

# Imperfect Information and Outside Options in Centralized School Choice\*

Felipe Arteaga<sup>†</sup>

May 9, 2024

[Click here for the latest version](#)

## Abstract

This paper studies the consequences of limited information among school choice participants and the influence of outside options in application and enrollment decisions. We focus on the Chilean PK-12th application system, where 23% of the applicants who receive an offer 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 and estimate a model of the joint decision of school 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, 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 the ability to inform families about these schools thoroughly. 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 Admisión Escolar (SAE)* of the Ministry of Education in Chile.

<sup>†</sup>University of California, Berkeley. Mailing Address: 530 Evans Hall #3880, Berkeley, CA 94720. Email: [fharteaga@berkeley.edu](mailto: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 valued for their capacity to foster fairness, transparency, and efficiency in the allocation of students to schools. Economists have played a pivotal role in their evolution, not only by developing 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), strategize effectively based on the specific mechanism they encounter (Kapor et al., 2020), and navigate the complexities of application platforms. Furthermore, they must understand admission probabilities (Arteaga et al., 2022) and manage policymakers' decisions on information dissemination, system expansion, and handling of external schooling options.

This paper investigates 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 educational system using a deferred acceptance mechanism. Notably, 23% of applicants choose to enroll in different schools than their placement offers, most of them in a school they could have applied to. This non-compliance impacts 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. We explore the interplay of imperfect information, outside options, congestion externalities, and awareness-increasing policies within the framework of imperfectly informed families. Our analysis utilizes survey data from over 200,000 applicants and administrative records from three years of school choice processes involving nearly 1.5 million applicants.

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, which allows us to capture the role of risk preferences in family decision-making. Second, it accounts for the possibility of learning between the application and enrollment phases, enabling us to model how families update their beliefs based on new information. Third, it addresses the challenge of unobserved choice sets in environments with numerous options. Lastly, it considers heterogeneous outside options, enhancing our understanding of how families weigh the centralized system against alternative schooling choices.

**Results** Our survey data reveals a significant gap in families' knowledge about nearby schools and those they are applying to. Approximately 45% of the 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. The data indicates that applicants with limited information about their assigned school are up to 30% more likely to enroll in a different institution than well-informed families. This trend underscores the vital role that information plays in the decision-making process of families within centralized school choice systems.

Our theoretical framework sheds light on two key findings in the context of school choice. First, it demonstrates that in scenarios characterized by uncertainty aversion, a lack of comprehensive knowledge about a school significantly reduces its perceived value. Second, the framework reveals that the presence of more attractive outside options diminishes the incentive for families to engage in an extensive search of schools offered in the centralized system. These factors—reduced valuation and outside options—collectively decrease the likelihood that participants in the centralized school choice system will adhere to their initial enrollment offers.

Our model estimates suggest that the uncertainty levels are considerable in magnitude, translating into a penalty attributed to limited knowledge that is quantitatively equivalent to the effect of increasing the travel distance by three standard deviations. This uncertainty diminishes to 70% in the enrollment stage, indicating a significant degree of learning and information acquisition by families after placement.

[paragraph on model results of outside options value]

Our analytical framework identifies non-compliance behavior as a catalyst for policy innovation in two critical dimensions. First, 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 available within the centralized system but have yet to be considered. Enhancing awareness and knowledge about these alternatives could potentially reduce non-compliance rates. Second, 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 our model's estimated parameters and knowledge of the assignment mechanism to simulate the impact of policies targeted at these two dimensions, providing valuable insights for policymakers.

In the first set of our counterfactual analyses, we design a policy intervention to enhance families' knowledge about school options and examine its impact on compliance rates. We simulate a scenario where applicants are informed about a predicted school in which they would enroll, manipulating the accuracy of the school predictions and the depth of information provided. Our findings reveal 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 investigates 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 otherwise would have to apply to a school outside the system. These insights provide valuable evidence for policymakers considering the expansion of school choice frameworks to include a broader array of schooling options.

