Imperfect Information and Outside Options in Centralized School Choice*

Felipe Arteaga†

November 30, 2023

Click here for the latest version

Abstract

This paper studies the consequences of limited information among families and the presence of outside options on the take-up of centralized offers in school choice. We focus on the Chilean PK-12th application system, where 23% of the applicants receive an offer but choose to enroll elsewhere, unnecessarily blocking seats that would improve the allocation for 12% of the placed applicants and would offer placement to 11% of the non-placed students. We develop a model of the joint decision of choice and enrollment that incorporates uncertainty aversion and heterogeneous outside options. We show that imperfect information translates into penalization on the valuation of the schools, affecting application and search behavior, and decreasing the probability of enrollment. Concurrently, a greater availability of options outside the system diminishes the incentive for extensive search and lowers the cost of rejecting placement offers. Our counterfactual analyses reveal two critical insights: firstly, the success of information campaigns in suggesting alternative schools is highly dependent on their ability to thoroughly inform families about these schools. Secondly, integrating out-of-system options into the centralized application process could partially mitigate the impact of non-compliance externalities, underscoring the importance of after-market design in centralized school choice systems.

*I am thankful to Nano Barahona, Ben Handel, Christopher Neilson, and Christopher Walters for their support and guidance. I also thank Claudia Allende, Matthew Backus, Federico Echeñique, Pat Kennedy, Cristóbal Otero, Sebastián Otero, Damián Vergara, and Harry Wheeler for useful discussions, comments, and suggestions. Also to IO, Development Lunch, and Labor Lunch seminar participants at UC Berkeley. I thank the implementation team of data scientists and developers at ConsiliumBots for their help throughout the project. I also 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.

†University of California, Berkeley. Mailing Address: 530 Evans Hall #3880, Berkeley, CA 94720. Email: fharteaga@berkeley.edu.

I. Introduction

Centralized school choice systems have been increasingly adopted by numerous cities around the globe (Neilson, 2019). These systems are particularly esteemed for their capacity to foster fairness, transparency, and efficiency in educational allocations. Economists have played a pivotal role in their evolution, not only by devising student-school matching algorithms (Abdulkadiroglu and Sönmez, 2003; Pathak, 2017) but also by rigorously assessing their welfare implications (Abdulkadiroğlu et al., 2017).

However, the widespread adoption of these systems has introduced various challenges, as highlighted by Agarwal and Budish (2021). Families are required to be well-informed about available schooling options (Hastings and Weinstein, 2008), to strategize effectively based on the specific mechanism they encounter (Kapor et al., 2020), and to navigate the complexities of application platforms. These challenges extend to understanding admission probabilities (Arteaga et al., 2022) and managing policymakers' decisions on information dissemination, system expansion, and handling of external schooling options.

This paper delves into a less explored challenge posed by imperfect compliance with centralized offers, specifically examining the impact of imperfect information among families and the influence of outside options. We analyze the Chilean single-offer centralized application system, which allocates over 87% of seats in the PK-12th grade system using a deferred acceptance mechanism. Notably, 23% of applicants choose to enroll in different schools than the placement offer, most of them in a school they could have applied to, impacting the system's overall effectiveness by hindering potential placements for lower-priority students or those with unfavorable lottery numbers.

To understand the non-compliance behavior to the school offer, we develop and estimate a model of the joint decision of school choice and compliance to the placement offer. Utilizing survey data from over 200,000 applicants and administrative records from three years of school choice processes involving nearly 1.5 million applicants, we explore the interplay of imperfect information, outside options, congestion externalities, and awareness-increasing policies within the framework of imperfectly informed families.

Our empirical model introduces four key innovations to the literature of school choice modeling (Agarwal and Somaini, 2020; Abdulkadiroglu and Andersson, 2022). First, it incorporates uncertainty aversion, mapped using detailed survey data. Second, it accounts for the possibility of learning between the application and enrollment phases. Third, it addresses the challenge of unobserved choice sets in environments with numerous options. Lastly, it considers heterogeneous outside options, enhancing our understanding of family decisions in the context of school choice

Results Our survey data reveals a significant gap in families' knowledge about nearby schools and those to which they are applying. Approximately 45% of surveyed families were unfamiliar with a randomly selected school located within 1.2 miles of their home address. Furthermore, a notable 33% lacked comprehensive knowledge of their first-choice school, and this figure rose to

69% for their third preference. Crucially, the data indicates that applicants with limited information about their assigned school are up to 30% more likely to enroll in a different institution. This trend underscores the vital role that information plays in the decision-making process of families within centralized school choice systems.

Our theoretical framework elucidates two pivotal findings in the context of school choice. Firstly, it demonstrates that in scenarios characterized by uncertainty aversion, a lack of comprehensive knowledge about a school significantly reduces its perceived value. This reduction in perceived value is a direct consequence of families' uncertainty-averse preferences in the face of incomplete information. Secondly, the framework reveals that the presence of more attractive outside options diminishes the incentive for families to engage in extensive search and evaluation of in-system schools. Consequently, both these factors - diminished school valuation due to insufficient information and the allure of more appealing outside options - collectively decrease the likelihood that participants in the school choice system will adhere to their initial enrollment offers. This theoretical insight underscores the complex interplay of information, valuation, and decision-making in centralized school choice systems.

Our model estimates suggest that the uncertainty levels are considerable in magnitude, and it translates into a penalization attributed to limited knowledge that is quantitatively equivalent to the effect of increasing the travel distance by three standard deviations. Intriguingly, this uncertainty diminishes to 70% in the enrollment stage, indicating a significant degree of learning and information acquisition by families during the school selection process. This observation highlights the dynamic nature of families' decision-making, underscoring the critical role of information in shaping educational choices within centralized school choice systems.

Our analytical framework identifies non-compliance behavior as a catalyst for policy innovation in at least two critical dimensions. Firstly, the phenomenon of non-compliance underscores the need for enhanced dissemination of information about school options. This necessity stems from the observation that a significant proportion of non-compliant families ultimately enroll their children in schools that were accessible options within the centralized system but were not initially considered. Enhancing awareness and knowledge about these alternatives could potentially reduce non-compliance rates. Secondly, the inclination of certain families to opt for schools outside the centralized system motivates the integration of these outside options into the central mechanism. In our study, we leverage the estimated parameters from our model and our knowledge of the assignment mechanism to simulate the impact of policies targeted at these two dimensions. Such simulations are instrumental in projecting the effects of potential policy interventions, thereby offering valuable insights for the strategic evolution of centralized school choice systems.

In the first set of our counterfactual analyses, we design a policy intervention aimed at enhancing families' knowledge about school options, examining its impact on compliance rates. This intervention involves simulating a scenario where applicants are informed about a predicted school in which they would enroll. We manipulate two key variables in this simulation: the accuracy of the school predictions and the depth of information provided about these schools. This approach allows us to approximate the effects of real-world interventions. Our findings indicate a critical insight: merely raising awareness of a school, without providing comprehensive information about it, can diminish the effectiveness of such a policy by at least 50%. This result highlights the nuanced relationship between information depth and policy efficacy in influencing school choice behaviors.

Our second suite of counterfactual simulations delves into the potential effects of incorporating outside schooling options into the centralized system. By varying the levels of awareness and knowledge about these options, we assess the impact of this policy on aligning school placements with family preferences. Our simulations suggest that integrating outside options is likely to reduce mismatches among non-compliant families and improve placement outcomes for both compliant families and those initially not placed. Moreover, this policy lessens the application burden for families who exhibit a preference for schools outside the traditional centralized system. These insights provide valuable evidence for policymakers considering the expansion of school choice frameworks to include a broader array of schooling options.

The realm of market design, particularly when applied in practice, presents numerous challenges, as highlighted by our study. A key issue we identify is the effectiveness of information dissemination about alternative school options. While our results indicate that increased awareness could lead to more efficient school matches, this efficiency is contingent upon the depth of understanding families attain about the schools. Merely providing information is insufficient; families need to assimilate this information to make informed choices. Furthermore, our research suggests that the inclusion of a wider array of options within the centralized school choice system is likely to yield positive outcomes. However, this brings to the fore the necessity for further research. There is a critical need to explore how the practical implementation of assignment mechanisms and associated policies influence the incentives for families to seek and process information. This exploration is essential to understand the dynamics of search behavior in the context of school choice, as underscored by studies such as (Immorlica et al., 2020). Addressing these challenges will be pivotal in refining market design strategies and enhancing the efficacy of centralized school choice systems.

Related literature Our research contributes to the empirical market design literature, particularly concerning deviations from full information and the potential for mismatches in school choice systems (Agarwal and Budish, 2021). While studies like Narita (2018) have explored post-match reassignments in the New York City context, identifying preference flipping as a key issue, our work diverges by focusing on the impact of information acquisition, or the lack thereof, as a primary driver of mismatches.

Moreover, our research both aligns with and expands upon existing literature concerning applications within centralized systems under conditions of incomplete information, as exemplified by Grenet et al. (2022). Research by Hastings and Weinstein (2008), Ajayi et al. (2017), and Andrabi et al. (2017) demonstrates how providing information about school performance influences family choices. Our work builds on this by using survey data to estimate a choice model that incorporates imperfect information, highlighting how interventions in school information can yield unexpected outcomes, especially in congested environments.

Additionally, our research is closely related to empirical models of school choice, as discussed in Agarwal and Somaini (2020), particularly those utilizing micro-data from strategy-proof mechanisms (Narita, 2018; Fack et al., 2019; Abdulkadiroglu et al., 2020; Ainsworth et al., 2023). Our methodology is akin to studies examining preferences for schools and their subsequent effects on academic achievement (Abdulkadiroğlu et al., 2011; Deming et al., 2014; Walters, 2018; Neilson, 2020), and to those integrating survey data for a more nuanced analysis (Kapor et al., 2020; Arteaga et al., 2022; De Haan et al., 2023; Budish and Cantillon, 2012). Our innovative approach includes considering outside options as part of future enrollment decisions and addressing noisy school valuations.

Lastly, our work intersects with the broader discourse on discrete choice under uncertainty, exemplified in sectors like health insurance (Handel, 2013) and auto insurance (Cohen and Einav, 2007), and the study of unobserved choice sets (Crawford et al., 2021; Abaluck and Adams-Prassl, 2021). Notably, it aligns with (Barseghyan et al., 2021)'s proposition of a discrete choice model that accommodates unobserved heterogeneity in consideration sets and incorporates risk aversion.

Organization of the Paper The paper begins with an overview of the Chilean school admission system in Section II, followed by an analysis of survey results in Section III. Section IV introduces a joint model of school choice and compliance, incorporating noisy school valuation. Section V explains the integration of this model with our data, with estimation results presented in Section VI. Counterfactual scenarios are explored in Section VII, and the paper concludes with a summary of findings and implications in Section VIII.

II. Setting

We study the school choice and enrollment decisions in the context of Chile, where a nationwide centralized system has regulated the admission for 87% of the seats since 2019. This section describes the institutional details of the education system, the application process, and the assignment mechanism. It continues with a description of the demographics of applicants, application behaviors, placement, and enrollment results.

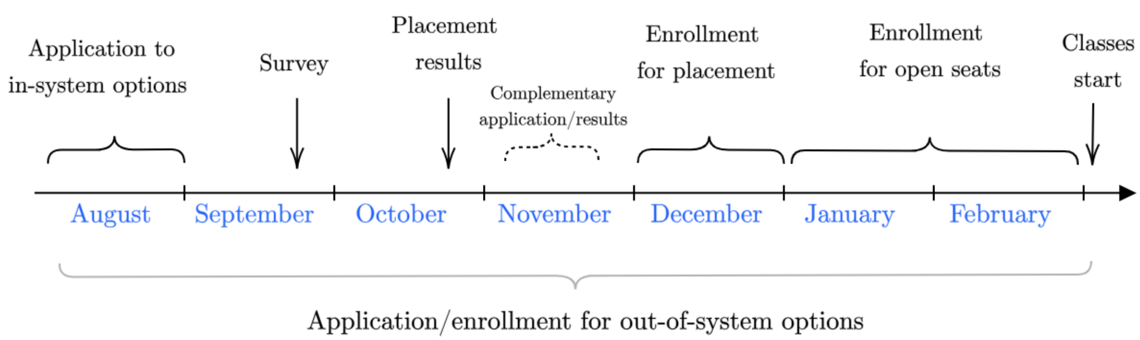
II.A. The Chilean admission system: institutional details

Chile has a long tradition of school choice. Student vouchers were introduced in 1981, with the competition-based idea that schools would improve the service provided in order to attract more students (Hsieh and Urquiola, 2006). By then, the choice process was decentralized; families applied independently to each school they liked. A bill passed in 2015 initiated the roll-out of a centralized school choice system, starting as a pilot in 2016 in a region that represents 1% of enrollment but expanded gradually to reach national coverage by 2019.

The centralized system regulates the admission for most of the public and private-voucher schools. Those schools represented 87.2% of the PK-12th enrollment in 2022 (total of 3.4 million students), and we will call them "in-system" schools. Two types of schools receive public funding (i.e., are publicly owned or private that receive vouchers) that do not participate in the centralized admission system, and we will name them "out-of-system public" schools. The first type is schools that offer kindergarten as their highest grade ("preschools", 2.8% of PK-12 enrollment), and the second type is schools with an artistic or sports specialty or hospital-based (0.5% of PK-12 enrollment). Additionally, there are private schools that do not accept vouchers and have their own admission rules,[1] that will name "out-of-system private" schools.[2]

Preschools are relevant in our setting because they are a meaningful outside option for families applying young students through the centralized process. They are composed of regular PK-K schools and language schools, representing 4.3% and 18.2%, respectively, of the total PK-K enrollment in 2022. Language schools have a particular curriculum oriented to students with language deficiencies, and they receive a voucher that is two to three times the voucher for regular education. Anecdotal evidence suggests that families can easily get a certificate to apply to those schools, independently of the language development of the kid, suggesting that it can be an alternative for most of the applicants.[3]

Figure I
Timeline of the application process



The centralized application process for the "in-system" schools usually happens during August, as Figure I shows. The government sets up an online platform that not only registers the applications but also provides information about the process and about each school.[4] All participant schools must declare the number of seats they want to fill for the next academic year. The system makes them available to participants after saving enough seats for the already enrolled

[1] Private-non-voucher schools represent 9.5% of the national enrollment and are expensive. A very small share of families apply to both voucher and non-voucher systems, making them almost separate markets.

[2] See Table B.I in Appendix B for a summary of the classification of schools and their share of the total enrollment.

[3] Another suggesting fact that language schools that educate a diverse pool of students is that they represent almost 20% of the total enrollment of PK-K, and most of their students transit to regular schools after Kinder.

[4] Figure B.I on appendix A shows screenshots of the application website from 2020.

students. There is a complementary application round for students who (1) did not participate in the main round,(2) were not assigned to any school, (3) or rejected the placement,[5] that lasts one week over November. Applicants can only apply to schools with seats not filled in the main round; hence, the menu is significantly reduced, with very few "popular" options.

Families can apply to as many schools as they want and modify their application while the process is still open. The government uses a student-proposing deferred acceptance (DA) mechanism to match students with schools (details on Correa et al., 2019). These conditions create a strategy-proof mechanism: the best strategy for the families is to rank schools according to their *true* preferences (Roth, 1982).[6]

If there are more applicants than seats for a specific school, the system uses a combination of coarse priorities and lottery numbers to ration seats. The priorities, in order, are siblings, school employee guardians, and alumni. There is also a quota for low SES students and, in a few cases, a quota for high-performing students in high-school grades.[7] Lotteries are drawn independently in every school, following a multiple tie-breaking rule (Ashlagi et al., 2019).

Results from the matching process of the main round are available in late October. Families that were assigned to a preference have to log in and accept or decline the offer.[8] The default for applicants that do not make a decision is to accept.

Applicants with an offer from the centralized process have the last two weeks of December to exercise their option to enroll in the school. If not, from January, they can enroll in any publicly funded school with spare capacity or go to a private school.[9]. All seats assigned in the centralized process in which placed students did not enroll are made available in a first-in-first-serve decentralized manner.[10]

On Figure B.II in Appenix A we explore additional heterogeneity on non-compliance behavior. Panel B.IIc

[5]Among the ones those placed in main round that we don't observe enrolled in their placement, only 14% participated of the complementary round.

[6]There is a growing literature on how applicants' behavior departs from truth revealing in settings with Deferred Acceptance. Hassidim et al. (2017) examine data from various nations and markets, discovering that a significant proportion of participants fail to disclose their authentic preferences. Hakimov and Kübler (2021) offer a comprehensive review of experimental studies in the field of centralized school choice and college admissions, highlighting findings related to deviation from truth-telling. Our main specification is robust to any behavior related to not including desired schools because we only infer preferences from the ranked alternatives. It's not robust to changes in the order of the ranking, and we abstract from that.

[7]The quota for low SES students reserve 15% of the seats for the applicants from the poorest tercile families (called *priority students*), and the quota for high-performing students saves 20% of the seats for the student from the highest grade quintile on their previous school.

[8]There is a third option to wait for the waitlist on higher preferences, but it is not commonly used.

[9]By law, every school that receives vouchers has to accept students if the enrollment is less than the capacity they declared for the centralized assignment. Private non-voucher schools have costly and selective admission process that usually begins earlier than the centralized admission system.

[10]The Ministry of Education keeps track of the available seats in the website vacantes.mineduc.cl. This website calculates the number of available seats as the difference between capacity and current enrollment, updated on a daily basis.

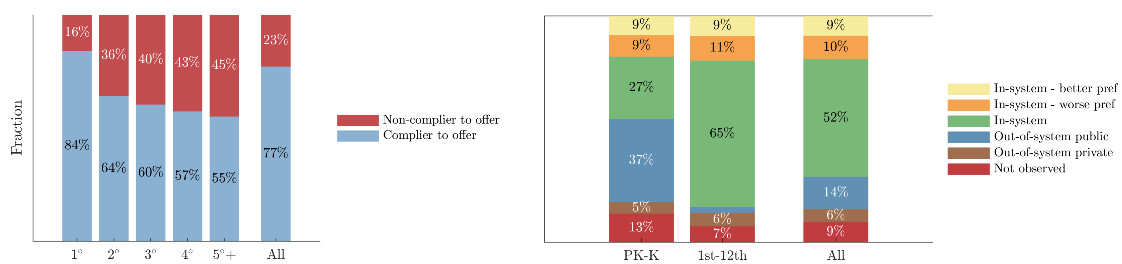
II.B. The Chilean admission system: sample description

In this work, we use the data from the application processes of 2020, 2021, and 2022, with around half a million participants per year. We complement the dataset with corresponding enrollment data from 2021 to 2023 and a novel large-scale survey that we collected between the application period and the publication of the placement results, during the three years of the application process.

We describe the population of applicants in Column 1 of Table I. 54% of the applicants are classified as low SES according to the Ministry of Education,[11] and 95% apply in first preference to a school located in an urban zone. 76% of the participants are placed in a preference in the main round. Of those, 68, 18, and 8% are assigned to 1st, 2nd, and 3rd preference, respectively. 23% of assigned applicants do not enroll in the placed school, a fraction that increases to 26% for the city of Santiago. We will call "compliers" placed applicants who enroll on placement and "non-compliers" the ones who do not enroll. Figure IIa shows that compliance rates decrease sharply with the placement ranking. 84% of applicants placed on 1st preference enrolled in the offer, while 55% of the participants assigned to a 5th or worse option complied with it.

Figure II
Compliance to placement offers and final enrollment

(a) Compliance decision by placement ranking (b) Where do non-compliers enroll?



To better understand the noncompliance behavior, we look at the next academic year's enrollment of participants who did not comply with placement. Figure IIb shows the fraction of students that went to in-system and out-of-system schools, as defined at the beginning of this section. 52% enrolled in an in-system school that they did not apply to, 19% in a school they applied to but where they were not assigned to. We observe that 14% attend a publicly funded off-platform school (out-of-system public), while 6% attend a private off-platform school (out-of-system private). A relevant 9% of students with an offer from the centralized application system are not observed in any regular school.[12] The figure depicts that it is much more common for PK-K noncompliers to enroll in an out-of-system public school than for 1st-12 applicants. This is

[11]Applicants from the poorest tercile families are considered Low SES students, and they receive a higher voucher. The term that the Ministry of Education uses is *priority students*.

