.896	3.603	0.561	0.684	0.730	0.362
Observations	3095	2814	3502	3502	3502	3502

	Reject Assignment	Not Assigned	Complementary Period	Has + 3	Has + 4	Assigned to 1st
Panel B: Assignment						
Treatment 1	0.003	-0.008	-0.009	-0.039*	-0.024	0.016
	(0.011)	(0.011)	(0.010)	(0.021)	(0.018)	(0.021)
Treatment 2	-0.005	-0.002	-0.002	-0.003	-0.006	0.033
	(0.010)	(0.011)	(0.011)	(0.021)	(0.018)	(0.021)
pval: T1 = T2	0.431	0.582	0.465	0.075	0.298	0.393
Outcome Mean	0.055	0.069	0.068	0.398	0.226	0.667
Observations	2805	2805	3095	2806	2806	2625

	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	0.014	0.013	0.009	0.013	0.027	0.018
	(0.020)	(0.020)	(0.015)	(0.019)	(0.023)	(0.021)
Treatment 2	0.035*	0.020	0.033**	0.039**	0.044**	0.030
	(0.019)	(0.019)	(0.014)	(0.019)	(0.022)	(0.020)
pval: T1 = T2	0.254	0.700	0.092	0.177	0.441	0.551
Outcome Mean	0.764	0.788	0.873	0.701	0.552	0.705
Observations	2736	2561	2736	3095	2736	2736

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

TABLE A13. Feedback treatment affects assignment

	Unassigned	Assigned to added	Assigned to Highlighted
	(1)	(2)	(3)
Panel A: Pooled			
Open Feedback	0.012	0.017*	0.015
	(0.017)	(0.008)	(0.052)
Control Group Mean	0.074	0.005	0.605
Observations	2116	2116	2116
Panel B: Heterogeneity by Parental Education			
Open Feedback × High SES	0.038	0.017*	0.035
	(0.043)	(0.009)	(0.070)
Open Feedback × Low SES	0.004	0.017*	0.009
	(0.019)	(0.009)	(0.066)
p-value: Open Feedback x High SES = Open Feedback x Low SES	0.497	0.990	0.794
Control Group Mean (High SES)	0.099	0.000	0.605
Control Group Mean (Low SES)	0.067	0.006	0.605
Observations 1 (High SES)	460	460	460
Observations 2 (Low SES)	1656	1656	1656

TABLE A14. Feedback treatment affects knowledge

	N schools at 2km		Correct ROL Performance Category				Correct ROL Price Category			
	Belief	Ln(Belief/Real)	1st	2nd	3rd	4th	1st	2nd	3rd	4th
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Pooled										
Open Feedback	2.649*	0.324*	0.206*	0.100	0.307*	0.178	0.176	0.133	0.239*	0.154
	(1.247)	(0.142)	(0.123)	(0.126)	(0.131)	(0.168)	(0.127)	(0.125)	(0.127)	(0.159)
Control Group Mean	7.880	-0.844	0.563	0.533	0.511	0.635	0.379	0.384	0.255	0.333
Observations	368	368	209	212	151	100	217	215	153	100
Panel B: Heterogeneity by Parental Education										
Open Feedback × High SES	1.498	0.242	0.285*	0.162	0.291*	0.212	-0.170	-0.012	-0.158	-0.077
	(1.987)	(0.227)	(0.164)	(0.175)	(0.167)	(0.229)	(0.168)	(0.179)	(0.164)	(0.225)
Open Feedback × Low SES	2.981*	0.345*	0.161	0.063	0.308*	0.074	0.334*	0.190	0.431*	0.280
	(1.478)	(0.171)	(0.157)	(0.161)	(0.175)	(0.221)	(0.163)	(0.160)	(0.175)	(0.228)
p-value: Open Feedback x = Open Feedback x Low SES	0.526	0.703	0.563	0.662	0.942	0.655	0.026	0.385	0.014	0.257
Control Group Mean (High SES)	9.133	-0.789	0.654	0.583	0.600	0.786	0.423	0.462	0.350	0.429
Control Group Mean (Low SES)	7.550	-0.858	0.541	0.523	0.486	0.592	0.368	0.366	0.230	0.306
Observations 1 (High SES)	88	88	43	43	37	27	45	45	37	27
Observations 2 (Low SES)	280	280	166	169	114	73	172	170	116	73

TABLE A15. Summary of Surveys

Survey	Sections	Questions
PreBaseline	Respondent	SEP Belief. Number of Kids
N = 13,721	Student Roster + Info	Student contact information. Mother education. Interest to apply to SAE
May 25 - Jul 2	Map + Beliefs	Address. Distribution of schools in neighborhood
Baseline	Priority	SEP Belief. Parent staff priority in the application
N = 3,948	Awareness Set	Knowledge level of schools (2 fake)
Jul 7 - Jul 16	Rank Ordered List (ROL)	Perceived ROL. Perception on risk of application
	School Level Information	Belief fee + academic performance of 1st ranked, random school in application, random known school not in application
	Search	Probability of engaging in search and adding schools to application
	Beliefs	Distribution of schools in neighborhood
Midline	Explorer Usage	Satisfaction with explorer. Was able to find new schools
N = 1,670	Application + Virtual Report Card	If applied to SAE. Satisfaction with report card, if changed application after report card
Aug 24 - Oct 25	Beliefs	Real ROL. Distribution of schools in neighborhood
	Awareness Set	Knowledge level of schools (1 fake)
	School Level Beliefs	Belief fee + academic performance + risk of up to 5 schools
Satisfaction	Application	Satisfaction with application process. Knowledge level of schools in application. Belief number of schools in 2km. Mother education. SEP Belief
N = 44,673	Beliefs/Risk	Perceptions on risk of application
	Behavior	Was report card useful and for what purpose
	School Level Beliefs	Knowledge and academic performance for 5 schools (1 fake)
	Information level 1	Belief academic perfomance + fee of schools in application
	Information level 2	Likelihood of adding a school
	Siblings	Siblings application, Likelihood of rejecting assignment based on siblings

Notes: This table presents a description of all surveys used in this study. SEP Belief refers to the perceived SES status for targeted vouchers. Distribution of schools in neighborhood refers to questions about the number of schools in each price and academic performance category, and the number of schools with primary education within 2km. Knowledge level of schools gives parents the alternatives: I don't know it, Know it by name, Know it well. Information level 1 in *Satisfaction Survey* refers to the randomization made to parents when answering the survey. This randomization ocurred to keep the survey short and easier to respond.