Market design 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 reduce the mismatch rate. However, 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 crucial for the market *re*-design, 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. We extend this line of inquiry by employing survey data to estimate a choice model that incorporates imperfect information, highlighting how interventions in school information can yield unexpected outcomes.

Moreover, our research aligns with and expands upon existing literature concerning applications within centralized systems under incomplete information, as exemplified by Grenet et al. (2022). We build upon the findings of Hastings and Weinstein (2008), Ajayi et al. (2017), and Andrabi et al. (2017), which demonstrate how providing information about school performance influences family choices. Our innovative approach includes considering outside options for future enrollment decisions and addressing noisy school valuations, offering a more comprehensive understanding of the decision-making process.

Our methodology is similar 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). By leveraging micro-data from strategy-proof mechanisms (Narita, 2018; Fack et al., 2019; Abdulkadiroğlu et al., 2020; Ainsworth et al., 2023), we contribute to the growing body of empirical models of school choice, as discussed in Agarwal and Somaini (2020).

Lastly, our work intersects with the broader discourse on discrete choice under uncertainty, exemplified in works related to health insurance (Handel, 2013) and auto insurance (Cohen and Einav, 2007). We also engage with the study of unobserved choice sets (Crawford et al., 2021; Abaluck and Adams-Prassl, 2021), aligning with (Barseghyan et al., 2021)’s proposition of a discrete choice model that accommodates unobserved heterogeneity in consideration sets and incorporates risk aversion. By bridging these various strands of literature, our research offers a novel perspective on the complex dynamics of school choice systems and the role of information in shaping outcomes.

**Organization of the Paper** The paper is structured as follows. Section II provides an overview of the Chilean school admission system, setting the context for our research. In Section III, we present an in-depth analysis of survey results, highlighting key insights into families’ knowledge and decision-making processes. Section IV introduces our joint model of school choice and compliance, which incorporates noisy school valuation to better capture the complexities of the choice process. The integration of this model with our rich dataset is detailed in Section V, followed by the presentation of estimation results in Section VI. Building upon these findings and model estimates, Section VII explores a range of counterfactual scenarios, shedding light on the potential impact of policy interventions. Finally, Section VIII concludes the paper by summarizing our key findings and discussing their implications for policy and future research.

## II. SETTING

We study school choice and enrollment decisions in the context of Chile, where a nationwide centralized system has regulated the admission for 87% of the seats available for the PK-12th education system since 2019. This section describes the institutional details, the application process, and the assignment mechanism. It then provides a description of the demographics of applicants, their application behaviors, placement, and enrollment results.

### II.A. The Chilean admission system: institutional details

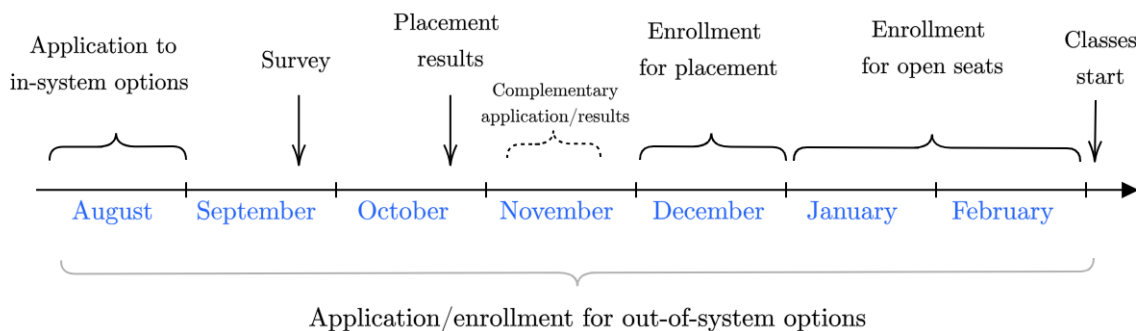
Chile has a long tradition of school choice, with student vouchers introduced in 1981 based on the idea that competition would drive schools to improve their services and attract more students (Hsieh and Urquiola, 2006). Initially, the choice process was decentralized, with families applying independently to each school they preferred. However, a bill passed in 2015 initiated the rollout of

a centralized school choice system. The system started as a pilot in 2016 in a region representing 1% of enrollment and gradually expanded to achieve national coverage by 2019.