[12]In Chile PK and K are not mandatory, part of the non-observed students could be preschoolers staying at home. Also, parents can opt for a non-traditional school or home-schooling option at any level and validate the studies at the end of each year or educational cycle. We do not have access to the list of families that take that option, so we cannot distinguish between students who are not receiving education or are homeschooling.

Table I
Descriptive Statistics for Choice Applicants

	(1) All	(2) Survey respondent	(3) Estimation sample
A. Applicants demographics			
Female	0.50	0.50	0.50
Low SES	0.54	0.48	0.42
Voluntary applicant	0.35	0.31	0.25
Santiago (main urban zone)	0.35	0.39	0.50
Rural	0.05	0.04	0.00
PK-K	0.32	0.36	0.40
1st-6th	0.29	0.27	0.27
7th-12th	0.39	0.37	0.34
2020	0.31	0.35	0.35
2021	0.31	0.29	0.30
2022	0.38	0.36	0.35
Reliable goecoding	0.60	0.67	1.00
Survey respondent	0.14	1.00	1.00
B. Application and placement			
Length portfolio	3.00	3.17	3.66
Placed in any preference	0.76	0.76	0.77
Placed in 1st preference|placed	0.68	0.66	0.60
Placed in 2nd preference|placed	0.18	0.18	0.21
Placed in 3rd preference|placed	0.08	0.08	0.10
Enroll in placement|placed	0.77	0.81	0.80
N	1,486,529	203,252	99,642

Notes. All statistics are means in the population defined by the column header. Selected row variable definitions are as follows. "Low SES" is a socio-economic status measure computed by Mineduc, representing roughly the poorest tercile of families. "'Voluntary applicant" indicates students applying from a school where they may continue studying. "Rural" is an indicator if students apply on first preference to a school located in a rural area. ""Voluntary applicant" indicates students applying from a school where they could continue studying. "Reliable geocoding" represents home addresses we could successfully geolocate. "Length portfolio" is the number of schools on an applicant's final choice application.

probably related to the availability of regular and language preschools that are exempted from participating in the centralized process, as described before.

We highlight three heterogeneous behaviors of non-compliance. First, low-SES compared to mid and high-SES students enroll in a better than placement preference less frequently (8 vs 12%), in a worse option more frequently (12 vs 8%), and rarely enroll in an out-of-system private school (1 vs 11%). Second, 83% of voluntary applicants enroll in an in-system school that they did not apply to, compared to 35% of the non-voluntary group. Third, we observe a higher fraction of non-compliers going to out-of-system private schools when placement goes down in ranking: 3% of applicants assigned to 1st, 13% of applicants assigned to 5th or lower. Refer to Figure B.II in Appendix A for more details.

III. Survey

To gain insights into how families navigated the school choice process, we collaborated with the Ministry of Education to survey choice participants, as related work has done it (Kapor et al., 2020; Wang and Zhou, 2020; Arteaga et al., 2022; De Haan et al., 2023). The survey examined preferences, information about options, beliefs on placement chances, search behavior, and other facets of the choice experience.[13] We use an expanded version of the 2020 survey sample utilized in Arteaga et al. (2022) that includes respondents from 2021 and 2022 choice processes.

During the three years of surveys, the Ministry of Education sent an invitation to participate to 1,249,298 families after the application process finished but before placements were announced, as Figure I suggests.[14] This timing allowed applicants to recall their experience while avoiding the influence of the results on the answers. Of those contacted, 203,252 (16%) completed the survey. Respondents closely resembled the overall population in application patterns, though they were slightly less likely to be low SES or rural (see column 2 of Table I).

III.A. Survey findings

Our survey analysis focuses on diagnosing the level of information families have during the application and the relation with the application and enrollment decisions. Applicants were asked how well they know the school on their rank order lists, but also about schools that they did not apply to. Responses were collected before placements were announced, avoiding ex-post rationalization. The key finding that emerges is that applicants have limited knowledge about both the schools they are applying to and nearby schools that they did not include in the portfolio.

Figure IIIa shows the answers to the question "How well do you know the schools in your application?".[15] 30% of the families answered that they *don't* know well the first ranked school, so either they said "I know it by name" (26%) or " I don't know it" (4%). [16]. Going down the ranking

[13]See Appendix H for a translated version of the survey questions.

[14]The number of surveys sent differs from the total number of applicants because parents who filed the application for multiple students were surveyed on only one applicant, and some families didn't have valid e-mail addresses.

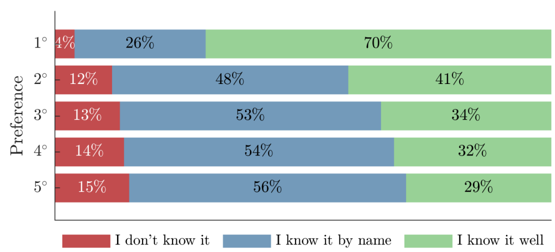
[15]A screenshot of the question from the implemented survey is shown in Figure B.III on Appendix A

[16]Since we asked about schools they included on the ranking, and some respondents said "I don't know it", we will

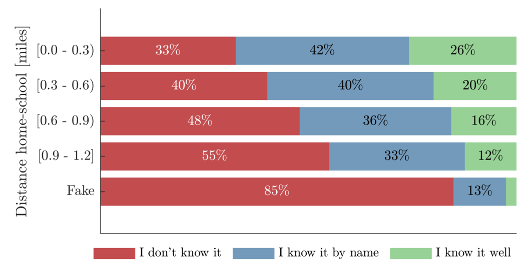
from 1st to 2nd, we observe a sharp increase in the fraction of respondents who don't know the school well from 30% to 60%, reaching 71% for the 5th choice. When we split our sample into students with mothers with at most secondary education (48%) and more educated (52%), we find that answers are similar, the former group declared slightly less knowledge (see Figures B.IVa and B.IVb in Appendix A).

Figure III
Knowledge level about schooling option

(a) Ranked options



(b) Non-ranked nearby options



We also find field evidence of limited knowledge about schools that families didn't include in the ranking. For each applicant, we picked between 1 to 5 random schools that were not part of the application but were located close to the home address. We asked about the level of knowledge of those schools in the same fashion we did for the ranked options. Figure IIIb shows in the vertical axis how far the school was from home, while the bars represent the average response. First, we can see that distance correlates with knowledge. 33% of respondents were unfamiliar with schools located within 0.3 miles, while 55% of families indicated they did not know the schools situated between 0.9 to 1.2 miles away from home. Second, considering that the median distance to the first ranked option is 0.86 miles and to the third is 1.2 miles, it is surprising how limited the awareness is: only 26% of families declared being well informed about a random school less than 0.3 miles from home. Third, to gauge the attentiveness of our respondents, we inquired about a fake school. Fortunately, most people responded "I don't know it".

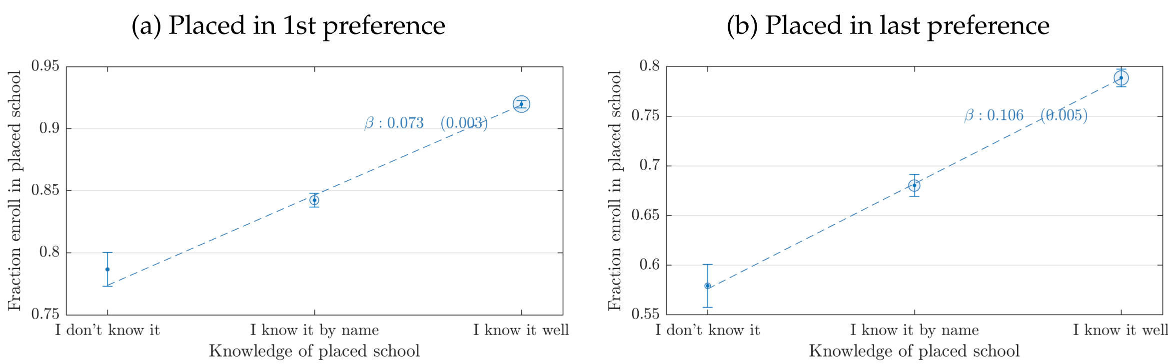
Since the meaning of knowing a school well is subjective, we gave survey respondents a list of eight different steps and asked them which of those they considered necessary to get to know a school well. Participants were allowed to choose multiple items. Figure Va shows the fraction of respondents that selected each alternative. It seems that acquiring a comprehensive understanding of a school necessitates accessing extensive information, which can sometimes be expensive. Notably, certain pieces of this information can be easily sourced from public records and platforms, like educational mission (93% said it's necessary) or academic performance (93%), but others require a more significant effort. 89% of respondents claim that knowing the infrastructure is essential, and 66% answered that interview with a staff member is a necessary step.

We use the survey responses to explore the correlation of knowledge about the placed school

not interpret the answers literally. Instead, we think of answers "I don't know it", "I know it by name" and "I know it well" as three ordinal levels of knowledge.

with the enrollment decision. Since families may decide to learn more about schools they initially liked or to stop learning about schools with a bad initial assessment, it's plausible that knowledge about a school correlated with preferences, and preferences matter for enrollment. As an attempt to isolate the relationship between knowledge and enrollment, we use as control the responses to the question of hypothetical satisfaction, a proxy for preferences. Figures IVa and IVb show the fraction of applicants that enroll in placement conditional on their level of knowledge after residualizing for enrollment satisfaction. For students placed in 1st preference, a decrease in knowledge from our highest to the lowest is related to a decrease in 15pp. For the ones placed in the last ranked option, the decrease is 21pp. These results suggest that knowledge at the application stage matters in enrollment.

Figure IV
Correlation of enrollment and knowledge about the placed school

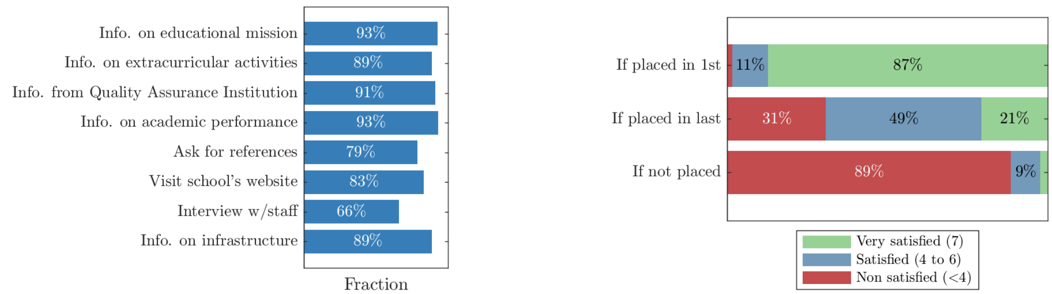


Notes: These plots show the fraction of applicants that enroll in the placement offer conditional on their answer of knowledge about the schools. Panel (a) shows for students placed on 1st preference, and panel (b) placed on last preference. Means are computed after controlling for the probability of being assigned to the first (panel a) or last preference (panel b) and the subjective satisfaction with the school if it was the placement, collected in the survey before the placement result.

Lastly, the survey evidence indicates that receiving a placement offer matters to families, and, as expected, they also care how far in the ranking are placed. We asked the satisfaction for three hypothetical results: being placed in first preference, last, or not receiving an offer. As Panel Vb shows, almost 90% of the families told us they would be completely satisfied if they got their first option. As expected, satisfaction drops significantly if placed at last; only 21% would be fully satisfied, and 31% gave a grade that is below the passing number on Chilean standard. This tells us that placement has first-order relevance despite all the potential outside options that families could have.

Figure V
Steps to get to know a school and satisfaction with placement

(a) Relevant steps to get to know a school well (b) Satisfaction with hypothetical placements



IV. Model

We continue our analysis by introducing a dynamic model of school choice and enrollment, which takes into account the possibility of noisy valuations of schools. The theoretical analysis serves two objectives. Firstly, it aims to demonstrate how uncertainty in school valuations and the availability of outside options impact students' decisions to (1) search for schools to add to their applications, and (2) enroll in placed schools. Secondly, it aims to define a model that can be estimated and used to evaluate the externalities of non-compliance and to estimate hypothetical changes in the market design.

Our analysis focuses on an individual student who is searching for schools to add to her school choice application. She is aware that in the future, she will choose between the centralized placement offer and an outside option. Our approach is similar to the model of multischool portfolio formation in Arteaga et al. (2022), which is based on job search models (McCall, 1970). The key difference is that our model has a dynamic component that allows applicants to consider the outside option as an alternative to any placement outcome, and incorporate uncertainty about the true school valuations.

IV.A. Setup

We propose a two-stage model of the rank order list formation and enrollment decision for families participating in centralized school choice. At **stage 1**, applicants form beliefs in utilities of the *enrollment option* at the schools they know and collect other inputs for the application decision. They decide the ranking, search for additional schools, and submit the rank order list to the centralized platform.[17]. At **stage 2**, each applicant receives a unique placement[18], and potentially learn more about this option, and decide to enroll on the placement offer or take the outside option. We now describe both stages in detail.

[17]Since we are modeling applications to PK to 12th grades, parents and students have a crucial role in the decision, but for simplicity, we will refer indistinctly as an applicant, student, or family. Also, despite 27% of guardians filing an application for 2 or more students; our choice model does not directly consider the joint decision. We do consider the joint placement with siblings as input for the enrollment decision.

[18]Non-placement is also a possible outcome. In our setting, 24% of applicants are not placed to any preference.

IV.B. Stage 1: subjective utility formation under uncertainty, search, and application.

At **stage 1**, applicants form beliefs in the utilities of the schools they know. This is the utility derived from the enrollment option at each school since placement offers are not binding but provide a guaranteed seat. At this stage, families have a noisy valuation of schools, so they decide which schools to rank based on an expectation. Formally, the utility derived from enrollment at school j for family i is U_{ij}, but the perceived utility is composed by the true utility plus a noise: $U_{ij}^{p1} = U_{ij} + \eta_{ij}$. The noise distributes $\sim N(0, \sigma_{k(i,j)})$, and families are aware of its distribution conditional on knowledge about the school ($k(i,j)$), but they can't differentiate their particular realization of noise from U_{ij}. Many aspects affect the valuation of a school, and families may miss and value different attributes when they have imperfect knowledge. Resolving the noise implies shifts in utility in different directions, some applicants will discover good news, others bad ones.

When families include this random term on their valuation, the Bernoulli utility in stage 1 is:

$$U_{ij}^{b1} = f(\underbrace{U_{ij}^{p1}}_{\substack{\text{Perceived}\\ \text{utility}}} - \underbrace{\eta_{ij}}_{\substack{\text{Noise}\\ \text{random}\\ \text{variable}}})$$

Where $f()$ is a Bernoulli utility function representing the attitude towards uncertainty. Families in our model are uncertainty averse, hence $f'' < 0$.

Given the uncertainty, families choose based on a measure that is monotonically related to the expected value of U_{ij}^{b1}, which we will call EU_{ij}^{s1}. The curvature of the function $f()$ reflects the effect of the uncertainty on the utility. Observe that if families were risk neutral ($f'' = 0$), then the expectation of the utility at stage 1 would be just the perceived utility ($\mathbb{E}[U_{ij}^{b1}] = U_{ij}^{p1}$), since η_{ij} has mean zero.

In order to empirically estimate this function, we will assume that $f()$ is the constant absolute risk aversion (CARA) function with a risk parameter $r > 0$, giving the following expression for the expectation:

$$\begin{aligned}
\mathbb{E}[U_{ij}^{b1}] &= \mathbb{E}[-\frac{1}{r}\exp(-r(U_{ij}^{p1} - \eta_{ij})] \\
&= \mathbb{E}[-\frac{1}{r}\exp(-rU_{ij}^{p1})\exp(r\eta_{ij})] \\
&= -\frac{1}{r}\exp(-rU_{ij}^{p1})\mathbb{E}[\underbrace{\exp(r\eta_{ij})}_{\sim LogN(0,(r\sigma_\eta)^2)}] \\
&= -\frac{1}{r}\exp(-rU_{ij}^{p1})\exp\left(\frac{(r\sigma_{\eta_{ij}})^2}{2}\right)
\end{aligned} \tag{1}$$

Notice that U_{ij}^{p1} is known for the families, so it is constant with respect to the expectation operator. The last line of equation 1 uses the fact that $\exp(r\eta_{ij})$ distributes log-normal since

$r\eta_{ij} \sim N(0, r\sigma_{\eta_{ij}})$. We then replace the expectation with the known analytical expression of the first moment.

The measure in which families based their choice in our model is $EU_{ij} = g(\mathbb{E}[U_{ij}^{b1}])$, where $g()$ is the rank-preserving transformation $g(x) = -log(-rx)/r$:

$$
\begin{aligned}
EU_{ij}^{s1} &= g\left(-\frac{1}{r}\exp(-rU_{ij}^{p1})\exp\left(\frac{(r\sigma_{\eta_{ij}})^2}{2}\right)\right) \\
&= U_{ij}^{p1} - \frac{r\sigma_{\eta_{ij}}^2}{2}
\end{aligned}
$$

EU_{ij}^{s1} is similar to the expectation in the risk-neutral case but for $-\frac{r\sigma_{\eta_{ij}}^2}{2}$. The new term tells us that uncertainty-averse families ($r > 0$) perceive a lower subjective utility if the uncertainty aversion is higher ($\frac{\partial EU_{ij}^{s1}}{\partial r} < 0$), or if the variance of the uncertainty is higher ($\frac{\partial EU_{ij}^{s1}}{\partial \sigma_{\eta_{ij}}^2} < 0$). [19]

As we will describe, in stage 2, families have the *option* to comply with the assigned school or take an outside option. Then, what matters to them in stage 1 is the *enrollment option* utility of each school w_{ij}, defined as the expected maximum between the utility of school j and the outside option:

$$
w_{ij} = \mathbb{E}\left[\max\left(\underbrace{\lambda EU_{ij}^{s1} + \xi_{ij}}_{\text{Utility of school j}}, \underbrace{\lambda U_{i0} + \xi_{i0}}_{\text{Utility of outside option}}\right)\right]
$$

Where ξ_{ij} and ξ_{i0} are future preference shocks, known to the families only in stage 2, assumed $\overset{\text{iid}}{\sim} EVI$ and uncorrelated with ϵ_{ij}. U_{i0} is the observed utility of the outside option, which depends on the geographic supply of schools in the after-market.[20] The scale factor λ multiplying EU_{ij}^{s1} and U_{i0} allows the unobserved part of U_{ij}^{p1} (not introduced yet) and ξ_{ij} to have different variances.

Given the distributional assumption for ξ_{ij} and ξ_{i0}, the expected value of the maximum between the two utilities, from the applicants' perspective, has the following closed form:

$$
w_{ij} = \log\left(\exp(\lambda EU_{ij}^{s1}) + \exp(\lambda U_{i0})\right)
$$

We assume that families optimally rank schools they know based on w_{ij}, the dominant strategy

[19] Apesteguia and Ballester (2018) notes that combining standard expected utility theory with additive unobserved utility results in non-monotonicity of choice probabilities in the risk preferences, an undesirable feature. Our framework is immune to this critique since ϵ is an additive component of U_{ij}^{p}, that is embedded into the Bernoulli utility function, as we detail in Section V. Posterior algebraic manipulation of the expected utility generates a convenient additive unobserved utility component.