The centralized system regulates admission for most public and private-voucher schools, which we refer to as “in-system” schools. In 2022, these schools represented 87.2% of the PK-12th enrollment (a total of 3.4 million students). However, among publicly funded schools (i.e., publicly owned or private schools that receive vouchers), two types do not participate in the centralized admission system, which we classify as “out-of-system public” schools. The first type is schools that offer kindergarten as their highest grade (“preschools”), accounting for 2.8% of PK-12 enrollment. The second type is schools with an artistic or sports specialty or those that are hospital-based, representing 0.5% of PK-12 enrollment. Additionally, some private schools that do not accept vouchers and have their own admission rules are classified as “out-of-system private” schools.<sup>1</sup>

Preschools are relevant in our setting because they serve as a meaningful outside option for families applying for young students through the centralized process. These preschools consist 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 specialized curriculum oriented towards students with language deficiencies and receive a voucher that is two to three times the amount allocated for regular education. Anecdotal evidence suggests that families can easily obtain a certificate to apply to these schools, regardless of the child’s language development, indicating that language schools can be an alternative for most applicants.<sup>2</sup>

Figure I  
Timeline of the application process



The centralized application process for the “in-system” schools typically takes place during August, as shown in Figure I. The government provides an online platform that not only registers applications but also offers information about the process and each school.<sup>3</sup> All participating

<sup>1</sup>Private-non-voucher schools represent 9.5% of the national enrollment and are typically expensive. A very small share of families apply to both voucher and non-voucher systems, making them almost separate markets. See Table B.I in Appendix B for a summary of the classification of schools and their share of the total enrollment.

<sup>2</sup>Another fact suggesting that language schools educate a diverse pool of students is that they represent almost 20% of the total PK-K enrollment, and most of their students transition to regular schools after kindergarten.

<sup>3</sup>Figure B.I in Appendix A shows screenshots of the application website from 2020.

schools must declare the number of seats they wish to fill for the upcoming academic year. The system makes these seats available to participants after reserving enough seats for currently enrolled students. There is a complementary application round lasting one week in November for students who (1) did not participate in the main round, (2) were not assigned to any school, or (3) rejected their placement.<sup>4</sup> In the complementary round, applicants can only apply to schools with seats not filled in the main round, resulting in a significantly reduced menu with very few “popular” options.

Families can apply to as many schools as they desire and modify their application while the process remains open. The government employs a student-proposing deferred acceptance (DA) mechanism to match students with schools (Correa et al., 2019). These conditions create a strategy-proof mechanism, meaning that the best strategy for families is to rank schools according to their *true* preferences (Roth, 1982).<sup>5</sup>

When a specific school receives more applications than available seats, the system employs a combination of coarse priorities and lottery numbers to allocate seats. The priorities, in order, are given to siblings, children of school employees, and alumni. Additionally, there is a quota for low-SES students and, in a few cases, a quota for high-performing students in high-school grades.<sup>6</sup> Lotteries are conducted independently for each school, following a multiple tie-breaking rule (Ashlagi et al., 2019).

The results from the main round matching process are released in late October. Families assigned to a preferred school must log in and either accept or decline the offer.<sup>7</sup> For applicants who do not make a decision, the default is to accept the assigned school.

Applicants who receive an offer from the centralized process have the last two weeks of December to exercise their option to enroll in the assigned school. If they do not enroll, they can, from January onwards, enroll in any publicly funded school with available capacity or opt for a private school.<sup>8</sup> All seats assigned in the centralized process that are not filled by the placed students are made available through a first-come, first-served decentralized system.<sup>9</sup>

---

<sup>4</sup>Among those placed in the main round who did not enroll in their assigned school, only 14% participated in the complementary round.

<sup>5</sup>There is a growing literature on how applicants’ behavior deviates from truth-telling in settings with Deferred Acceptance. Hassidim et al. (2017) examine data from various nations and markets, finding that a significant proportion of participants fail to disclose their true preferences. Hakimov and Kübler (2021) provide a comprehensive review of experimental studies in the field of centralized school choice and college admissions, highlighting findings related to deviations 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. However, it is not robust to changes in the order of the ranking, which we abstract from in our analysis.