[20] We detail how we model the utility of the outside option on section V.D.

when the allocation mechanism is Deferred Acceptance, as in our context. Defining Ω_i as the set of known schools for family i, $\mathcal{C}_i \subset \Omega_i$ as the current rank order list that contains $N = |\mathcal{C}_i|$ schools, and p_{ij} as the subjective placement probability in school j if applying as a first option, the expected utility derived from the rank order list $\mathcal{C}_i$ is the following:[21]

$$\mathcal{V}(\mathcal{C}_i) = w_{i1}p_{i1} + w_{i2} \underbrace{p_{i2}R_{i1}}_{\text{Prob. placed to 2nd}} + \ldots + w_{iN} \underbrace{p_{iN} \prod_{j<N} R_{ij}}_{\text{Prob. placed to Nth}} + EU_{i0} \underbrace{\prod_{j\leq N} R_{ij}}_{\text{Prob. not placed}}$$

This expression requires relabeling schools for each applicant such that a $w_{i1} > w_{i2} > w_{i3} > \ldots > w_{iN}$, so school 1 is the most preferred school in the choice set, but not necessarily the same school for all applicants. $R_{ij} \equiv 1 - p_{ij}$ is the probability of not being placed in j, hence $\prod_{j\leq N} R_{ij}$ is the probability of not being placed to any school in the ROL. At stage 1 of our model, the utility derived from non-placement is the expected utility of the outside option $EU_{i0} = \mathbb{E}[\lambda U_{i0} + \xi_{i0}]$.

Families will engage in a new iteration of a sequential search process if the increase in value of the portfolio with an additional school is higher than the cost of the search for the new school. Defining w_{is} as the utility of enrollment option for the "next school to be found", an unknown object for families, and κ_i as the search cost, a new search iteration will happen if:

$$E[\mathcal{V}(\mathcal{C}_i \cup s) - \mathcal{V}(\mathcal{C}_i)] - \kappa_i > 0$$

Assuming that the new school to be found is added as last in the new portfolio[22], the expected value of a new search iteration is:

$$\mathbb{E}[(w_{is} - EU_{i0})p_{is} \prod_{j\leq N} R_{ij}] - \kappa_i > 0$$
$$\int (w_{is} - EU_{i0})p_{is} \, \mathrm{d}F_i(EU^{s1}_{is}, p_{is}) \prod_{j\leq N} R_{ij} - \kappa_i > 0$$

The probability of realization of a new search iteration depends on (1) family i's beliefs on the joint distribution of EU^{s1}_{is} and p_{is} ($F_i(EU^{s1}_{is}, p_{is})$)[23], (2) the expected utility of the outside option, (3) the subjective belief about being placed on the already included schools ($\prod_{j\leq N} R_{ij}$), and (4) the search cost (κ_i).

Once the search is over, the family submits the ROL $\mathcal{C}_i$ to the centralized platform and waits

[21] Ordering schools from higher to lower w_{ij} is the result of the maximization of $\mathcal{V}(\mathcal{C}_i)$ conditional on the choice set Ω_i, and knowing the DA rules. Any other order would shift placement probability from a preferred school toward a less preferred school.

[22] We extend the analysis for cases where families add a school in a different position than last in Appendix F. Arteaga et al. (2022) shows that of all the families who added a school to the initial portfolio, 86% added one in the last position.

[23] Remember that $w_{is} = \log\left(\exp(\lambda EU^{s1}_{is}) + \exp(\lambda U_{i0})\right)$, the part of w_{is} that is unknown to the families is only EU_{is}, hence the expectation operator is over EU^{s1}_{is} (and p_{is}).

for the allocation process.

The way we model stage 1 predicts at least five application behaviors. First, families will search more if they think schools out there are better and less congested ($\uparrow \mathbb{E}[(w_{is} p_{is}])$. Second, a better outside option ($\uparrow U_{i0}$) makes the search less appealing.[24] Third, families that think they are going to be placed in some school in $\mathcal{C}_i$ ($\prod_{j \leq N} R_{ij} \approx 0$) do not benefit from additional search. Fourth, uncertainty about schools ($\uparrow \sigma_\eta$) makes searching and extending the portfolio less likely through reducing w_{is}.[25] And fifth, if there are schools slightly known but not in ranking, there has to be a "reservation utility", that rationalizes that families are not benefiting from the enrollment option at some schools.

IV.C. Stage 2: placement offers, learning, and enrollment decision

At **stage 2**, students receive an offer $z(i)$ ($= 0$ if no offer). They potentially learn about it, which is reflected in our model as shrinkage at rate τ_i of the noise on the enrollment option utility. The perceived utility at stage 2 is $U_{ij}^{p2} = U_{ij} + \tau_i \times \eta_{ij}$. A preference shock $\xi_{iz(i)}$ is realized and reflects changes in preferences over characteristics of the placed school $z(i)$ but also life situations such as moving home or grade retention. At this stage the expected utility $EU_{iz(i)}^{s2}$ takes the following form:

$$EU_{iz(i)}^{s2} \equiv \lambda \left(U_{iz(i)}^{p2} - \frac{r(\tau_i \sigma_{\eta_{iz(i)}})^2}{2} \right) + \xi_{iz(i)}$$

Now the uncertainty-penalization term depends on the variance of the distribution of $\tau_i \times \eta_{iz(i)}$, the shrunk noise component. Placed families ($z(i) > 0$) learn the remaining unknown part of the outside option ξ_{i0}, and decide to attend the offered school if $EU_{iz(i)}^{s2} > \lambda U_{i0} + \xi_{i0}$, or take the outside option otherwise.

The modeling definitions of stage 2 imply that higher uncertainty about $z(i)$ or facing a more valuable outside option decreases the probability of enrollment.

V. Bringing the Model to the Data

Our goal is to recover the parameters of the model of joint decisions of school choice and enrollment in placement offers. To achieve this, we use the observed set of applicants with their ROL, placement results, enrollment, and survey data.[26]

[24] Proof and details of the effect of the valuation of outside option on search on Appendix F.

[25] Assuming that search cost κ_i is the cost of getting to know a school, but not necessarily being fully informed about it.

[26] We will not estimate parameters of the school search and portfolio construction, since we only observed the result of this process (ROLs), and don't have data on the sequential process described. Also, we didn't collect survey data on beliefs of options not considered (to approximate $F_i(EU_{is}^{s1}, p_{is})$), nor related to search costs (κ_i). We can't identify them without imposing heroic assumptions.

In our model, families decide the ROL by comparing the enrollment option utility w_{ij} between the schools on their choice set Ω_i. Hence, ROLs provide multiple pseudo-choices per applicant of the form $w_{ir} > w_{ij}$, $j \in \Omega_i \backslash \{1 \dots r\} \, \forall r \in \mathcal{C}_i$ (Train, 2009). If we take the rank-preserving transformation $g_i(x) = \frac{1}{\lambda} \log(\exp(x) - \exp(\lambda U_{i0}))$ to each w_{ij}, we get a similar relation but based on EU^{s1} instead of w: $EU^{s1}_{ir} > EU^{s1}_{ij}$, $j \in \Omega_i \backslash \{1 \dots r\} \, \forall r \in \mathcal{C}_i$.[27]

As we detailed in Section IV.B, EU^{s1}_{ij} depends on the perceived utility U^{p1}_{ij} and the uncertainty penalization $\frac{r\sigma^2_{\eta_{ij}}}{2}$. The perceived utility comprises the *true* utility plus the noise. We don't observe the noise, so we model it as a random component. We follow a standard practice on the school choice literature (Agarwal and Somaini, 2020), and distinguish between the observed V_{ij} to the unobserved (for us) part of the utility $\epsilon_{ij} \overset{\text{iid}}{\sim} EVI$. Putting together these definitions we get to $EU^{s1}_{ij} = V_{ij} + \eta_{ij} - \frac{r\sigma^2_{\eta_{ij}}}{2} + \epsilon_{ij}$. Conditioning on the random terms, it gives us an analytical expression for the probability of the pseudo-choices $r \in \mathcal{C}_i$ (McFadden, 1974):

$$P(w_{ir} > w_{ij}, \forall j \in \Omega_i \backslash \{1 \dots r\} | \boldsymbol{\beta}^{\sigma}, \boldsymbol{\eta}) = \frac{\exp\left(V_{ir} + \eta_{ir} - \frac{r\sigma^2_{\eta_{ir}}}{2}\right)}{\sum_{j \in \Omega_i \backslash \{1 \dots r-1\}} \left(V_{ij} + \eta_{ij} - \frac{r\sigma^2_{\eta_{ij}}}{2}\right)}$$

The likelihood of the ROL $\mathcal{C}_i$ is the product of the probability of the individual pseudo-choices (Beggs et al., 1981).

The second choice in our model is the decision to enroll in the placement offer $z(i)$ or take the outside option. Families enroll in $z(i)$ if the expected utility at stage 2, $EU^{s2}_{iz(i)}$ [28], is greater than the utility derived from the outside option, $\lambda U_{i0} + \xi_{i0}$. Given that both utilities have a component that follows a EVI distribution, once we condition on the random terms $\epsilon_{iz(i)}, \boldsymbol{\beta}^{\sigma}, \boldsymbol{\eta}$ the probability of $EU^{s2}_{iz(i)} > \lambda U_{i0} + \xi_{i0}$ also has an analytic expression:

$$P(EU^{s2}_{iz(i)} > U_{i0} + \xi_{i0} | \epsilon_{iz(i)}, \boldsymbol{\beta}^{\sigma}, \boldsymbol{\eta}) = \frac{1}{1 + \exp\left(\lambda V_{i0} - \lambda\left(V_{iz(i)} + \tau \times \eta_{iz(i)} - \frac{r(\tau\sigma_{\eta_{iz(i)}})^2}{2} + \epsilon_{iz(i)}\right)\right)}$$

In the following subsections, we detail how we map data with the different components of the enrollment option utility EU^{s1}_{ij}, the outside option utility U_{i0}, and the choice set definition Ω_i. We conclude the section with the likelihood function that we maximize to retrieve our estimated parameters.

[27] The intuition of why the transformation $g_i(\dot{})$ works is the fact that for applicant i the outside option of each choice j is the same.

[28] As a reminder, $EU^{s2}_{iz(i)}$ (stage 2) is different to $EU^{s1}_{iz(i)}$ (stage 1) in three aspects. (1) the uncertainty penalization depends on the variance of $\tau \times \eta_{iz(i)}$ instead of $\eta_{iz(i)}$, which allows potential learning. (2) it includes a new preference shock $\xi_{iz(i)} \sim EVI$, and (3) the utility but $\xi_{iz(i)}$ is scaled by λ, to allow $\epsilon_{iz(i)}$ and $\xi_{iz(i)}$ have different variances.

V.A. Observed utility of schools (V_{ij})

We follow the literature and assume a linear functional form for the observed portion of the utility $V_{ij} = v_i(W_j, D_{ij}, X_i)$. W_j is a vector of characteristics of school j that includes number of grades offered, mean size of cohorts, fee[29], fraction of enrolled low SES students, a dummy if is charter (private with public subsidy), math test score, and language test score. D_{ij} is a matrix of individual-specific school attributes that includes distance to the school, and a dummy if there is a sibling enrolled. X_i as a vector of applicant i's attributes that includes a dummy for low SES, female, voluntary change, application year, and school level.[30]

$$\begin{aligned} V_{ij} &= (\boldsymbol{\gamma} + \boldsymbol{\gamma}_X \boldsymbol{X}_i)\boldsymbol{D}_{ij} + (\boldsymbol{\beta} + \boldsymbol{\beta}_X \boldsymbol{X}_i + \boldsymbol{\beta}_i^{\sigma})\boldsymbol{W}_j + \zeta_j \\ &= (\boldsymbol{\gamma} + \boldsymbol{\gamma}_X \boldsymbol{X}_i)\boldsymbol{D}_{ij} + (\boldsymbol{\beta}_X \boldsymbol{X}_i + \boldsymbol{\beta}_i^{\sigma})\boldsymbol{W_j} + \delta_j \end{aligned}$$

To account for the fact that families may differ on how important those characteristics are in their valuation, the functional form of V_{ij} incorporates observable and unobservable preference heterogeneity. The vector of parameters γ_X and β_X represent the observed preference heterogeneity. The unobserved heterogeneity is captured by the random component $\boldsymbol{\beta}_i^{\sigma}$, which we assume comes from a normal distribution with diagonal variance matrix Σ_{β}. The term δ_j represents the components of the indirect utility of school j that are equally perceived among applicants: the common preference over schools' observed attributes $\boldsymbol{\beta W}_j$, and the unobserved attributes summarized on ζ_j.

V.B. Noise term (η_{ij})

In our model, families cannot distinguish between the noise and the *true* utility. They have a sense of how big this noise could be, reflected in the belief about the second moment of its distribution. We also don't observe the noise, so from an econometric perspective, it is a random term. We use our survey responses to the question "How well do you know the school" to map the noise term to one of three distributions. Denoting $k(i,j) \in \{1 \dots 3\}$ the answer of applicant i about school j, we assume that the distribution of η_{ij} in stage 1, is the following:

$$\eta_{ij} = \eta_{k(i,j)} = \begin{cases} \eta_1 \sim N(0, \sigma^2_{\eta 1}) & \text{if } k(i,j) = 1 : \text{“I don't know it”} \\ \eta_2 \sim N(0, \sigma^2_{\eta 2}) & \text{if } k(i,j) = 2 : \text{“I know it by name”} \\ \eta_3 \sim N(0, 0) & \text{if } k(i,j) = 3 : \text{“I know it well”} \end{cases}$$

[29] Only 21% of the schools that participate in the centralized admission system charge fee, and they represent 25% of the enrollment among participants schools in 2022. Details on Table B.II in Appendix B.

[30] School level in our context does not map perfectly to the U.S. system of elementary, middle, and high school. Chile has a system of pre-básica (4 to 5 years old), básica (6 to 14 years old), and media (15 to 18 years old). We use the divisions 4 to 5, 6 to 12, and 13 to 18 years old.

This implies that families that answered "I know it well" face zero noise for that particular school, hence the perceived utility is equivalent to the true utility.

The uncertainty enters into the expected utility in stage 1 through the uncertainty-penalization term $\frac{r\sigma^2_{\eta_{ij}}}{2}$, where $\sigma_{\eta_{ij}}$ is the belief that family i about the second moment of the noise distribution, and r is the risk parameter of the CARA Bernoulli utility function. We are not able to separately identify r from $\sigma_{\eta_{ij}}$, hence we will estimate the parameters ρ^{s1}_1 and ρ^{s1}_2 that represents the uncertainty-penalization term:

$$\frac{r\sigma^2_{\eta_{ij}}}{2} = \rho^{s1}_{k(i,j)} = \begin{cases} \rho^{s1}_1 = \frac{r\sigma^2_{\eta_1}}{2} & \text{if } k(i,j) = 1: \text{ "I don't know it"} \\ \rho^{s1}_2 = \frac{r\sigma^2_{\eta_2}}{2} & \text{if } k(i,j) = 2: \text{"I know it by name"} \\ \rho^{s1}_3 = 0 & \text{if } k(i,j) = 3: \text{"I know it well"} \end{cases}$$

And since the beliefs of the variance of the noise in stage 2 may change, we will estimate ρ^{s2}_1 and ρ^{s2}_2 as the uncertainty-penalization terms included in the enrollment decision ($\rho^{s2}_3 = 0$).

V.C. Choice Sets (Ω_i)

Identification of the parameters that define the indirect utility function relies on comparing attributes of chosen options with other considered ones, and that requires a choice set definition (Agarwal and Somaini, 2020). Unfortunately, we do not observe applicants' complete choice sets Ω_i, but only the subset they applied to ($\mathcal{C}_i \subset \Omega_i$). A common approach in the school choice literature is to assume that choice sets are composed of all schools available within a limited geographic zone where applicants live (Abdulkadiroglu et al., 2020; Ainsworth et al., 2023; Bodéré, 2023, for example). Based on the limited awareness about options that families declared, this approach will likely result in biased estimates, because we would consider irrelevant alternatives (McFadden, 1978). We are on the grounds of heterogeneous unobserved choice sets (Crawford et al., 2021).

There is another reason to be cautious when constructing the choice set, also related to the risk of including irrelevant alternatives. Researchers have observed in the field and proposed theoretical reasons for the behavior of omitting viable options from rankings in settings with strategy-proof mechanisms. For instance, in the context of Mexico, Chen and Sebastián Pereyra (2019) found that some high-school applicants choose not to apply to certain desirable schools. Our model aligns with this finding, suggesting that if the subjective placement probability for an attractive school is perceived as zero, then including it in the list bears no value. This notion echoes the argument posited by Haeringer and Klijn (2009). Adopting a different perspective, Meisner and von Wangenheim (2023) rationalizes the decision of not including a preferred but highly popular alternative in the ranking through expectation-based loss aversion. They argue that potential disappointment may play an essential role in the application decision. Fack et al. (2019) acknowledge this fact in a scenario where limited rankings create even stronger incentives to deviate from truth-telling, and rely on stability to estimate preference, arguing is a more robust assumption.

As attempts to overcome the problem of unobserved choice sets, we estimate our model with two choice sets of definitions. First, and as our main specification, we define the choice set as the schools on the ROLs ($\Omega_i = \mathcal{C}_i$). This guarantees that we are inferring preferences from real trade-offs that families make. The downside is that we can only use applications with more than one school ($|\mathcal{C}_i| > 1$), and we need to rely on more assumptions since the inclusion of unobserved taste heterogeneity and random noise in our choice model framework brakes the independence of irrelevant alternatives (IIA) property of plain Logits (Guevara and Ben-Akiva, 2013).[31] This empirical approach of using subsets of the true choice choices has been labeled "differencing out" (Crawford et al., 2021) and was pioneered in McFadden (1978).

The second approach is based on the alternative specific consideration (ASC) model started by Manski (1977).[32] This procedure has been used in economics (Manzini and Mariotto, 2014; Kawaguchi et al., 2021) and marketing (Swait and Ben-Akiva, 1987; Ben-Akiva and Boccara, 1995; van Nierop et al., 2010) to estimate preferences with heterogeneous unobserved choice sets. The process requires integration over all the potential choice sets that contain the choices, becoming computationally infeasible with many options (Abaluck and Adams-Prassl, 2021; Crawford et al., 2021). For settings like ours, they suggest to follow a simulated choice sets approach implemented by Sovinsky Goeree (2008), who estimated a demand model for home PCs, in a universe with 2112 options and unobserved choice sets. Our approach is similar to hers.

The method uses simulation to approximate the integration over all potential choice sets. The procedure starts by calculating a consideration probability $\hat{p}^c_{ij}$ for each potential option $j \in \{1 \ldots J_i\}$ of applicant i. On each simulation s, we draw J_i uniform random variables u_{ijs}, for all i. The inclusion of alternative j on the *simulated* choice set of i in simulation s is defined by the Bernoulli variable $b_{ijs} = \mathbb{1}(\hat{p}^c_{ij} > u_{ijs})$. Our approach differs to Sovinsky Goeree (2008) in how we calculate $\hat{p}^c_{ij}$. In Sovinsky Goeree (2008) the probabilities are calculated endogenously using advertisement measures as consideration shifters that don't affect choice probabilities.[33] We use our survey data to estimate the consideration probability in a previous step, approximating consideration with answers to our questions about knowledge of schools not in the ranking but in the neighborhood. The procedure is detailed in Appendix C.[34]

[31] Abdulkadiroglu et al. (2020) use this approach as a robustness check of their main specification (Table A10), in a context with Logit models estimated at granular levels. For Mixed Logit there seems to be no exact procedures to estimate consistent parameters from subsets of true choice sets. Citing Crawford et al. (2021): "To the best of our knowledge, in the context of cross-sectional, results of this kind (estimating discrete choice models from subsets of true choice sets) are not available for mixed logit models with continuous distributions of random coefficients, even though some interesting approximations have been proposed by Keane and Wasi (2013) and Guevara and Ben-Akiva (2013)."

[32] The method is also labeled as "integrating over approach" in Crawford et al. (2021) or "ARC" in Barseghyan et al. (2021). Abaluck and Adams-Prassl (2021); Barseghyan et al. (2021) describe it and derive identification results.

[33] Abaluck and Adams-Prassl (2021) proves that parameters of the consideration and choice model are identified even without the need for a consideration probability shifter that is excluded from the choice model.

[34] The results using this method are not included in this version of the paper. I expect to have a version by mid-December.

V.D. Utility of the outside option (U_{i0})

In our model, the outside option is student-specific. The observed utility of the outside option depends on the number of alternatives around the area of interest of the applicant. We define the area of interest as the centroid of the schools that the student is applying to. We distinguish between four types of schools, counting only alternatives that offer education for the grade and gender of the student. First, count the density of schools that participated in the centralized platform, as a measure of the richness of the process, which we expect to be negatively related to the value of the outside option. The second set is the publicly funded schools that are not part of the centralized system. Those include schools that only offer elementary education (PK to K), which, for idiosyncratic reasons, are not part of the centralized offer, and the schools with ad-hoc admission processes as artistic or sport-focused. Third, we group the schools that provide education for students with special language needs.[35] And fourth, we count the number of fully private schools around the centroid.

We allow for observable preference heterogeneity. Low SES families and applicants that are voluntarily changing schools[36] may asses differently the outside options. We interact with the availability of each type of alternative with dummies that reflect if the student has low SES and if she is applying voluntarily.

There are two additional attributes that we include in the outside option but are genuinely a characteristic of the placement. The first is related to whether siblings were placed in the same school. For each applicant i, we checked if the same guardian filed an application for a sibling of i and if the ROL of the sibling had any overlap with $\mathcal{C}_i$. Then, we checked if the applicant got placed in the same school as the sibling. If families prefer to have their siblings in the same school, then being placed in different should reduce the likelihood of enrolling in the placed school. Second, we add a dummy for the placement ranking to count for potential behavioral motives of non-enrollment related to disappointment (Meisner and von Wangenheim, 2023).

V.E. Likelihood function

The individual likelihood of the joint decision of school choice and enrollment for an applicant that enrolls in her placement offer takes the following form:[37]

[35] Schools that offer a curriculum for students with language problems have a different admission process that lets them screen the applicants based on their disability level and are not part of a centralized system. Anecdotal evidence suggests that screening is not very rigorous, and families could easily get a medical certificate that will allow them to apply.

[36] We define an applicant as voluntary if she is currently enrolled in a school that offers the next grade.

[37] The likelihood for applicants with placement but decides not to enroll is very similar but, with $s(i) \neq z(i)$ instead of $s(i) = z(i)$, which implies $EU^{s2}_{iz(i)} < U_{i0} + \xi_{i0}$. For applicants that are not placed ($z(i) = 0$) the likelihood is just $P(\mathcal{C}_i)$.

$$
\begin{aligned}
L_i =& P(\mathcal{C}_i \wedge s(i) = z(i)) \\
=& P(w_{ir} > w_{ij}, \forall j \in \Omega_i \backslash \{1 \ldots r\}, \forall r \in \mathcal{C}_i \wedge EU^{s2}_{iz(i)} > U_{i0} + \xi_{i0})
\end{aligned}
$$

Once we condition on the random components $\boldsymbol{\beta}^{\sigma}, \boldsymbol{\eta}, \epsilon_{iz(i)}$, and integrate over its distribution, we get the following expression:

$$
\begin{aligned}
L_i =& \int P\left(w_{ir} > w_{ij}, \forall j \in \Omega_i \backslash \{1 \ldots r\}, \forall r \in \mathcal{C}_i | \boldsymbol{\beta}^{\sigma}, \boldsymbol{\eta}\right) \times \\
& \qquad \left(\int P\left(EU^{s2}_{iz(i)} > U_{i0} + \xi_{i0} | \epsilon_{iz(i)}, \boldsymbol{\beta}^{\sigma}, \boldsymbol{\eta}\right) \mathrm{d}F(\epsilon_{iz(i)} | \boldsymbol{\beta}^{\sigma}, \boldsymbol{\eta}) \right) \mathrm{d}F(\boldsymbol{\beta}^{\sigma}, \boldsymbol{\eta}) \\
=& \int \left(\prod_{r \in \mathcal{C}_i} \frac{\exp\left(V_{ir} + \eta_{k(i,r)} - \rho^{s1}_{k(i,r)}\right)}{\sum_{j \in \Omega_i \backslash \{1 \ldots r-1\}} \exp\left(V_{ij} + \eta_{k(i,j)} - \rho^{s1}_{k(i,j)}\right)} \times \right. \\
& \left. \int \frac{1}{1 + \exp\left(\lambda V_{i0} - \lambda\left(V_{iz(i)} + \tau \times \eta_{k(i,z(i))} - \rho^{s2}_{k(i,z(i))} + \epsilon_{iz(i)}\right)\right)} \mathrm{d}F(\epsilon_{iz(i)} | \boldsymbol{\beta}^{\sigma}, \boldsymbol{\eta}) \right) \mathrm{d}F(\boldsymbol{\beta}^{\sigma}, \boldsymbol{\eta})
\end{aligned}
$$

The log-likelihood function is defined as $ll = \sum_{i=1}^{I} \log(L_i)$. We estimate the parameters via simulated maximum likelihood (Train, 2009), following the standard procedure except for the generation of draws for $\epsilon_{iz(i)}$. Given that the Deferred Acceptance algorithm places each student in her most preferred school where her lottery is over the cutoff, it's very likely that the unobserved part of the placed school $\epsilon_{iz(i)}$ is not $iid\,EVI$.[38] This is because schools with higher unobserved utility ϵ_{ij} are more likely to be ranked at the top of the list. To generate approximate draws from the distribution of $\epsilon_{iz(i)} | \boldsymbol{\beta}^{\sigma}, \boldsymbol{\eta}$, we follow a 2-step procedure that we described in detail in Appendix E.

VI. Results

We now turn to describe the estimated parameters of the model detailed in Section IV. We use the Simulated Maximum Likelihood procedure over the function defined at the end of Section V. First, we show the results of the choice process that describe the weights on school characteristics defining the indirect utility function of school enrollment. Then, we go over the attributes that define the valuation of the outside option. Finally, we review the estimated parameters related to the noise faced by families with limited knowledge about the options they were applying to.

VI.A. Weights on school attributes

The estimates of the weights on school characteristics that define the indirect utility function of school attendance are shown in Table II. Following Abdulkadiroğlu et al. (2017), all estimates

[38] As Abdulkadiroğlu et al. (2017) points out, the assumption on the relationship between ranking and utilities restricts the values of the unobserved terms.

are relative to the effect of 1 mile for the $\boldsymbol{X}_i = \mathbf{0}$ student (male, mid-high-SES, non-voluntary, applying to elementary school in 2020), so it could be interpreted as the willingness to travel. For example, the $\boldsymbol{X}_i = \mathbf{0}$ student values having a sibling in the school as much as having the school 7.39 miles closer. The model in Abdulkadiroğlu et al. (2017) assumes a common distaste for distance while we allow for observed preference heterogeneity, which makes interpretation less direct. As an example, in our case, an extra mile for the $\boldsymbol{X}_i = \mathbf{0}$ student affects the valuation of a school $-1/(-1+0.57) = 2.3$ times more than for his equivalent who is applying to high school (HighSch$=1$).[39]

From the first row of Panel A, we observe that families dislike schools that are further away from home, and high school applicants give less than half the importance than younger applicants. Females penalize distance marginally more than male applicants, same relation for applicants of 2021 and 2022 compared with 2020.[40] The second row shows that families strongly prefer schools with siblings, but the relevance decreases in upper grades.

Applicants seem to put different weights on school characteristics. Panel B of Table II reflects this heterogeneity. Female students then put more weight on math test scores and less on language test scores compared with males. Low SES applicants have higher taste intensity for charter and schools with larger enrollment, and pay less attention to language test scores. Voluntary applicants prefer schools that offer more grades, are more sensitive to tuition fees, and care more about language test scores than non-voluntary applicants. For high-schoolers the size of the school is more relevant -number of grades and enrollment- they care less about the SES of the student body or math test scores and are more inclined towards charter school than applicants to lower grades.

[39] As an example, the additional willingness to travel for attend a Charter school for the last described high school applicant is $-(0.184)/(-1+0.57) = 0.3$.

[40] As a response to the COVID-19 pandemic, classes in Chile were fully remote from mid-March 2020 to June 2021. Applicants of the 2021 process experienced a partial return to in-person classes, while for 2022 applicants, all schools had mandatory in-person teaching. These extraordinary experiences might have influenced how families gathered information and applied to schools.

Table II
School Choice Estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	γ	Observable heterogeneity (γ_X)						
		Female	Low SES	Voluntary	MidSch	HighSch	2021	2022
Distance (1 mile)	-1.000	-0.060	0.091	-0.051	0.011	0.564	-0.119	-0.081
	(0.026)	(0.020)	(0.021)	(0.024)	(0.032)	(0.026)	(0.025)	(0.024)
Sibling	7.390	-0.097	-0.836	0.019	-1.780	-3.451	0.359	-0.572
	(0.167)	(0.139)	(0.140)	(0.210)	(0.182)	(0.176)	(0.176)	(0.164)
	σ_β	Observable heterogeneity (β_X)						
		Female	Low SES	Voluntary	MidSch	HighSch	2021	2022
# of grades offered		-0.008	0.047	0.133	-0.047	0.524	0.076	0.133
		(0.040)	(0.040)	(0.047)	(0.076)	(0.104)	(0.048)	(0.046)
Fee		-0.006	-0.044	-0.155	0.000	0.059	0.062	0.019
		(0.029)	(0.030)	(0.034)	(0.039)	(0.039)	(0.035)	(0.033)
Share of low SES		-0.087	0.191	-0.138	0.151	0.438	-0.061	-0.136
		(0.050)	(0.053)	(0.061)	(0.069)	(0.071)	(0.062)	(0.058)
Charter		-0.057	0.135	-0.030	0.123	0.184	-0.053	-0.127
		(0.041)	(0.042)	(0.050)	(0.059)	(0.067)	(0.050)	(0.047)
Enrollment per grade		0.006	0.063	0.037	-0.015	0.182	0.049	-0.035
		(0.027)	(0.026)	(0.029)	(0.052)	(0.050)	(0.031)	(0.031)
Math test score	1.578	0.189	-0.007	0.020	0.086	-0.481	-0.097	-0.006
	(0.044)	(0.073)	(0.074)	(0.086)	(0.104)	(0.112)	(0.087)	(0.083)
Language test score		-0.139	-0.179	0.123	-0.245	-0.089	-0.027	0.013
		(0.062)	(0.063)	(0.073)	(0.091)	(0.096)	(0.075)	(0.071)

Notes. Estimates of the parameters that define the observed utility of enrolling in a school. Column 1 contains estimates of common preference for school characteristics (upper panel) or unobserved heterogeneous preference for school attributes (lower panel). Columns 2 to 8 contain parameters that reflect preference heterogeneity by applicants' attributes (columns) for schools' characteristics (rows). Distance is calculated as the Euclidean distance between the home address and the school. "Sibling" indicates having a sibling enrolled in the school. Math and language tests are standardized national-level tests. "Low SES" is a socio-economic status measure computed by Mineduc, representing roughly families in the poorest tercile. "'Voluntary" indicates students applying from a school where they could continue studying. "MidSch" and "HighSch" are students applying to 1st to 6th and 7th to 12th grade, respectively. Standard errors in parentheses.

VI.B. Noise, uncertainty penalization and learning

Table III shows the estimates for the noise distribution. Families that declared the lowest level of knowledge about the school ($k(i, j) = 1$) perceive a utility that has a noisy component with a 35% larger standard deviation than applicants that answer the middle level of knowledge ($k(i, j) = 2$). Estimates suggest that the noise, compared to the people who declared that they know the school well ($k(i, j) = 3$), is relevant. A noise realization from percentile 20 of the distribution is equivalent to moving the school 1.7 miles further from home (or closer if the noise comes from percentile 85).

Families's beliefs about the noise variance are coherent with our estimates of the actual vari-

Table III
Noise standard deviation and uncertainty penalization estimates

Noise	Stage 1 η	Stage 2 $\tau \times \eta$
A. Standard deviation of noise		
σ_1	-5.768	4.404
	(0.146)	
σ_2	-4.251	3.246
	(0.087)	
B. Uncertainty-penalization		
$-\rho_1$	-8.169	-10.375
	(0.087)	(0.444)
$-\rho_2$	-5.226	-6.097
	(0.055)	(0.273)
C. Other parameters		
λ		0.528
		(0.019)
τ		0.764
		(0.069)

Notes. Panel A: estimates of the standard deviation of the noise distribution faced by families with imperfect knowledge. Panel B: estimates of the uncertainty penalization terms that affect the valuation of schools. Panel C: estimates for λ and τ. λ is the ratio of standard deviations of the unobserved portion of $EV^{s1}_{iz(i)}$ and the preference shock realized in stage 2 ($\xi_{iz(i)}$. τ reflects the shrinkage of the noise distribution from stage 1 to stage 2. Standard errors in parentheses. Estimates in column 2 of Panel A are the product of column 1 and τ.

ance of the noise. Remember that $\rho_1^{s1} = \frac{\sigma_{\eta 1}^2}{2}$ is the uncertainty-penalization term that comes from families taking the expectation over the CARA Bernoulli utility function. If beliefs of families are correct, then ratio $\frac{\rho_1}{\rho_2} = 1.6$ (or 1.7 for stage 2) should be similar to the ratio of the variances, that is 1.8.

The estimated shrink parameter τ is 0.71, suggesting that the dispersion of the noise is reduced by 30% from stage 1 to stage 2, which we interpret as learning. At the same time, we also observe that the penalization term, if any, increases from one stage to the other. Our model predicts that it should shrink at $\tau^2 = .55$ rate. We provide two hypotheses for this result. First, it could be that families' beliefs about σ_η do not update at the same rate as the shrink of the true variance of η. Second, the stakes in this stage are higher since it is a decision to enroll instead of a decision to get an option to enroll; hence the aversion to uncertainty might be higher. A value of the CARA risk parameter r that doubles from one stage to the other will rationalize the estimated uncertainty-penalization terms, assuming beliefs about σ_η are correct.x

VI.C. Valuation of the outside option

The compliance model compares the updated expected utility of enrolling in the assigned school $EU_{iz(i)}^{s2}$ with the value of the outside option $U_{i0} + \xi_{i0}$. The former is defined as the perceived utility at the moment of application but allowing the noise present at the application stage to shrink (at rate τ), representing potential learning about the school done by the family after placement.

The value of the outside option U_{i0} depends on the number of alternatives that the family faces in the application platform and outside. Table IV shows families that face more availability of schools in the centralized application platform, in-system school, are less likely to decline the placement offer. The presence of out-of-system schools has the opposite effect. Mid and high-SES non-voluntary applicants that live around private schools that don't receive voucher value more the outside option. The same happens to families with more out-of-system public schools. Preschool language schools are especially valuable for low-SES families.

When family members apply together, being placed in the same school matters. 27% of the guardians are responsible for two or more students with at least one school in common between applications. We observe that applicants placed without other family members are less likely to enroll in placement. The effect is equivalent to moving the placed school 7.5 miles away for students applying to elementary school and 6.0 miles away for middle school applicants. Interestingly, the effect vanishes for high school applicants.

VII. COUNTERFACTUALS

We design counterfactual scenarios to approximate the congestion cost imposed by non-compliers, and to explore the role of outside options and information on the non-compliance behavior. We start with simple scenarios, in which we run the assignment algorithm after dropping the prefer-

Table IV
Outside option estimates

Value of outside option	(1) ϕ	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Observable heterogeneity (ϕ_X)						
		Female	Low SES	Voluntary	MidSch	HighSch	2021	2022
Constant		-0.190	-0.493	3.487	-1.284	-2.318	0.418	2.177
		(0.185)	(0.192)	(0.294)	(0.439)	(0.404)	(0.233)	(0.222)
A. Concentration of options around home address								
In-system schools	-0.981	-0.174	0.175	-0.014	0.980	1.052	-0.220	0.442
	(0.266)	(0.172)	(0.176)	(0.210)	(0.276)	(0.275)	(0.217)	(0.206)
Out-of-system private	0.363	0.142	-0.900	-1.243	0.707	0.315	0.102	0.590
	(0.244)	(0.180)	(0.205)	(0.213)	(0.265)	(0.231)	(0.227)	(0.215)
Out-of-system preschool language	0.335	-0.111	0.437	0.190			0.152	-0.282
	(0.252)	(0.185)	(0.189)	(0.419)			(0.234)	(0.222)
Out-of-system public	0.756	-0.198	0.189	-1.753	0.096	-0.764	-0.137	-0.055
	(0.190)	(0.179)	(0.186)	(0.380)	(0.457)	(0.400)	(0.226)	(0.213)
B. Placement outcomes								
Placed without sibling	5.958	0.715	-1.628	-2.305	0.170	-2.491	1.534	0.754
	(0.812)	(0.699)	(0.710)	(0.918)	(0.905)	(0.992)	(0.941)	(0.824)

Notes. Estimates of the parameters that define the utility of the outside option. Column 1 contains estimates of common preferences for characteristics of the outside option. Columns 2 to 8 contain parameters that reflect preference heterogeneity by applicants' attributes (columns) for outside option's characteristics (rows). Panel A shows the parameters related to the number of schools available in a radius of 1.2 miles from the home address. Panel B includes placement outcomes different from the placed school that affect the enrollment on placement decision. "Sibling" indicates having a sibling enrolled in the school. Math and language tests are standardized national-level tests. "Low SES" is a socio-economic status measure computed by Mineduc, representing roughly the poorest tercile of families. "'Voluntary" indicates students applying from a school where they could continue studying. "MidSch" and "HighSch" are students applying to 1st to 6th and 7th to 12th grade, respectively. "Placed without sibling" refers to applicants who applied with a sibling but were assigned to different schools. We also included Standard errors in parentheses.

ences that non-compliers will not enroll. We continue with model-based counterfactuals, where we evaluate the changes in allocation when an information campaign is implemented, varying its effectiveness and the inclusion of out-of-system alternatives in the centralized process.

VII.A. Baseline scenario

In order to calculate changes in utility-based measures for our counterfactual, we first construct a baseline scenario that emulates the application and assignment (stage 1 of the model) and enrollment decision (stage 2). Since we don't estimate participation or a search model, our starting point is the original set of applicants, and we assume that their choice set is the set of schools on their ROLs. We construct the observed portion of expected utility (V_{ij}) using the model's estimated parameters, schools, and applicants' characteristics (W_j and X_i). We predict the level of knowledge $k(i,j)$ to map to a specific distribution of noise ($\eta_{k(i,j)}$) and uncertainty penalization ($\rho_{k(i,j)}$), and we simulate the unobserved portions of the utility (ϵ_{ij}). With those inputs, we calculate the expected utility (EU_{ij}^{s1}) of each school, and build the ROLs ($\mathcal{C}_i$).

After obtaining the ROLs, we proceed to run the Deferred Acceptance algorithm using the real school capacities, and we recover the placements $z(i)$ of the students. We then construct the utility of the outside option U_{i0}, simulate the preference shocks realized in stage 2 ($\xi_{iz(i)}$ and ξ_{i0}), and generate compliance decisions ($EV_{iz(i)} > U_{i0} + \xi_{i0}$?).[41]

Our simulated baseline scenario mimics the real scenario in two relevant dimensions. First, the percentage of students assigned to preferences or not assigned is almost identical. Second, the compliance rate is similar, differing only by 1 pp. The good fit holds when we disaggregate the measures at the urban zone level.[42]

Next, we describe the counterfactuals that we will compare with the baseline scenario. We start with the simple exercise of calculating the allocation if non-compliers do not apply to the placed school, and we move to model-based counterfactuals, where we simulate an information campaign and changes in the incorporation of out-of-system outside options in the centralized process.

VII.B. Mechanical counterfactuals

Families that don't comply with the placement offer need to apply to a school without the convenience of the centralized system, which may be an optimal decision given the new information acquired and/or the presence of out-of-system schools. But non-compliance also generates a negative externality on other families that would have preferred the school assigned to a non-complier than their own placement.