<sup>6</sup>The quota for low-SES students reserves 15% of the seats for applicants from the poorest tercile of families (referred to as *priority students*), while the quota for high-performing students allocates 20% of the seats to students from the highest grade quintile in their previous school.

<sup>7</sup>There is a third option to remain on the waitlist for higher preferences, but this option is not commonly used.

<sup>8</sup>By law, every school that receives vouchers must accept students if the enrollment is less than the capacity they declared for the centralized assignment. Private non-voucher schools have a costly and selective admission process that usually begins earlier than the centralized admission system.

<sup>9</sup>During the period of our study (2020-2022), the Ministry of Education tracked the available seats on the website [vacantes.mineduc.cl](http://vacantes.mineduc.cl). This website calculated the number of available seats as the difference between capacity and current enrollment, and it was updated daily.

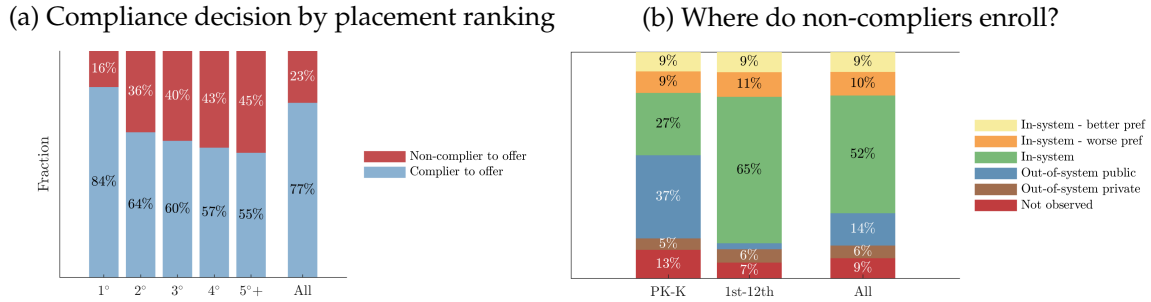


## II.B. The Chilean admission system: sample description

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

Column 1 of Table I presents the descriptive statistics of the applicant population. According to the Ministry of Education, 54% of the applicants are classified as low-SES students, who receive a higher voucher.<sup>10</sup> Additionally, 95% of applicants list an urban school as their first preference. In the main round, 76% of the participants are placed in one of their preferred schools, with 68%, 18%, and 8% assigned to their 1st, 2nd, and 3rd preferences, respectively. However, 23% of assigned applicants do not enroll in their placed school, and this fraction increases to 26% for the city of Santiago. We define “compliers” as placed applicants who enroll in their assigned school and “non-compliers” as those who do not enroll. Figure IIa illustrates that compliance rates decrease sharply with the placement ranking, with 84% of applicants placed in their 1st preference enrolling in the offered school, while only 55% of participants assigned to their 5th preference or lower comply with their placement.

Figure II  
Compliance to placement offers and final enrollment



Notes: Panel (a) shows in red the percentage of students who received a placement offer in the centralized application system but enrolled in a different school (non-compliant applicants). The first four columns are the subgroups' shares placed on 1st to 4th preference. The fifth column represents applicants assigned to the 5th or lower preferences, while the last column is the aggregate fraction for all placed applicants. Panel (b) describes where non-compliant applicants enroll. The first column represents the subgroup applying to PK or K, the second is those applying between 1st and 12th grade, and the third is the aggregate result for all non-compliers. “In-system - better(worse) pref” reflects a school that was in the ranking on a better(worse) preference than the placement offer. “In-system” represents a school not in the student’s ranking but could have applied to. “Out-of-system public(private)” is a school with an application process outside the centralized system and does(not) receive public funding.

To better understand non-compliance behavior, we examine the enrollment of participants who did not comply with their placement in the next academic year. Figure IIb shows the fraction of these students who enrolled in in-system and out-of-system schools, as defined at the beginning

<sup>10</sup>The Ministry of Education considers applicants from the poorest tercile of families as low-SES students, and they are referred to as *priority students*.



Table I  
Descriptive Statistics for Choice Applicants

	(1) All	(2) Survey respondent	(3) Estimation sample
<i>A. Applicants demographics</i>			
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 geocoding	0.60	0.67	1.00
Survey respondent	0.14	1.00	1.00
<i>B. Application and placement</i>			
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.

of this section. We find that 52% of non-compliant applicants enrolled in an in-system school they did not apply to, while 19% enrolled in a school they applied to but were not assigned to. Additionally, 14% of non-compliant applicants attend a publicly funded off-platform school (out-of-system public), and 6% attend a private off-platform school (out-of-system private). Notably, a significant 9% of students with an offer from the centralized application system are not observed in any regular school.<sup>11</sup> The figure also reveals that PK-K non-compliers are much more likely to enroll in an out-of-system public school compared to 1st-12th grade applicants. This difference is likely related to the availability of regular and language preschools that are exempt from participating in the centralized process, as previously described.

We highlight three heterogeneous behaviors of non-compliance. First, compared to mid and high-SES students, low-SES students are less likely to enroll in a school better than their placement preference (8% vs 12%), more likely to enroll in a worse option (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 they did not apply to, compared to only 35% of the non-voluntary group. Third, we observe a higher fraction of non-compliers enrolling in out-of-system private schools as the placement rank decreases: 3% of applicants assigned to their 1st preference, compared to 13% of applicants assigned to their 5th preference or lower. Figure B.II in Appendix A provides more details on these heterogeneous behaviors.

### 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 studies have done (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 aspects of the choice experience.<sup>12</sup> We use an expanded version of the 2020 survey sample utilized in Arteaga et al. (2022), which includes respondents from the 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 had concluded but before placements were announced, as shown in Figure I.<sup>13</sup> This timing allowed applicants to recall their experience while avoiding the influence of the results on their responses. Of those contacted, 203,252 (16%) completed the survey. Respondents closely resembled the overall population in terms of application patterns, though they were slightly less likely to be low-SES or rural (see column 2 of Table I).

<sup>11</sup>In Chile, PK and K are not mandatory; some of the non-observed students could be preschoolers staying at home. Furthermore, 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 choose this option, so we cannot distinguish between students who are not receiving education and those who are being homeschooled.

<sup>12</sup>See Appendix H for a translated version of the survey questions.

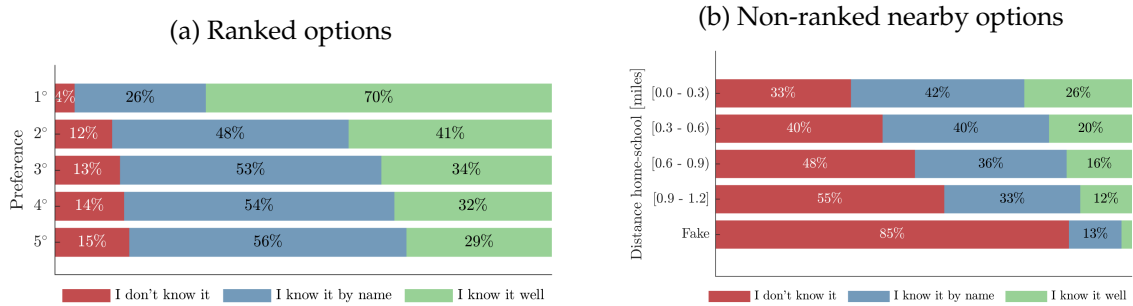
<sup>13</sup>The number of surveys sent differs from the total number of applicants because parents who filed applications for multiple students were surveyed on only one applicant, and some families did not have valid e-mail addresses.

### III.A. Survey findings

Our survey analysis focuses on diagnosing the level of information families have during the application process and its relation to application and enrollment decisions. Applicants were asked about their level of familiarity with the schools on their rank order lists and schools 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 they did not include in their portfolio.