We would like to know the hypothetical placement if assigned applicants who do not enroll

[41] We run the entire process 100 times. We provide details of the construction of the indirect utility and overall simulation in Appendix D. Details of the function that predicts the level of knowledge can be found in Appendix G.1.

[42] For the counterfactuals we consider 70 urban zones, omitting very small ones not included in the estimation sample of our main model.

on the placement offer would not have applied to that school. Since we observe preferences, placement, and compliance, we can replicate the assignment using the allocation rules of the implemented version of the Deferred Acceptance in Chile. We can evaluate any change in the ROLs and compare the placement outcome with the original.

We evaluate two changes: (1) when non-compliers do not apply to the school they were assigned, and (2) when they don't apply to the assigned school or any preference below. These counterfactuals tell us how much externalities non-compliers impose on the rest of the applicants. Our group of interest is students with room for improvement: applicants who were placed on second or lower preference and complied with the offer or applicants who were not placed in any preference (46% of all the applicants).

VII.C. Information campaign

Ideally, we would like to simulate the effect of policies oriented to have more informed families on application and enrollment decisions. But standard policies implemented in school choice settings such as information campaigns (Hastings and Weinstein, 2008; Allende et al., 2019) or modifications in the market design that change the cost of non-compliance (a "tax" for non-compliance) would change the incentives or costs of search, potentially inducing changes in the composition of the choice set. Our model does not allow us to predict that kind of behavioral response.

Instead, we are going to use the fact that we the observed outcome of the search process that noncompliers do: the school where they end up enrolled after dismissing the centralized placement offer. If applicants had applied to the enrolled school in the first place, then compliance would have been less of a problem.

We introduce a counterfactual policy wherein an information campaign endeavors to anticipate and inform applicants of the schools they are most likely to enroll in, which we call school $q(i)$. In the ideal scenario, referred to as the "oracle campaign," the prediction function exhibits perfect accuracy, akin to having foresight into the applicants' eventual enrollment choice $s(i)$. However, recognizing the practical impossibilities of such precision, we incorporate "prediction errors" to mirror real-world unpredictability. We do this by varying the "prediction accuracy" denoted by $\alpha \in [0, 1]$, which reflects the percentage of families that we correctly predict the enrollment decision. For the $1 - alpha$ fraction of applicants, we give a "naive recommendation": the most popular feasible school within 2 miles that was not included in the ranking.

When recommending a school, families form beliefs about it and decide if they will add it to its ranking. When the recommended school is the enrolled school ($q(i) \equiv s(i)$), we exploit a revealed preference argument to approximate its expected utility. If the enrolled school $s(i)$ is preferred to the outside option, then the expected utility at stage 2 of the former ($s(i)$) has to be greater or equal to the utility of the outside option $\lambda U_{i0} + \xi_{i0}$. In practice, we draw the unobserved portion of the expected utility of the enrolled schools constrained to the following utility inequality: $\epsilon_{iq(i)}\, st.\, EU^{s2}_{iq(i)} > \lambda U_{i0} + \xi_{i0})$. When the suggested school is the naive recommendation, we build the observed utility with the model estimates and draw the unobserved portion from an unconditional EVI distribution.

To analyze the impact of the "intensive margin", specifically, the extent to which families are informed about school $q(i)$, we examine varying levels of familial knowledge about the school the policy recommends. We introduce a parameter, β, to quantify this variation. At one extreme, where $\beta = 1$, families possess comprehensive knowledge about school $q(i)$, expressed as $k(i, q(i)) = 3$, eliminating any uncertainty penalization in the expected utility ($EU^{s1}_{iq(i)}$). Conversely, when $\beta = 0$, families possess only the predicted level of knowledge attributed to a school that does not feature in the rankings. The function to predict this knowledge level is estimated using survey data concerning non-ranked schools, as detailed in Appendix in G.2.

VII.D. Out-of-system outside options available in centralized choice

In our last counterfactual, we add on top of the information campaign a policy that would include out-of-system publicly funded schools in the centralized platform, making them "in-system". We will call them the "included schools." 20% of the non-compliers enroll in an out-of-system school. Among those, 70% enroll in a school that receives public funding but is exempt to participate from the centralized system. This counterfactual emulates the inclusion of those schools in the system.

We need to add assumptions about school capacities and the preferences over these schools. In-system schools have to declare their seats to the centralized authorities since it is a key input of the allocation mechanism. Out-of-system schools do not need to, so we don't observe their capacity. As an approximation, we will assume the capacity of the included schools is the same as the observed enrollment, which is a lower bound for the actual number of seats.

We assume that only applicants informed about the included schools will add them to their list. By design, the information campaign is only targeted to non-compliers, making this a restrictive assumption, since having new schools in the centralized system will potentially affect the ROL of all students. We argue that our results in congestion of the counterfactual will reflect a lower bound since lifting the assumption and making them available to everybody will reduce the pressure in the original set of in-system schools. The combination of the capacity and preference assumptions results in every applicant to the included schools getting a seat with probability one.

VII.E. Counterfactual results

We aim to evaluate the externalities produced by families placed by the centralized mechanism in a school but enrolled in a different place and the effect of including out-of-system schools in the centralized mechanism. The model allows use to evaluate the effect on overall placement, but also considering the enrollment on assignment decision. This is important because the outcome that finally matters is enrollment. For example, improving the placement of a student will not be welfare-relevant if he chooses the outside option anyway. We start by showing the changes in the placement and enrollment decisions for the overall population of applicants. During the analysis, we will refer to several groups of particular interest, defined by the placement and enrollment decision at baseline: compliers assigned to 1st preference (i.e., no room for improvement), compliers assigned to 2nd or lower preference, non-placed applicants, and non-complier applicants.

Table V
Counterfactual Results for all Applicants

	(1)	(2)	(3)	(4)	(5)	(6)
	Placement			Enrollment		
	Better	Same	Worse	Better	Same	Worse
A. Mechanical counterfactuals						
Non-complier not applying to offer	0.047	0.805	0.147	0.040	0.947	0.012
Non-complier not applying to offer or lower preference	0.066	0.792	0.143	0.051	0.939	0.011
B. Model-based counterfactuals: oracle information campaign						
Oracle recommendation, full knowledge ($\alpha = 1, \beta = 1$)	0.082	0.912	0.006	0.075	0.920	0.004
Oracle recommendation, predicted knowledge ($\alpha = 1, \beta = 0$)	0.060	0.934	0.005	0.024	0.972	0.004
C. Model-based counterfactuals: naive information campaign						
Naive recommendation, full knowledge ($\alpha = 0, \beta = 1$)	0.108	0.849	0.043	0.094	0.866	0.039
Naive recommendation, predicted knowledge ($\alpha = 0, \beta = 0$)	0.070	0.894	0.035	0.048	0.920	0.032
D. Model-based counterfactuals: including out-of-system options in centralized platform						
Oracle recomendation + internalizing out-of-system ($\alpha = 1, \beta = 1$)	0.141	0.853	0.006	0.129	0.866	0.005

Notes. This table shows the changes in placement (columns 1 to 3) and the enrollment decision (columns 4 to 6), comparing counterfactuals to the baseline scenario for all applicants. Panel A contains the results for the mechanical counterfactuals (i.e. dropping preferences of non-compliers), while panel B the results for the oracle information campaign (i.e. suggesting school of future enrollment). Panel C has the results for the naive information campaign (i.e. suggesting a popular nearby school), while Panel D shows the simulation result when we incorporate out-of-system publicly funded schools into the centralized system

Results are summarized in Table V. Columns 1 to 3 show the fraction of applicants placed on a better, same, or worse preference in each counterfactual scenario. Columns 4 to 6 reflect the outcome post enrollment on placement decision. These two results could differ if, for example, an applicant gets assigned on a better preference than baseline, but in both cases, does not comply with the offer. It would be classified as "Better" placement but "Same" enrollment. In addition, we will discuss the results for the four groups of students mentioned above, results that are displayed on tables B.V to B.VI in Appendix B.

Looking at Panel A of Table V, we observe that if non-compliers do not apply to their placed schools, 5% of all applicants will improve allocation. If, in addition, non-compliers do not include any preference below their baseline placement, then 7% get a better preference. The results are mostly driven by the groups of applicants that, in baseline, were assigned to the 2nd or lower preference or were not assigned to any. Among them, the improvements are 12% and 11% respectively. Interestingly, 9% of non-compliers will also improve placement. When we focus on enrollment decisions, the results are smaller. 5% of all applicants are better off when non-compliers don't apply to their placement or any school below. Looking at the specific group, the fraction of non-compliers in a better placement preference than baseline is only 4%, less than half of the proportion assigned to a better school. The reduction between placement and enrollment improvement is less than 20% for students assigned to 2nd+ preference or not assigned to any. The difference in decline is because non-compliers are more inclined to reject any placement offer.

We observe that 15% of total applicants have a worse placement than the baseline. This is almost entirely explained by the non-compliant group, not assigned to any school after dropping their baseline offer from the rank order list. But, looking at enrollment results, only 1% of students are in a worse position under mechanical counterfactuals. This is explained by the fact that non-compliers rejected the baseline school anyway, so placement plus rejection is the same outcome for them as non-placement.

The results for the counterfactual that simulates the oracle information campaign are depicted in Panel B of Table V. This mimics non-compliers applying to the school they will enroll in the future, and they have full knowledge about them ($\beta = 1$). As a result, 8% of the total students will be placed in a better preference. Most of the gain comes from better results on the group of non-complier, 40% benefit from it. Still, 5% of the group of students with room for improvement are better placed because of less congestion in schools of their rankings. We observe that less than 1% of applicants end up in a worse school, which is an interesting result given that the oracle recommendation is in-system schools and could potentially have displaced other applicants.

When we allow the knowledge that families have to mimic the empirical knowledge of schools not included in the ranking ($\beta = 0$), the overall winners reduce from 8 to 6%. Only 12% of the target population of the policy, the non-compliers, are placed in a better school. Since the congestion alleviation is less, the effect on compliers reduces to one-fourth. These results do not translate into enrollment; the proportion of better-off applicants is only 2%. The reason behind this drop is that most of the new placements are rejected by the baseline non-compliers.

Looking at the results of the naive recommendation policy in Panel C, we observe that the fraction of better placed is higher than with the oracle campaign. This is natural since the campaign reaches every applicant, not only non-compliers. But, the proportion of students worse off is more than six times greater, giving a lower net benefit.

Panel D of Table V describes the effect when we include out-of-system publicly funded schools in the centralized platform, making them "in-system", and we implement the oracle campaign. We observe that 14% of applicants of the system would improve placement, and 13% would get a better enrollment. We can observe that there is no increase on the proportion of applicants that are worse off.

VIII. Conclusions

In this study, we have explored the effects of limited information and the availability of outside options on centralized school choice. Our empirical analysis yields three major findings. Firstly, we observed a significant lack of information among families about neighborhood schools and the ones they apply to. Secondly, non-compliance emerges as a significant issue, with 23% of applicants not enrolling in their placement offer and 70% of these non-compliers enrolling in schools they initially bypassed. Thirdly, there is a clear correlation between compliance and the level of information about a school, even when controlling for ranking and potential satisfaction.

Our theoretical model, accounting for imperfect knowledge and uncertainty aversion, leads to two crucial insights. It shows that uncertainty about schools adversely affects their perceived value, thereby decreasing compliance with placement offers. Additionally, the presence of appealing outside options is found to reduce the incentive for families to extensively search for alternatives.

Utilizing our model estimates, we simulate an information campaign aimed at informing applicants about additional schools. This policy, we find, primarily benefits non-targeted students by reducing congestion externalities from non-compliers and lowering application costs for targeted families.

Further, we evaluate the effectiveness of information campaigns based on the depth of information provided. Our findings suggest that a campaign with complete take-up but superficial information is only as effective as one with a 45% take-up rate but with comprehensive information provision.

Our research presents key challenges for policymakers. The success of information campaigns promoting new schools for application critically hinges on the depth and quality of the information provided. Moreover, including out-of-system options in centralized applications could alleviate the impact of non-compliance externalities, underscoring the importance of after-market design considerations in centralized school choice systems.

References

Abaluck, J. and Adams-Prassl, A. (2021). What do Consumers Consider Before They Choose? Identification from Asymmetric Demand Responses*. *The Quarterly Journal of Economics*, 136(3):1611–1663.

Abdulkadiroğlu, A., Agarwal, N., and Pathak, P. A. (2017). The Welfare Effects of Coordinated Assignment: Evidence from the New York City High School Match. *American Economic Review*, 107(12):3635–3689.

Abdulkadiroglu, A. and Andersson, T. (2022). School Choice. Technical Report w29822, National Bureau of Economic Research, Cambridge, MA.

Abdulkadiroğlu, A., Angrist, J. D., Dynarski, S. M., Kane, T. J., and Pathak, P. A. (2011). Accountability and Flexibility in Public Schools: Evidence from Boston's Charters And Pilots *. *The Quarterly Journal of Economics*, 126(2):699–748.

Abdulkadiroglu, A., Pathak, P. A., Schellenberg, J., and Walters, C. R. (2020). Do Parents Value School Effectiveness? *American Economic Review*, 110(5):1502–39.

Abdulkadiroglu, A. and Sönmez, T. (2003). School Choice: A Mechanism Design Approach. *American Economic Review*, 93(3):729–747.

Agarwal, N. and Budish, E. (2021). Market design. In Ho, K., Hortaçsu, A., and Lizzeri, A., editors, *Handbook of Industrial Organization*, volume 5 of *Handbook of Industrial Organization, Volume 5*, pages 1–79. Elsevier.

Agarwal, N. and Somaini, P. (2020). Revealed Preference Analysis of School Choice Models. *Annual Review of Economics*, 12(1):471–501.

Ainsworth, R., Dehejia, R., Pop-Eleches, C., and Urquiola, M. (2023). Why Do Households Leave School Value Added on the Table? The Roles of Information and Preferences. *American Economic Review*, 113(4):1049–1082.

Ajayi, K. F., Friedman, W. H., and Lucas, A. M. (2017). The Importance of Information Targeting for School Choice. *American Economic Review*, 107(5):638–643.

Allende, C., Neilson, C., and Gallego, F. A. (2019). Approximating the Equilibrium Effects of Informed School Choice. Technical Report 628.

Andrabi, T., Das, J., and Khwaja, A. I. (2017). Report Cards: The Impact of Providing School and Child Test Scores on Educational Markets. *American Economic Review*, 107(6):1535–1563.

Apesteguia, J. and Ballester, M. A. (2018). Monotone Stochastic Choice Models: The Case of Risk and Time Preferences. *Journal of Political Economy*, 126(1):74–106.

Arteaga, F., Kapor, A. J., Neilson, C. A., and Zimmerman, S. D. (2022). Smart Matching Platforms and Heterogeneous Beliefs in Centralized School Choice. *The Quarterly Journal of Economics*, 137(3):1791–1848.

Ashlagi, I., Nikzad, A., and Romm, A. (2019). Assigning more students to their top choices: A comparison of tie-breaking rules. *Games and Economic Behavior*, 115:167–187.

Barseghyan, L., Molinari, F., and Thirkettle, M. (2021). Discrete Choice under Risk with Limited Consideration. *American Economic Review*, 111(6):1972–2006.

Beggs, S., Cardell, S., and Hausman, J. (1981). Assessing the potential demand for electric cars. *Journal of Econometrics*, 17(1):1–19.

Ben-Akiva, M. and Boccara, B. (1995). Discrete choice models with latent choice sets. *International Journal of Research in Marketing*, 12(1):9–24.

Bodéré, P. (2023). Dynamic Spatial Competition in Early Education: An Equilibrium Analysis of the Preschool Market in Pennsylvania. Job Market Paper.

Budish, E. and Cantillon, E. (2012). The Multi-unit Assignment Problem: Theory and Evidence from Course Allocation at Harvard. *American Economic Review*, 102(5):2237–2271.

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

Cohen, A. and Einav, L. (2007). Estimating Risk Preferences from Deductible Choice. *The American Economic Review*, 97(3):57.

Correa, J., Epstein, R., Escobar, J., Rios, I., Bahamondes, B., Bonet, C., Epstein, N., Aramayo, N., Castillo, M., Cristi, A., and Epstein, B. (2019). School Choice in Chile. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pages 325–343, Phoenix AZ USA. ACM.

Crawford, G. S., Griffith, R., and Iaria, A. (2021). A survey of preference estimation with unobserved choice set heterogeneity. *Journal of Econometrics*, 222(1):4–43.

De Haan, M., Gautier, P. A., Oosterbeek, H., and van der Klaauw, B. (2023). The Performance of School Assignment Mechanisms in Practice. *Journal of Political Economy*, 131(2):388–455.

Deming, D. J., Hastings, J. S., Kane, T. J., and Staiger, D. O. (2014). School Choice, School Quality, and Postsecondary Attainment. *American Economic Review*, 104(3):991–1013.

Fack, G., Grenet, J., and He, Y. (2019). Beyond Truth-Telling: Preference Estimation with Centralized School Choice and College Admissions. *American Economic Review*, 109(4):1486–1529.

Grenet, J., He, Y., and Kübler, D. (2022). Preference Discovery in University Admissions: The Case for Dynamic Multioffer Mechanisms. *Journal of Political Economy*, 130(6):1427–1476.

Guevara, C. A. and Ben-Akiva, M. E. (2013). Sampling of alternatives in Logit Mixture models. *Transportation Research Part B: Methodological*, 58:185–198.

Haeringer, G. and Klijn, F. (2009). Constrained school choice. *Journal of Economic theory*, 144(5):1921–1947.

Hakimov, R. and Kübler, D. (2021). Experiments on centralized school choice and college admissions: A survey. *Experimental Economics*, 24(2):434–488.

Handel, B. R. (2013). Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts. *American Economic Review*, 103(7):2643–2682.

Hassidim, A., Marciano, D., Romm, A., and Shorrer, R. I. (2017). The Mechanism Is Truthful, Why Aren't You? *American Economic Review*, 107(5):220–224.

Hastings, J. S. and Weinstein, J. M. (2008). Information, School Choice, and Academic Achievement: Evidence from Two Experiments. *The Quarterly Journal of Economics*, 123(4):1373–1414.

Hsieh, C.-T. and Urquiola, M. (2006). The effects of generalized school choice on achievement and stratification: Evidence from Chile's voucher program. *Journal of Public Economics*, 90:1477–1503.

Immorlica, N., Leshno, J., Lo, I., and Lucier, B. (2020). Information Acquisition in Matching Markets: The Role of Price Discovery. SSRN Working Paper.

Kapor, A. J., Neilson, C. A., and Zimmerman, S. D. (2020). Heterogeneous Beliefs and School Choice Mechanisms. *American Economic Review*.

Kawaguchi, K., Uetake, K., and Watanabe, Y. (2021). Designing Context-Based Marketing: Product Recommendations Under Time Pressure. *Management Science*, 67(9):5642–5659.

Keane, M. and Wasi, N. (2013). Comparing Alternative Models of Heterogeneity in Consumer Choice Behavior. *Journal of Applied Econometrics*, 28(6):1018–1045.

Manski, C. F. (1977). The structure of random utility models. *Theory and Decision*, 8(3):229–254.

Manzini, P. and Mariotto, M. (2014). Stochastic Choice and Consideration Sets. *Econometrica*, 82(3):1153–1176.

McCall, J. J. (1970). Economics of Information and Job Search. *The Quarterly Journal of Economics*, 84(1):113–126.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In Zarembka, P., editor, *Frontiers in Econometrics*, pages 105–142, New York. Academic Press.

McFadden, D. (1978). Modeling the choice of residential location. In Karlqvist, A., Lundqvist, L., Snickars, F., and Weibull, J., editors, *Spatial Interaction Theory and Planning Models*, volume 1, pages 75–96. North-Holland.

Meisner, V. and von Wangenheim, J. (2023). Loss aversion in strategy-proof school-choice mechanisms. *Journal of Economic Theory*, 207:105588.

Narita, Y. (2018). Match or Mismatch? Learning and Inertia in School Choice. SSRN Scholarly Paper ID 3198417, Social Science Research Network, Rochester, NY.

Neilson, C. (2019). The Rise of Centralized Choice and Assignment Mechanisms in Education Markets Around the World. Technical report, Princeton Economics.

Neilson, C. (2020). Targeted Vouchers, Competition Among Schools, and the Academic Achievement of Poor Students. Job Market Paper, Yale University.

Pathak, P. A. (2017). What Really Matters in Designing School Choice Mechanisms. In Honore, B., Pakes, A., Piazzesi, M., and Samuelson, L., editors, *Advances in Economics and Econometrics*, pages 176–214. Cambridge University Press, Cambridge.

Roth, A. E. (1982). The economics of matching: Stability and incentives. *Mathematics of operations research*, 7(4):617–628.

Sovinsky Goeree, M. (2008). Limited Information and Advertising in the U.S. Personal Computer Industry. *Econometrica*, 76(5):1017–1074.

Swait, J. and Ben-Akiva, M. (1987). Incorporating random constraints in discrete models of choice set generation. *Transportation Research Part B: Methodological*, 21(2):91–102.

Train, K. E. (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press.

van Nierop, E., Bronnenberg, B., Paap, R., Wedel, M., and Franses, P. H. (2010). Retrieving Unobserved Consideration Sets from Household Panel Data. *Journal of Marketing Research*, 47(1):63–74.

Walters, C. R. (2018). The Demand for Effective Charter Schools. *Journal of Political Economy*, 126(6):2179–2223.

Wang, T. and Zhou, C. (2020). High school admission reform in China: A welfare analysis. *Review of Economic Design*, 24(3):215–269.

Online Appendix to “Imperfect Information and Outside Options in Centralized School Choice”

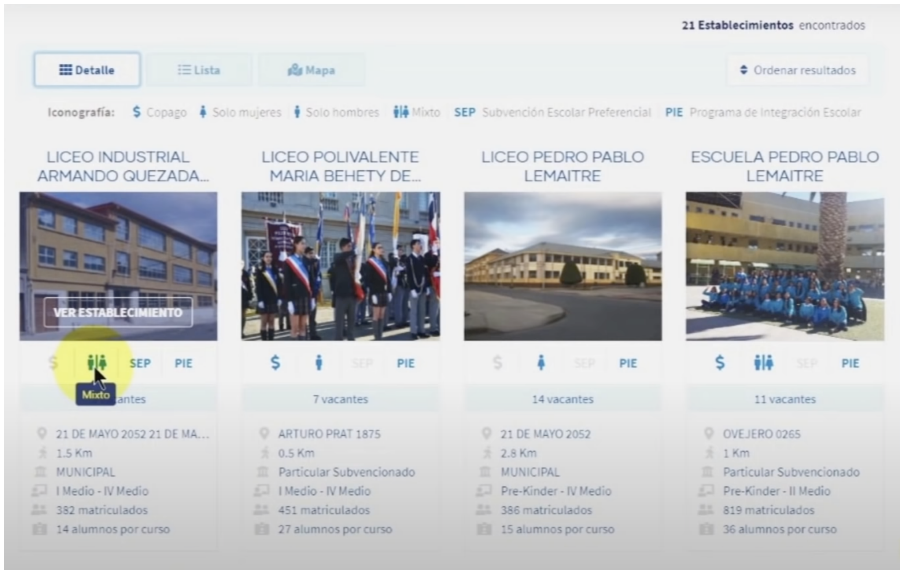
Felipe Arteaga

November 30, 2023

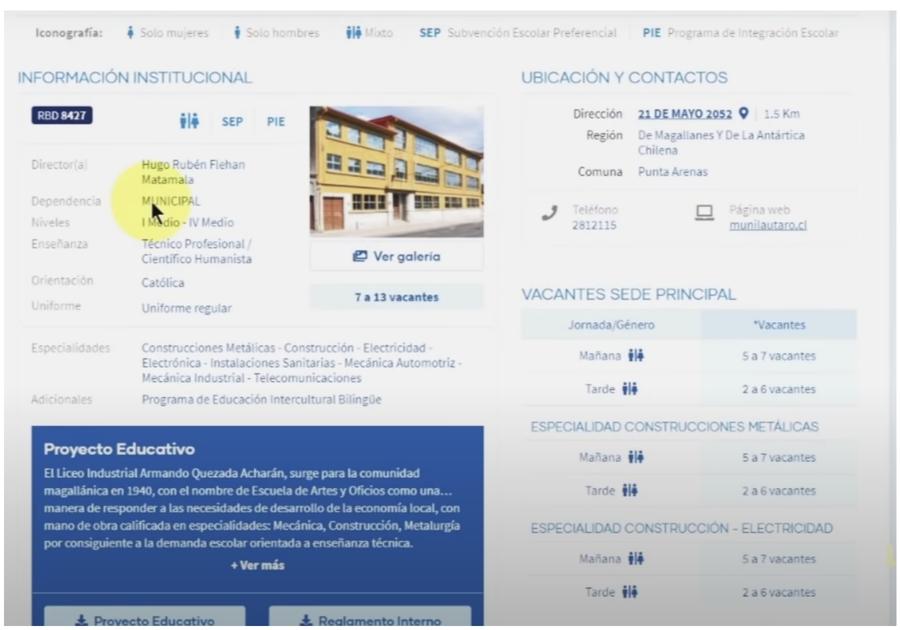
Contents

Figure B.I
2022 Online Application Platform Screenshots

(a) Gallery of schools



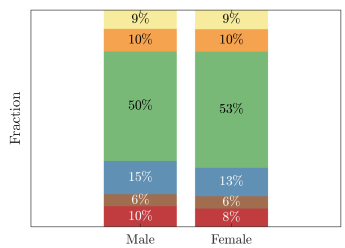
(b) Detailed Information of a School



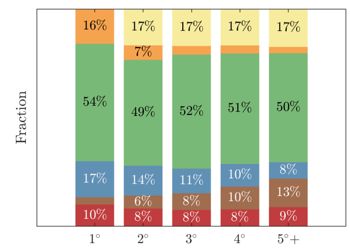
Notes: Panel (a) displays a sample view that an applicant would see in the gallery of schools, featuring a primary photo and several attributes such as proximity to home, enrollment size, and cost. Users have the option to view the schools in a list format or on a map, showcasing all nearby educational options. Panel (b) presents a screenshot containing detailed information about a particular school, including its educational program, estimated availability of seats, religious affiliation, among others.

Figure B.II
Enrollment decision for non-compliers

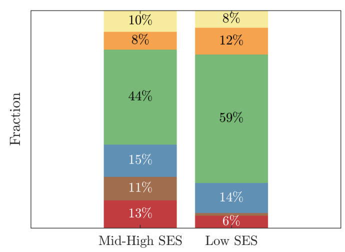
(a) By applicant's gender



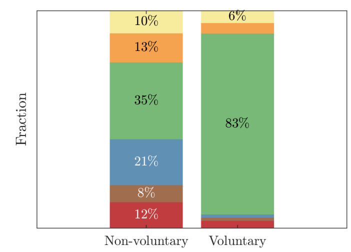
(b) By placement ranking



(c) By SES



(d) By voluntary status



(e)

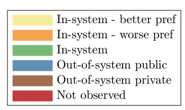
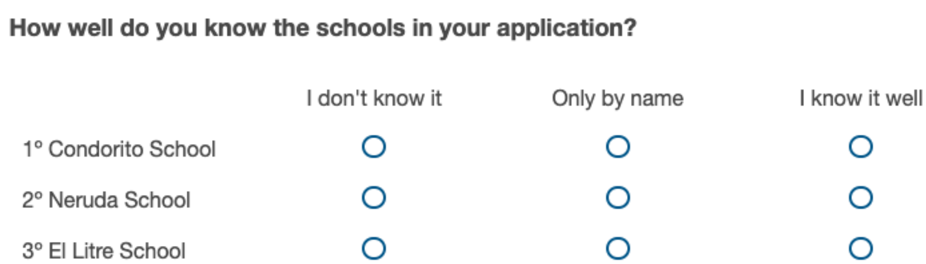
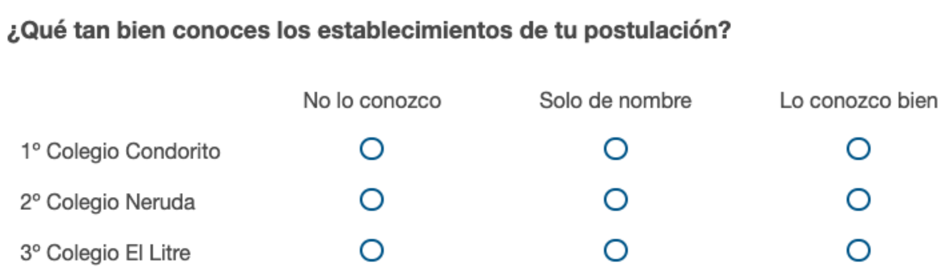


Figure B.III
Screenshot of question about knowledge level of schools

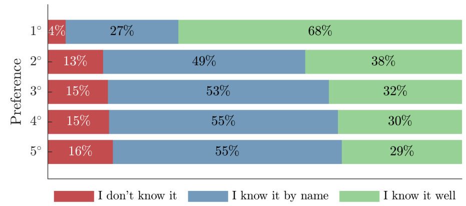
(a) Translated version to English



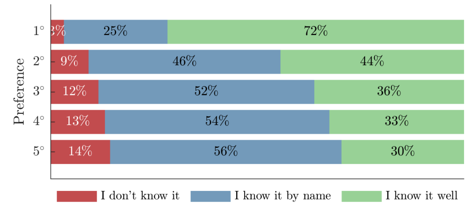
(b) Original version in Spanish



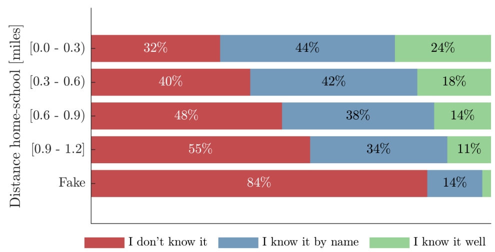
(a) Knowledge of ranked options - Low education



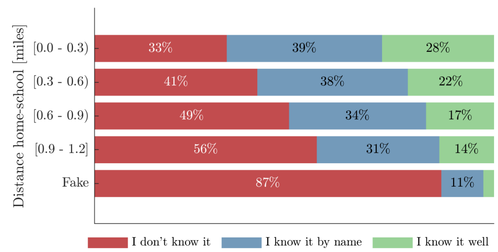
(b) Knowledge of ranked options - High education



(c) Knowledge of non-ranked options - Low education



(d) Knowledge of non-ranked options - High education



Notes: "Low education" refers to families whose mother has at most secondary education (48% of the survey sample). "High education" refers to families whose mother has more than secondary education, ie complete or incomplete technical tertiary education or a college degree (52% of the survey sample).

B. Additional Tables

Table B.I
Composition of in-system and out-of-system schools

		% of enrollment	
		PK-12th	PK-K
In-system schools	Public	36%	27%
	Private voucher ("charter")	52%	41%
Out-of-system public schools	Regular preschools	<1%	4%
	Language preschools	2%	18%
	Artistic, sport, or hospital based	<1%	<1%
Out-of-system private schools	Private non-voucher	9%	9%

Notes. "In-system" are schools that participate in the centralized admission system, and out-of-system schools that have their own admission process. Out-of-system public schools are publicly funded. They may be owned by a non-profit (private voucher) or by a state agency or municipality (public). Source: enrollment data 2022, Ministry of Education, Chile.

Table B.II
Descriptive Statistics for Schools

	(1)	(2)	(3)	(4)	(5)
	In estimation				Out of estimation
	All	PK to K	1st to 6th	9th to 12th	
A. Unweighted					
# of grades offered	10.55	11.90	11.71	11.21	7.64
Enrollment per grade	70.31	57.62	57.54	85.85	17.28
Share of low SES	0.65	0.65	0.64	0.62	0.83
Charter	0.61	0.64	0.67	0.71	0.26
Rural	0.01	0.01	0.01	0.01	0.68
Math test score	0.07	0.14	0.15	0.08	-0.03
Language test score	0.02	0.05	0.07	0.10	-0.30
Missing math test score	0.01	0.01	0.00	0.00	0.08
Missing language test score	0.01	0.01	0.00	0.00	0.08
Charges monthly fee	0.21	0.20	0.23	0.30	0.01
Monthly fee (USD)	17.51	17.24	19.71	26.55	0.40
B. Weighted by enrollment					
# of grades offered	11.27	12.60	12.42	11.89	9.92
Enrollment per grade	94.61	79.46	78.88	107.41	42.15
Share of low SES	0.62	0.62	0.61	0.60	0.78
Charter	0.67	0.72	0.74	0.76	0.29
Rural	0.01	0.01	0.01	0.01	0.35
Math test score	0.15	0.21	0.22	0.17	-0.04
Language test score	0.15	0.18	0.19	0.23	-0.19
Missing math test score	0.00	0.00	0.00	0.00	0.01
Missing language test score	0.00	0.00	0.00	0.00	0.01
Charges monthly fee	0.25	0.25	0.28	0.32	0.02
Monthly fee (USD)	21.74	22.45	24.57	29.01	1.36
N	3,344	2,357	2,792	1,953	4,682

Notes. All statistics are means in the population defined by the column header. Columns 2 to 4 consider schools that at least offer the grades defined by the header. Panel A shows unweighted means, panel B displays weighted means by the school enrollment. Selected row variable definitions are as follows. "Rural" is an indicator for schools located outside urban areas defined by the 2017 census. "Math and Language test scores" are standardized national tests.

Table B.III
Counterfactual Results for Non-placed Applicants

	(1)	(2)	(3)	(4)	(5)	(6)
	Placement			Enrollment		
	Better	Same	Worse	Better	Same	Worse
A. Mechanical counterfactuals						
Non-complier not applying to offer	0.085	0.915	0.000	0.063	0.937	0.000
Non-complier not applying to offer or lower preference	0.112	0.888	0.000	0.083	0.917	0.000
B. Model-based counterfactuals: oracle information campaign						
Oracle recommendation, full knowledge ($\alpha = 1$, $\beta = 1$)	0.053	0.947	0.000	0.039	0.961	0.000
Oracle recommendation, predicted knowledge ($\alpha = 1$, $\beta = 0$)	0.039	0.961	0.000	0.029	0.971	0.000
C. Model-based counterfactuals: naive information campaign						
Naive recommendation, full knowledge ($\alpha = 0$, $\beta = 1$)	0.141	0.859	0.000	0.115	0.885	0.000
Naive recommendation, predicted knowledge ($\alpha = 0$, $\beta = 0$)	0.141	0.859	0.000	0.088	0.912	0.000
D. Model-based counterfactuals: including out-of-system options in centralized platform						
Oracle recomendation + internalizing out-of-system ($\alpha = 1$, $\beta = 1$)	0.088	0.912	0.000	0.065	0.935	0.000

Notes.

Table B.IV
Counterfactual Results for Complier Placed in 2nd+ Preference

	(1)	(2)	(3)	(4)	(5)	(6)
	Placement			Enrollment		
	Better	Same	Worse	Better	Same	Worse
A. Mechanical counterfactuals						
Non-complier not applying to offer	0.076	0.902	0.022	0.074	0.900	0.026
Non-complier not applying to offer or lower preference	0.116	0.865	0.020	0.114	0.862	0.024
B. Model-based counterfactuals: oracle information campaign						
Oracle recommendation, full knowledge ($\alpha = 1$, $\beta = 1$)	0.047	0.945	0.008	0.046	0.944	0.010
Oracle recommendation, predicted knowledge ($\alpha = 1$, $\beta = 0$)	0.034	0.959	0.007	0.033	0.959	0.008
C. Model-based counterfactuals: naive information campaign						
Naive recommendation, full knowledge ($\alpha = 0$, $\beta = 1$)	0.138	0.800	0.062	0.137	0.799	0.064
Naive recommendation, predicted knowledge ($\alpha = 0$, $\beta = 0$)	0.068	0.878	0.054	0.063	0.877	0.060
D. Model-based counterfactuals: including out-of-system options in centralized platform						
Oracle recomendation + internalizing out-of-system ($\alpha = 1$, $\beta = 1$)	0.089	0.902	0.008	0.088	0.901	0.011

Notes.

Table B.V
Counterfactual Results for Complier Placed in 1st Preference

	(1)	(2)	(3)	(4)	(5)	(6)
	Placement			Enrollment		
	Better	Same	Worse	Better	Same	Worse
A. Mechanical counterfactuals						
Non-complier not applying to offer	0.000	0.983	0.017	0.000	0.982	0.018
Non-complier not applying to offer or lower preference	0.000	0.986	0.014	0.000	0.985	0.015
B. Model-based counterfactuals: oracle information campaign						
Oracle recommendation, full knowledge ($\alpha = 1, \beta = 1$)	0.000	0.994	0.006	0.000	0.994	0.006
Oracle recommendation, predicted knowledge ($\alpha = 1, \beta = 0$)	0.000	0.995	0.005	0.000	0.995	0.005
C. Model-based counterfactuals: naive information campaign						
Naive recommendation, full knowledge ($\alpha = 0, \beta = 1$)	0.066	0.878	0.056	0.065	0.867	0.068
Naive recommendation, predicted knowledge ($\alpha = 0, \beta = 0$)	0.022	0.933	0.045	0.020	0.928	0.052
D. Model-based counterfactuals: including out-of-system options in centralized platform						
Oracle recomendation + internalizing out-of-system ($\alpha = 1, \beta = 1$)	0.000	0.994	0.006	0.000	0.993	0.007

Notes.

Table B.VI
Counterfactual Results for Non-complier

	(1)	(2)	(3)	(4)	(5)	(6)
	Placement			Enrollment		
	Better	Same	Worse	Better	Same	Worse
A. Mechanical counterfactuals						
Non-complier not applying to offer	0.066	0.000	0.934	0.062	0.938	0.000
Non-complier not applying to offer or lower preference	0.089	0.000	0.911	0.041	0.959	0.000
B. Model-based counterfactuals: oracle information campaign						
Oracle recommendation, full knowledge ($\alpha = 1, \beta = 1$)	0.400	0.587	0.013	0.381	0.619	0.000
Oracle recommendation, predicted knowledge ($\alpha = 1, \beta = 0$)	0.295	0.692	0.013	0.066	0.934	0.000
C. Model-based counterfactuals: naive information campaign						
Naive recommendation, full knowledge ($\alpha = 0, \beta = 1$)	0.121	0.823	0.056	0.074	0.926	0.000
Naive recommendation, predicted knowledge ($\alpha = 0, \beta = 0$)	0.076	0.878	0.046	0.026	0.974	0.000
D. Model-based counterfactuals: including out-of-system options in centralized platform						
Oracle recomendation + internalizing out-of-system ($\alpha = 1, \beta = 1$)	0.682	0.304	0.014	0.649	0.351	0.000

Notes.

C. SIMULATED CHOICE SETS

In this section we explain how we implement our version of the specific consideration (ASC) model started by Manski (1977) in the estimation.[1] The original process requires integration over all the potential choice sets that contain the choices, which is computationally infeasible with many options (Abaluck and Adams-Prassl, 2021; Crawford et al., 2021). We follow the recommendation of Abaluck and Adams-Prassl (2021), and take an approach of simulated choice sets based on Sovinsky Goeree (2008).

The method uses simulation to approximate the integration over all potential choice sets. The procedure starts by calculating a consideration probability for each potential option for all applicants. Then, each simulated choice set is defined by a vector of iid uniform draws of length equal to the number of the potential options. If the draw is lower than the consideration probability, then the school is considered. Otherwise, it is not. Since the level of knowledge affects the utility in our framework, for the considered schools, we impute the knowledge using the prediction function described in Appendix G.2. In Sovinsky Goeree (2008) the consideration probabilities are calculated endogenously using advertisement measures as consideration shifters that don't affect choice probabilities. We use our survey data to estimate the consideration probability *offline*, approximating consideration with answers to our questions about knowledge of schools not in the ranking but in the neighborhood.

The detailed steps of the procedure are the following:

For each applicant i in the estimation sample:

1. Find the set of potential schools. Call J_i the cardinality of the set.
2. Predict the consideration probability $\hat{p}^c_{ij}$ with the function described on section C.1 for each potential option $j \in \{1 \ldots J_i\}$.
3. For each simulation $s \in \{1 \ldots S\}$ times:
   (a) Draw J_i iid uniform random variables, call them u_{ijs}
   (b) The inclusion of alternative j on the *simulated* choice set of i in simulation s is defined by the Bernoulli variable $b_{ijs} = \mathbb{1}(\hat{p}^c_{ij} > u_{ijs})$.

C.1. Consideration probability of unranked school

Considered schools for applicant i (or schools on her choice set Ω_i) are all the alternatives that she compares to build the rank order list. We partially observe the set considered schools through the rank order list ($\mathcal{C}_i$), but not the ones outside it. Since our survey did not ask directly about schools considered during the application that were not included in the ranking, we are gonna

[1] The method is also labeled as "integrating over approach" in Crawford et al. (2021). Abaluck and Adams-Prassl (2021) describes it and derives identification results.

proxy "consideration" with knowledge. We will assume that schools known by the families are in the choice set.

We aim to build a function to predict the consideration probability of non-ranked schools. We use this to build simulated choice sets, as introduced in Section V.C, and explained in detail in Appendix C.

To achieve this, we fit a binary logit model (Train, 2009) using the responses to the survey described in Section III. We are going to assume that school is considered if the answer is 2: *I know it by name*, 3: *I know it well*, and not considered if the answer is 1: *I don't know it*. We assume that the consideration $c(i,j) \in \{0,1\}$ depends on an underlying continuous index C_{ij} defined as:

$$C_{ij} = \alpha_1 \times distance_{ij} + \alpha_2 \times distance_{ij}^2 + \boldsymbol{\beta} \times connection_{ij} + \delta_{f(i)j} + \epsilon_{ij}$$

$distance_{ij}$ is the euclidean distance between home of applicant i and school j. $connection_{ij}$ is a vector that includes dummies representing a familiar connection with the school (a currently enrolled sibling, employed parent, or alumni). $\delta_{f(i)j}$ is a *school-student type* fixed effect, where $f(i)$ maps the individual "i" to a bin defined by the combinations of the two binary variables $female_i$ and $LowSES_i$ (2×2). ϵ_{ij} is an unobserved (to us) portion of C_{ij} and is assumed $IID\ Logistic(0,1)$.

We will assume that there is a threshold κ, and the consideration $c(i,j)$ depends if C_{ij} is higher than this threshold:

$$c(i,j) = \begin{cases} 0 : \text{not considered} & \text{if } C_{ij} < \kappa \\ 1 : \text{considered} & \text{if } C_{ij} > \kappa \end{cases}$$

We observe the covariates that define C_{ij}, and we build c_ij from survey answers to the question "How well do you know the schools in your neighborhood?", with results summarized in Figure IIIb in Section III.

$$c_{ij} = \begin{cases} 0 : \text{not considered} & \text{if } a_{ij} = 1 : \text{"I don't know it"} \\ 1 : \text{considered} & \text{if } a_{ij} = 2 : \text{"I know it by name" or } a_{ij} = 3 : \text{"I know it well"} \end{cases}$$

The probability of observing the three types of answers is the following:

$$P(c(i,j) = a) \begin{cases} P(C_{ij} < \kappa) & \text{if } a = 0 \\ P(C_{ij} > \kappa) & \text{if } a = 1 \end{cases}$$

Given that $\epsilon_{ij} \sim Logistic(0,1)$, the probability $P(c(i,j) = a|\boldsymbol{\theta})$ has a simple analytical form once we condition on the vector θ, that contains the parameters that define C_{ij} and the threshold.[2] The

[2]See Train (2009) for details.

log-likelihood function of observing c_{ij} for each school $j \in \mathcal{S}_i$[3] and survey respondent $i \in \{1 \ldots I\}$ is the following:

$$ll(\boldsymbol{\theta}) = \sum_{i=1}^{I} \sum_{j \in \mathcal{S}_i} \log \left(P(c(i,j) = c_{ij} | \boldsymbol{\theta}) \right)$$

The estimate of the vector of parameters $\boldsymbol{\theta} = [\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\delta}, \kappa]$ is the argument that maximized ll. With the estimated parameters, we can predict the consideration probabilities $\hat{p}_c$ for each school j that is a potential alternative for applicant i.

Since our survey sample comes from a very heterogeneous set of places, we estimated the binary logit at the urban zone level. That results in a set of 70 different vector of parameters $\{\boldsymbol{\theta_z}\}_{z \in \{1...70\}}$.

[3] $\mathcal{S}_i$ is the set of non-ranked schools that we asked about their knowledge to applicant i.

D. COUNTERFACTUAL SIMULATIONS

This section describes (1) the inputs of the counterfactuals, (2) the details of the simulation procedure for the baseline and counterfactuals,[4] It also shows the fit of the simulated baseline compared to the observed (real) results of the assignment/enrollment processes, disaggregated at the urban zone level.

D.1. Inputs for counterfactuals

We need three inputs:

Choice model estimates : Parameters associated with the observed portion of the expected utility (V_{ij}) on stage 1, shown in Table **??** on the main article.

Compliance model estimates : Parameters associated with the observed portion of the utility of the outside option U_{i0}, shown in Table **??** on the main article, and parameters related to the expected utility of the enrolled schools in stage 2 that are not present in stage 1 (λ and τ).

Knowledge prediction estimates : Parameters of the ordered logit models estimated for each urban zone that predict probabilities of knowledge of ranked schools based on distance, and a school fixed effect interacted with applicant characteristics. Details of the estimation in Appendix G.1.

D.2. Simulation procedure

We borrow the home location[5], characteristics, and the schools in the ranking ($j \in \mathcal{C}_i$) from each participant of the application system, we also borrow the i index. The only thing we don't borrow is the unobserved part of the utility ($\epsilon_{i,j}$) because we don't know it.

On each s simulation out of S:

- For each pair applicant-school $\{i, j\}$, $i \in \{1 \dots I\} \wedge j \in \mathcal{C}_i$
  1. Compute the observable part of the (indirect) expected utility of each school in the choice set ($\hat{V}_{ij}$), based on the estimated parameters, characteristics of student i and schools j
  2. Predict the knowledge level. We use the estimated ordered logit model to predict a probability for all three levels of knowledge: $\{\hat{p}^1_{ij}, \hat{p}^2_{ij}, \hat{p}^3_{ij}\}$. Then draw a knowledge level $\hat{k}(i,j)$ randomly from levels 1, 2 or 3 with probabilities $\{\hat{p}^1_{ij}, \hat{p}^2_{ij}, \hat{p}^3_{ij}\}$.
  3. Simulate the noise component $\eta_{\hat{k}(i,j)} \sim N\left(0, \sigma_{\hat{k}(i,j)}\right)$

[4] This section is based on the explanation structure of **?**.

[5] Home location for applicants with unreliable geocoding are imputed with the method detailed in Section D.3 of this appendix.

4. Simulate the Gumbel errors that represent the unobserved portion of the utility $\epsilon_{ij} \sim EVI$.
5. Construct the rank order list (ROL) based on the Indirect expected utility $E\hat{U}_{ij} = \hat{V}_{ij} + \hat{\eta}_{\hat{k}(i,j)} - \hat{\rho}^{s1}_{k(i,j)} + \hat{\epsilon}_{ij}$

- Run the assignment algorithm and get the assigned school $z(i)$ for each applicant. Now, for each applicant $i \in \{1 \dots I\}$:

1. Simulate the enrollment preference shock $\xi_{iz(i)} \sim EVI$
2. Compute the expected indirect utility for assigned school $E\hat{U}^{s2}_{iz(i)} = \lambda(\hat{V}_{iz(i)} + \hat{\tau} \times \hat{\eta}_{k(i,z(i))} - \hat{\rho}^{s2}_{\hat{k}(i,j)} + \hat{\epsilon}_{iz(i)}) + \hat{\xi}_{iz(i)}$
3. Compute the utility of the outside option U_{i0}, based on the location of i, placement, and characteristics of i.
4. Simulate the Gumbel errors that represent the stage 2 realization of the unobserved portion of the outside option's utility $\xi_{i0} \sim EVI$.
5. Construct the utility of the outside option $\hat{U}_{i0} = \hat{\lambda}\hat{U}_{i0} + \hat{\xi}_{i0}$
6. Construct the enrollment decision

$$Z_i = \begin{cases} 0: \text{ do not comply (do not enroll)} & \text{if} \quad E\hat{U}^{s2}_{iz(i)} < \hat{\lambda}\hat{U}_{i0} + \hat{\xi}_{i0} \\ 1: \text{ comply (enroll)} & \text{if} \quad E\hat{U}^{s2}_{iz(i)} > \hat{\lambda}\hat{U}_{i0} + \hat{\xi}_{i0} \end{cases}$$

C1: non-compliers don't apply to placed schools

- For students who didn't comply with the placement offer ($Z_i = 0$), we drop from their ranking the placed school $z(i)$. We re-run the Deferred Acceptance algorithm and construct the new enrollment decision based on the new placement.

C2: non-compliers don't apply to placed schools or lower preferences

- For students who didn't comply with the placement offer ($Z_i = 0$), we drop from their ranking the placed school $z(i)$ and any school in a lower preference. We re-run the Deferred Acceptance algorithm and construct the new enrollment decision based on the new placement.

C3: Information campaign

- For students who didn't comply with the placement offer ($Z_i = 0$), we suggest a school $q(i)$, aiming to be a prediction of where they would enroll. The ideal suggestion is the in-system school they will actually enroll: $s(i)$. Since this is not observed, we called it an oracle campaign. To account for prediction error, we vary the fraction of applicants that receive a suggestion from the oracle or from a naive predictor that suggests the school with the largest number of applicants that was not included in the ranking. We restrict $q(i)$ to be an in-system school.

- Since we don't know the enrolled school for our simulated population of non-compliers, we impute $s(i)$ matching each simulated non-complier with a real non-complier based on geographic distance, and assume that the enrollment of the simulated non-complier will be the same as the matched real non-complier. We do this procedure on each stratum defined by gender, application grade, and geographic zone, following these steps:
  1. We count the amount of simulated and real non-compliers. If the set of simulated is larger, we bootstrap from the real until we get the same number.
  2. We first match all simulated applicants that share the same geolocation with a real applicant.[6]
  3. We generate a lottery for each remaining non-matched simulated non-complier. The simulated applicants are matched to the closest non-matched real applicant, following the order induced by the lottery.
  4. Then, we define $s(i)$ as the observed enrolled school of the matched real non-complier. It might be that $s(i) = 0$, that is the case when we observe the matched applicant enrolled none school.

- To locate $q(i)$ in the rank when is equal to $s(i)$, we exploit a revealed preference argument to approximate its expected utility. If the enrolled school $s(i)$ is preferred to the placed school $z(i)$, then the expected utility of the former ($s(i)$) has to be greater or equal to the latter ($s(i)$). In practice, we draw the unobserved portion of the expected utility of the enrolled schools constrained to the utility inequality ($\epsilon_{iq(i)}\, st.\, EU^{s1}_{iq(i)} > EU^{s1}_{iz(i)}$)). This guarantees that the suggested school $s(i)$ is ranked better than the placed school $z(i)$.

- To locate $q(i)$ in the rank when is the naive recomendation, we calculate the observed utility based on model's estiamted parameters, and we draw the unobserved portion of the expected utility from EVI distribution. This opens the possibility of that the suggested school $s(i)$ is ranked better than the placed school $z(i)$.

- We vary the type of recommendation applicants receive. A fraction α receives their future enrollment school ($q(i) = s(i)$), and a fraction by $(1 - \alpha)$ is a naive prediction of where they could enroll, based on popularity.

- To analyze the impact of the "intensive margin", specifically, the extent to which families are informed about school $q(i)$, we examine varying levels of familial knowledge about the school the policy recommends. We introduce a parameter, β, to quantify this variation. At one extreme, where $\beta = 1$, families possess comprehensive knowledge about school $q(i)$,

[6] All simulated non-complier applicants have the same geolocation of at least one real applicant, but not necessarily a non-complier.

expressed as $k(i, q(i)) = 3$, eliminating any uncertainty penalization in the expected utility ($EU^{s1}_{iq(i)}$). Conversely, when $\beta = 0$, families possess only the predicted level of knowledge attributed to a school that does not feature in the rankings. The function to predict this knowledge level is estimated using survey data concerning non-ranked schools, as detailed in Appendix in G.2.

C4: Oracle campaign + out-of-system included in centralized platform

- This is equivalent to C3, but now we restrict $q(i)$ to be an in-system school or an out-of-system publicly founded school.

D.3. Home location imputation procedure

We use the centroid of the applied schools[7], plus a random distance shifter drawn from the empirical distribution of distances centroid-home of students with reliable geocoding. Since traveled distances may differ city by city, and the centroid carries different information depending on the number of schools, we perform this process at the urban zone and length of ROL level.

To account for city geography and avoid imputed location in infeasible zones, for example, in the sea for coastal cities, the direction of the distance shifter is drawn from the empirical distribution of directions of well-geolocated families within 1 km.

[7] We consider at most the first 3 schools in the ranking.

We are interested in the parameters of our joint decision model, represented by the vector $\boldsymbol{\theta} = [\gamma, \gamma_X, \boldsymbol{\beta}_X, \boldsymbol{\beta}^{\sigma}, \delta, \boldsymbol{\sigma}^{\eta}, \boldsymbol{\rho}, \boldsymbol{\psi}, \lambda, \tau]$. To estimate them we follow a log-likelihood maximization procedure. In Section V we defined the individual likelihood function, conditional on $\boldsymbol{\theta}$, as:

$$L_i(\boldsymbol{\theta}) = \int \left(\prod_{r \in \mathcal{C}_i} \frac{\exp\left(V_{ir} + \eta_{k(i,r)} - \rho^{s1}_{k(i,r)}\right)}{\sum_{j \in \Omega_i \setminus \{1 \ldots r-1\}} \exp\left(V_{ij} + \eta_{k(i,j)} - \rho^{s1}_{k(i,j)}\right)} \times \int \frac{1}{1 + \exp\left(\lambda V_{i0} - \lambda\left(V_{iz(i)} + \tau \times \eta_{k(i,z(i))} - \rho^{s2}_{k(i,z(i))} + \epsilon_{iz(i)}\right)\right)} \mathrm{d}F(\epsilon_{iz(i)} | \boldsymbol{\beta}^{\sigma}, \boldsymbol{\eta}) \right) \mathrm{d}F(\boldsymbol{\beta}^{\sigma}, \boldsymbol{\eta})$$

Since the integral has no closed form, we use simulation to approximate it (Train, 2009). The primitives of the random terms (but $\epsilon_{iz(i)}$) are the following:

$$\begin{aligned}
\boldsymbol{\beta}^{\sigma}_i &= \boldsymbol{\phi}^{\beta}_i \cdot \boldsymbol{\sigma}^{\beta} & \qquad \boldsymbol{\phi}^{\beta}_i &\sim N(0, I_{|\boldsymbol{\beta}^{\sigma}|}) \\
\eta_1 &= \phi^{\eta}_i \sigma^{\eta}_1 & \phi^{\eta}_i &\sim N(0,1) \\
\eta_2 &= \phi^{\eta}_i \sigma^{\eta}_2 & \phi^{\eta}_i &\sim N(0,1)
\end{aligned}$$

Initially, for each applicant i in the set $\{1, \ldots, I\}$ and for each simulation s in the set $\{1, \ldots, S\}$, we obtain the draws ϕ^{η}_{is} and $\boldsymbol{\phi}^{\beta}_{is}$. We then compute the individual likelihood for each s and average these values to approximate the overall likelihood L_i:

$$\hat{L}_i(\boldsymbol{\theta}) = \frac{1}{S} \sum_{s=1}^{S} \left(\prod_{r \in \mathcal{C}_i} \frac{\exp\left(V_{irs} + \eta_{k(i,r)s} - \rho^{s1}_{k(i,r)}\right)}{\sum_{j \in \Omega_i \setminus \{1 \ldots r-1\}} \exp\left(V_{ijs} + \eta_{k(i,j)s} - \rho^{s1}_{k(i,j)}\right)} \times \frac{1}{S'} \sum_{s'=1}^{S'} \frac{1}{1 + \exp\left(\lambda V_{i0s} - \lambda\left(V_{iz(i)s} + \tau \times \eta_{k(i,z(i))s} - \rho^{s2}_{k(i,z(i))} + \epsilon_{iz(i)ss'}\right)\right)} \right)$$

Then, we search for the vector $\hat{\boldsymbol{\theta}}$ that maximizes the sum of the logarithm of $\hat{L}_i$:[8]

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \sum \log\left(\hat{L}_i(\boldsymbol{\theta})\right)$$

The vector $\hat{\boldsymbol{\theta}}$ represents our estimates. We calculate the covariance matrix as the inverse of the Fisher Information matrix, defined as the negative expectation of the Hessian matrix of the log-

[8] We use the $BHHH$ algorithm in the maximization process.

likelihood function. We use the outer product of the gradient (covariance matrix of the scores) to approximate the Hessian (Train, 2009).

E.1. 2-step procedure

To recover the empirical distribution of the unobserved portion of the utility of the placed school ($\epsilon_{iz(i)}$) conditional on other random parameters ($\boldsymbol{\beta}^{\sigma}$ and $\boldsymbol{\eta}$) we add a "step 1" that precedes the estimation of the full model ("step 2"). In step 1, we perform a preliminary estimation of the parameters related to EU_{ij}^{s1}, using only the ranking data and not the enrollment decision. We use those estimates to construct the observed portion of the utility conditional on $[\boldsymbol{\beta}^{\sigma}, \boldsymbol{\eta}]$, and then recover draws from the distribution of ϵ_{ij}, imposing a "coherence constraint" between ϵ_{ij} and the ranking we observe. This approximation has a flavor of the procedure used by Abdulkadiroğlu et al. (2017) to calculate the expected utilities of ranked alternatives. The detailed procedure of "step 1" is the following:

- We start by generating all the draws necessary to approximate our step 1 integrals by simulation. Those correspond to the random parameters associated with preference heterogeneity ($\boldsymbol{\beta}^{\sigma}$) and the noise term ($\boldsymbol{\eta}$). We will use the same set of draws for step 1 and step 2.
- We estimate the parameters of the rank choice model (stage 1 in the model), i.e., without including the enrollment decision (stage 2 in the model), using simulated maximum likelihood. The simulated log-likelihood function that we maximize is:

$$ll(\boldsymbol{\theta}) = \sum_{i=1}^{I} \log\left(\frac{1}{S}\sum_{s=1}^{S}\prod_{r\in\mathcal{C}_i}\frac{\exp\left(V_{irs} + \eta_{k(i,r)s} - \rho_{k(i,r)}^{s1}\right)}{\sum_{j\in\Omega_i\setminus\{1\ldots r-1\}}\exp\left(V_{ijs} + \eta_{k(i,j)s} - \rho_{k(i,j)}^{s1}\right)}\right)$$

- With the maximum likelihood estimates of the rank choice parameters in hand, for each applicant $i \in \{1\ldots I\}$ and simulation $s \in \{1\ldots S\}$ (i.e. conditional on $\boldsymbol{\beta}_{is}^{\sigma}$ and $\boldsymbol{\eta}_{ijs}$):
  1. We predict the observed part of the expected utility $V_{ijs} + \eta_{k(i,j)s} - \rho_{k(i,j)}$ using the estimated parameters and the generated draws ($\boldsymbol{\beta}_{s}^{\sigma}$ and $\boldsymbol{\eta}_{s}$) for every school in the ranking.
  2. We generate T approximate draws from $F(\epsilon_{ij}|\boldsymbol{\beta}_{s}^{\sigma}, \boldsymbol{\eta}_{s})$, performing the following procedure T times:
     (a) We create a set of candidates $\{\hat{\epsilon}_{ijs}\}_{j\in\mathcal{C}_i}$ sampling $|\mathcal{C}_i|$ iid EVI draws.
     (b) We use the candidates $\hat{\epsilon}_{ijs}$ to construct the expected utilities $\hat{EU}_{ijs}^{s1} = \hat{V}_{ijs} + \hat{\eta}_{k(i,j)s} - \hat{\rho}_{k(i,j)}^{s1} + \hat{\epsilon}_{ijs}$.
     (c) We check if the constructed expected utilities are coherent with the ranking: $\hat{EU}_{irs}^{s1} > \hat{EU}_{ijs}^{s1} \quad \forall j > r, \forall r \in \mathcal{C}_i$. If the order of the constructed $\hat{EU}_{ijs}^{s1}$ matches the ranking, then we save our candidates $\{\hat{\epsilon}_{ijs}\}_{j\in\mathcal{C}_i}$ as a realization of the unobserved part of the

expected utility of each school j in the ranking of i. If it does not match the ranking, we go back to step (a).[9]

- Since this is performed at s level and T times, at the end of the procedure we have a matrix $\hat{\boldsymbol{\epsilon}}_{ij}$ of length SxT. In the estimation of the full model we only use the vector of draws related to the placed school: $\hat{\boldsymbol{\epsilon}}_{iz(i)}$.

In step 2 we estimate the full model. We don't need to produce new draws for the approximation of the integrals, since we use the same set generated in step 1 for $\boldsymbol{\beta}^{\sigma}$ and $\boldsymbol{\eta}$, and the approximate draws of $\hat{\boldsymbol{\epsilon}}_{iz(i)}$ generated on the step 1.

[9]We go back to step (a) at most 500 times. We were able to recover coherent vector draws for 96% of the $I \times S \times S'$ rankings. For the remaining 4%, we save the "most-coherent" vector draw out of the 500 draws of $\hat{\boldsymbol{\epsilon}}_{is}$. The most coherent is defined as the vector which minimizes the "incoherent distance" between adjacent ranked schools, defined as $\sum_{r=1}^{|\mathcal{C}_i|-1} \max\{0, EU_{r+1} - EU_r\}$ Ex: If $EU_1 = 5$, $EU_2 = 3$, $EU_3 = 4$, the "incoherent distance" is $\max\{0, -2\} + \max\{0, 1\} = 1$, and reflects the fact that the expected utilities of options 2 and 3 are "incoherent" with respect to the preference order.

F. Additional Proofs

F.1. Better outside option reduces the value of search

The benefit from search is represented by the expression:

$$
\begin{aligned}
E[\mathcal{V}(\mathcal{C}_i \cup s) - \mathcal{V}(\mathcal{C}_i)] =& \mathbb{E}[(w_{is} - EU_{i0}) p_{is} \prod_{j \leq N} R_{ij}] \\
=& \int (w_{is} - EU_{i0}) p_{is} \, \mathrm{d}F_i(EU_{is}^{s1}, p_{is}) \prod_{j \leq N} R_{ij}
\end{aligned}
$$

Assuming that the support of EU_{is}^{s1} is positive[10], since p_{is} and $\prod_{j \leq N} R_{ij}$ are non-negative, a better outside options produce a change on the benefit of search with the same sign as:

$$
\begin{aligned}
\frac{\partial w_{is} - EU_{i0}}{\partial U_{i0}} =& (\log \Big(\exp(\lambda EU_{is}^{s1}) + \exp(\lambda U_{i0}) \Big) - \mathbb{E}[\lambda U_{i0} + \xi_{i0}]) \\
=& \log \Big(\exp(\lambda EU_{is}^{s1}) + \exp(\lambda U_{i0}) \Big) - \lambda U_{i0} + \mathbb{E}[\xi_{i0}]) \\
=& \lambda \left(\frac{\exp(\lambda U_{i0})}{\exp(\lambda EU_{is}^{s1}) + \exp(\lambda U_{i0})} - 1 \right) < 0
\end{aligned}
$$

The left term within parenthesis is always smaller than 0, and since $\lambda > 0$, the partial derivative is negative: a better outside option reduces the benefit of search.

[10] This is not restrictive. We can uniformly add any positive constant to all EU^{s1} and the choice decision is unaltered.

G. Knowledge and consideration prediction functions

The counterfactuals detailed in Section VII and the alternative choice set definition based on the work of Sovinsky Goeree (2008) explained in Section V.C rely on functions that predict a probability of knowledge level or consideration. This Section describes those functions.

G.1. Knowledge prediction function for ranked school

Our goal is to build a function to predict probabilities of knowledge level of ranked school for each school and applicants, given the position on the ranking. This will be used to build the expected utility of the ranked schools for the universe of simulated applicants.

To achieve this, we fit an ordered logit model (Train, 2009) using the responses to the survey described in Section III. The orderer discrete variable that we want to predict has three categories: 1: *I don't know it, 2: I know it by name, 3: I know it well* . We assume that the discrete level of knowledge $k(i,j) \in \{1 \dots 3\}$ depends on an underlying continuous index K_{ij} defined as:

$$K_{ij} = \alpha_1 \times distance_{ij} + \alpha_2 \times distance_{ij}^2 + \boldsymbol{\beta} \times connection_{ij} + \boldsymbol{\phi} \times rank_{ij} + \delta_{f(i)j} + \epsilon_{ij}$$

$distance_{ij}$ is the euclidean distance between home of applicant i and school j. $connection_{ij}$ is a vector that includes dummies representing a familiar connection with the school (a currently enrolled sibling, employed parent, or alumni). $rank_{ij}$ is a rank fixed effect. $\delta_{f(i)j}$ is a *school-student type* fixed effect, where $f(i)$ maps the individual "i" to a bin defined by the combinations of the two binary variables $female_i$ and $LowSES_i$ (2×2). ϵ_{ij} is an unobserved (to us) portion of K_{ij} and is assumed $IID\ Logistic(0,1)$.

We will assume that there are thresholds κ_1 and κ_2 that families use to map the underlying continuous index K_{ij} to the discrete level of knowledge $k(i,j)$ with the following rule:

$$k(i,j) = \begin{cases} 1: \text{"I don't know it"} & \text{if } K_{ij} < \kappa_1 \\ 2: \text{"I know it by name"} & \text{if } \kappa_1 \leq K_{ij} < \kappa_2 \\ 3: \text{"I know it well"} & \text{if } \kappa_2 < K_{ij} \end{cases}$$

We observe the covariates that define K_{ij}, and we collect $k(i,j)$ from survey answers to the question "How well do you know the schools in your application?", pictured in Figure B.III on Appendix A, with results summarized in Figure IIIa in Section III.

The probability of observing the three types of answers is the following:

$$P(k(i,j)=a)\begin{cases} P(K_{ij}<\kappa_1) & \text{if } a=1 \\ P(\kappa_1\leq K_{ij}<\kappa_2) & \text{if } a=2 \\ P(\kappa_2<K_{ij}) & \text{if } a=3 \end{cases}$$

Given that $\epsilon_{ij}\sim Logistic(0,1)$, the probability $P(k(i,j)=a|\boldsymbol{\theta})$ has a simple analytical form once we condition on the vector θ, that contains the parameters that define K_{ij} and the thresholds.[11] The log-likelihood function of observing responses a_{ij} for each school $j\in\mathcal{C}_i$[12] and survey respondent $i\in\{1\ldots I\}$ is the following:

$$ll(\boldsymbol{\theta})=\sum_{i=1}^{I}\sum_{j\in\mathcal{C}_i}\log\left(P(k(i,j)=a_{ij}|\boldsymbol{\theta})\right)$$

The estimate of the vector of parameters $\boldsymbol{\theta}=[\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\phi},\boldsymbol{\delta},\boldsymbol{\kappa}]$ is the argument that maximized ll. With the estimated parameters, we can predict the knowledge-level probabilities $[\hat{p}_1,\hat{p}_2,\hat{p}_3]$ for each school j on the rank order list of applicant i.

Since our survey sample comes from a very heterogeneous set of places, we estimated the ordered logit at the urban zone level. That results in a set of 70 different vector of parameters $\{\boldsymbol{\theta_z}\}_{z\in\{1...70\}}$.

G.2. Knowledge prediction function for unranked school

We aim to build a function to predict probabilities of knowledge levels of non-ranked schools. We use this function to build the expected utility of schools suggested by our simulated policy in the counterfactuals, described in Section VII.

Our methodology parallels the strategy outlined in the preceding section, which develops a function to forecast knowledge for ranked schools, albeit with two distinctions.

1. The underlying continuous index K_{ij} that defines the knowledge categories do not include the ranking fixed effects, nor the $distance^2$ term.

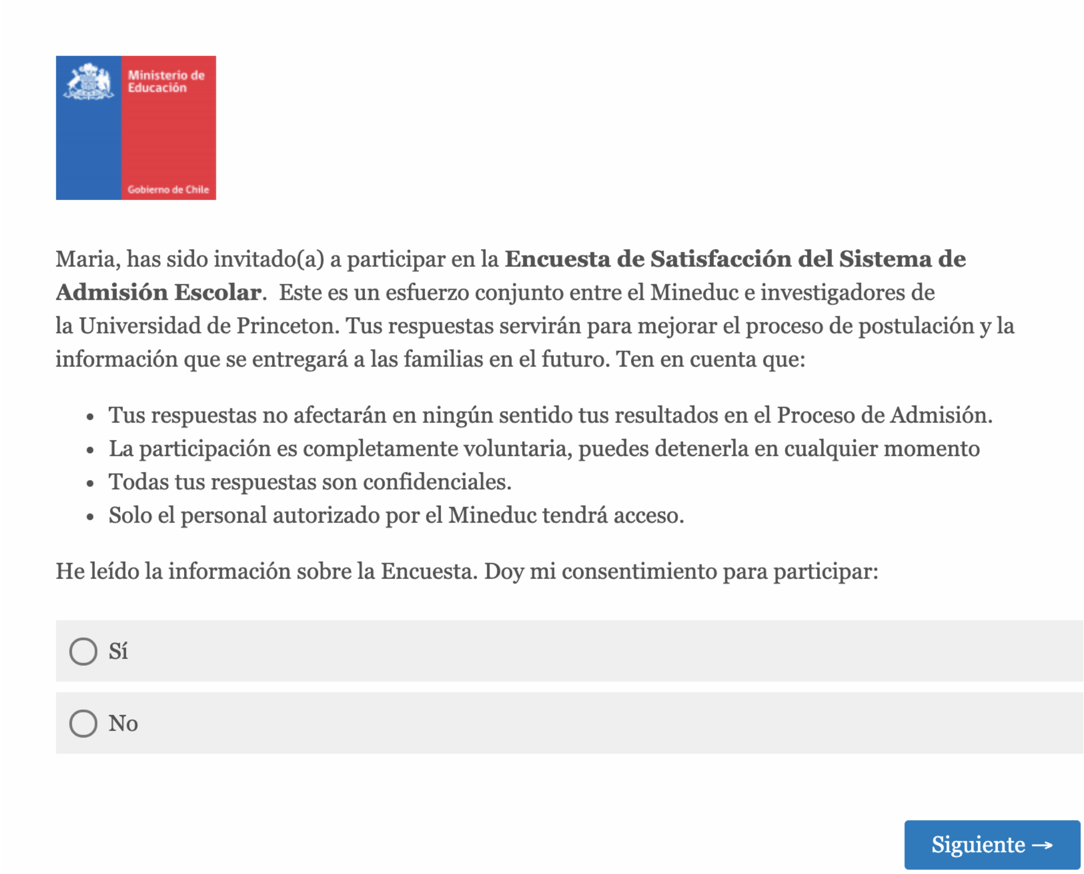
2. The data we used for the estimation comes from a different survey question. Besides asking for the knowledge of ranked options, we also asked about non-ranked options. We describe how we picked these options on Section III.

[11] See Train (2009) for details.

[12] In the surveys 2020 and 2021, we asked for at most five schools; in 2022, at most seven.

H. Survey Translation

Figure B.V
2020 Survey Landing Page



Ministerio de Educación

Gobierno de Chile

Maria, has sido invitado(a) a participar en la **Encuesta de Satisfacción del Sistema de Admisión Escolar**. Este es un esfuerzo conjunto entre el Mineduc e investigadores de la Universidad de Princeton. Tus respuestas servirán para mejorar el proceso de postulación y la información que se entregará a las familias en el futuro. Ten en cuenta que:

- Tus respuestas no afectarán en ningún sentido tus resultados en el Proceso de Admisión.
- La participación es completamente voluntaria, puedes detenerla en cualquier momento
- Todas tus respuestas son confidenciales.
- Solo el personal autorizado por el Mineduc tendrá acceso.

He leído la información sobre la Encuesta. Doy mi consentimiento para participar:

○ Sí

○ No

Siguiente →

Notes. This is the website displayed after applicants clicked the invitation link to participate in the 2020 survey, which is very similar to the 2021 and 2022 version. The link was sent by email. The translation to English is the following: *Maria, you have been invited to participate in the School Admission System Satisfaction Survey, a joint effort between Mineduc and Princeton University researchers. Your answers will help to improve the application process and the information that we will provide new applicants. Note that: (1) Your answers will not affect in any way your results in the Admission Process. (2) Participation is entirely voluntary; you can stop it at any time. (3) All your answers are confidential. (4) Only personnel authorized by Mineduc will have access. I have read the information about the Survey. I give my consent to participate. [Options: Yes or No]*

1. *(List of schools, a reminder of the filed application)*

2. First, we want to know how you evaluate the process of the School Admission System. Choose a grade from 1 to 7 for the following aspects
   [*Slider 1 to 20*]

   (a) Information on schools available (academic performance, collections, educational project, after school activities)

   (b) Availability of information on the application process (relevant dates, website, etc).

(c) In general, what rating would you put to the application process?

3. How did you get information about of the application process? Select all that apply
[*Select multiple*]

    (a) Through the Municipality
    (b) Through the current school/pre-school
    (c) Through the newspaper or radio
    (d) Through social networks (Facebook, Instagram, Twitter, Youtube)
    (e) Through friends or relatives
    (f) Through the website of the Ministry of Education (www.sistemadeadmisionescolar.cl)
    (g) Through the platform of the Ministry of Education Your Information
    (h) I did not inform myself

4. Select the social networks you used to get information about SAE?
[*Select multiple*]

    (a) Facebook
    (b) Twitter
    (c) Instagram
    (d) Youtube

5. Select the traditional media outlets you used to get information about SAE?
[*Select multiple*]

    (a) Newspaper
    (b) Radio
    (c) TV

6. When you add a school to your application, what do you consider a necessary steps to know well a school before applying?(Check all that apply).
[*Select multiple*]

    (a) Knowing the infrastructure
    (b) Interview with the principal or a teacher
    (c) Visit the website of the school
    (d) Get referrals from someone you know
    (e) Academic Performance information
    (f) Knowing indicators from the Agency for Quality Education

(g) Knowing the extracurricular activities offered
(h) Know your project Educational Institutional (PIE)

7. Any other relevant step that we have not included here?
[*Open text*]

8. How well do you know the schools in your application ?
[*Knowledge scale: (I don't know it, Only by name, I know it well)*]

 (a) [Name preference 1]
 (b) [Name preference 2]
 (c) [Name preference 3]
 (d) [Name preference 4]
 (e) [Name preference 5]

9. Because COVID-19, much of classroom activities have been suspended.Do you think this affected your application process in any of these dimensions?
[*Select one*]

 (a) COVID-19 did not affect my application process
 (b) Without COVID-19, I would have known better the schools that I already know, but I would not have applied to more schools
 (c) Without COVID-19, I would have known more schools and perhaps I would have added them to my application

10. We note that during the application process you added schools to your initial list.¿Did you know these schools before the start of the application process?
[*Knowledge scale (I didn't know it before applying, I knew it by name before applying, I knew it well before applying)*]

 (a) [Name preference added 1]
 (b) [Name preference added 2]
 (c) [Name preference added 3]

11. In order to convince yourself to add these schools:
[*Select one*]

 (a) It was necessary to find out more about them
 (b) It was not necessary to search for more information

12. You applied to [Name preference 1] in first preference:From 0 to 100, how likely or how sure are you that you will get a seat on that option?
[*Slider 0 to 100*]

13. Imagine if you would had put your second choice [Name preference 2] as your first choice:From 0 to 100, how likely or how sure are you that you would get a seat on that option?
[*Slider 0 to 100*]
14. Imagine if you had put your third choice [Name preference 3] as your first choice:From 0 to 100, how likely or how sure are you that you would get a seat on that option?
[*Slider 0 to 100*]
15. Some families are not placed in any option beacuse there is no sufficient seats.Using the same range of 0 to 100,How likely or how sure are you that [Applicant name] will be placed in one of the [Length application] schools in the application?
[*Slider 0 to 100*]
16. Why you did not add more schools to your application?
[*Select one*]
    (a) I know the other options well and I prefer to have no placement than to add those alternatives
    (b) I think I will definitely be placed in one of the schools I applied for
    (c) It is very difficult to find more schools
    (d) There are no more schools close enough (good or bad)
17. If you would had added more schools to your application. Do you think you would have higher changes to be placed to one school?
[*Select one*]
    (a) No
    (b) Yes
18. Here are five schools. How well do you think you know these schools?
[*Knowledge scale: (I don't know it, Only by name, I know it well)*]
    (a) [School not considered in application 1]
    (b) [School not considered in application 2]
    (c) [School not considered in application 3]
    (d) [School not considered in application 4]
    (e) [School not considered in application 5]
19. From 1 to 10,How easy it is to find information on the academic performance of schools?
[*Slider 1 to 10*]

20. Imagine that you spend time researching all schools that you do not know well.After you know them well, do you think you would add at least one of these schools to your application?
[*Select one*]
    (a) No
    (b) Yes
21. From 0 to 100, how likely would you add it as your first preference?
[*Slider 0 to 100*]
22. From 0 to 100, how likely would you add it below your last choice?
[*Slider 0 to 100*]
23. During the application process, did you get any recommendations about adding more schools to your list?
[*Select one*]
    (a) No
    (b) Yes
24. By what method did you receive the recommendation to add more schools?(Select all that apply)
[*Select multiple*]
    (a) SMS
    (b) WhatsApp
    (c) E-mail
    (d) Web page
    (e) Other
25. By what method did you receive the recommendation to add more schools?- Other
[*Open text*]
26. If [applicant name] get a seat in the following schools, from 1 to 7, how satisfied would you be?
[*Slider 1 to 7*]
    (a) First preference: [Name preference 1]
    (b) Last Preference: [Name Last preference]
    (c) If you are not in any school in the regular period

27. Would you like to have had the following information on schools that did not have at the time of application?
    [*Yes or No*]
    (a) Information about your chances of being accepted
    (b) Standarized test score
    (c) Performance category
    (d) Price
    (e) Priority for economically-vulnerable students
    (f) SAT scores
    (g) Seats available
28. What is your preferred method of contact during the application process?
    [*Select one*]
    (a) E-mail
    (b) Other
    (c) SMS
    (d) Telephone
    (e) WhatsApp
29. What is your preferred method of contact during the application process? - Other
    [*Open text*]
30. For registration purposes only, what is the highest educational level of the Mother (or Stepmother) of [applicant name]?
    [*Select one*]
    (a) Educación Básica Completa
    (b) Educación Básica Incompleta
    (c) Educación Media Completa
    (d) Educación Media Incompleta
    (e) Educación incompleta en una Universidad
    (f) Grado de magíster universitario
    (g) No estudió
    (h) Titulada de un Centro de Formación Técnica o Instituto Profesional
    (i) Titulada de una Universidad

31. Do you know if [Field-nomPostulante] is a priority student (SEP)?
    [*Select one*]
    (a) He/she is not a beneficiary of the preferential subsidy
    (b) I do not know
    (c) He/she is a beneficiary of the preferential subsidy
32. Do you have any other comments, complaints or suggestions to make us?
    [*Open text*]