Figure IIIa shows the responses to the question “How well do you know the schools in your application?”<sup>14</sup> 30% of the families indicated that they do not know their first-ranked school well, with 26% stating “I know it by name” and 4% stating “I don’t know it.”<sup>15</sup> Moving down the ranking from 1st to 2nd choice, we observe a sharp increase in the fraction of respondents who do not know the school well, from 30% to 60%, reaching 71% for the 5th choice. When we split our sample into students with mothers who have at most a secondary education (48%) and those with more educated mothers (52%), we find similar responses, with the former group declaring slightly less knowledge (see Figures B.IVa and B.IVb in Appendix A).

Figure III  
Knowledge level about schooling option



Notes: Panel (a) shows the responses to the question “How well do you know the schools in your application?” by position on the rank-order list. Panel (b) shows the answers to the question “Here are five schools. How well do you think you know these schools?” about schools that were not included in the rank-order list but were within 1.2 miles of the applicant’s home address. The last school is a made-up institution to check response quality.

We also find field evidence of limited knowledge about schools that families did not include in their ranking. For each applicant, we randomly selected between 1 to 5 schools that were not part of their application but were located close to their home address. We asked about the level of knowledge of these schools in the same manner as we did for the ranked options. Figure IIIb shows the distance of the school from home on the vertical axis, while the bars represent the average response. First, we observe that distance correlates with knowledge. 33% of respondents

<sup>14</sup> A screenshot of the question from the implemented survey is shown in Figure B.III in Appendix A

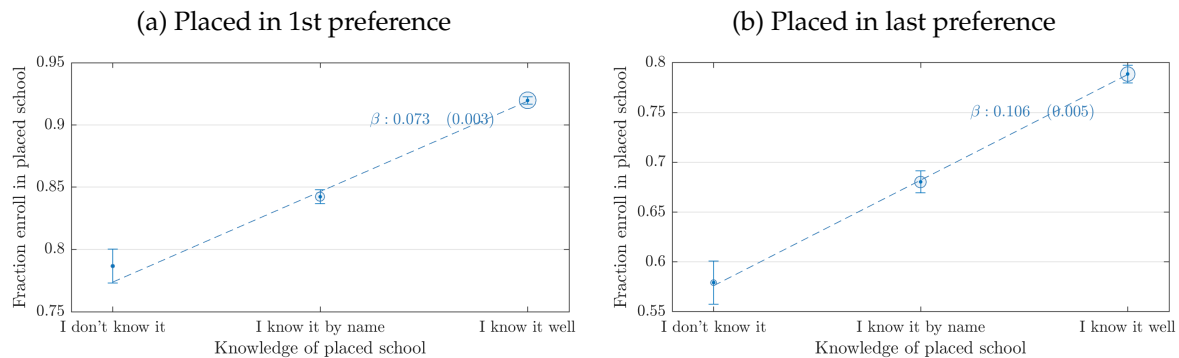
<sup>15</sup> Since we asked about schools they included in their ranking and some respondents said “I don’t know it,” we will not interpret the answers literally. Instead, we consider “I don’t know it,” “I know it by name” and “I know it well” as three ordinal levels of knowledge.

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 answered “I don’t know it.”

Since the meaning of “knowing a school well” is subjective, we provided survey respondents with a list of eight different steps and asked them to indicate 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 who selected each alternative. It appears that acquiring a comprehensive understanding of a school necessitates accessing extensive information, which can sometimes be costly. Notably, certain pieces of this information can be easily sourced from public records and platforms, such as the educational mission (93% said it’s necessary) or academic performance (93%), but others require more significant effort. 89% of respondents claim that knowing the infrastructure is essential, and 66% answered that an interview with a staff member is a necessary step.

We use the survey responses to explore the correlation between knowledge about the placed school and 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 is plausible that knowledge about a school is correlated with preferences, and preferences matter for enrollment. In an attempt to isolate the relationship between knowledge and enrollment, we use the responses to the question of hypothetical satisfaction as a control, which serves as a proxy for preferences. Figures [IVa](#) and [IVb](#) show the fraction of applicants who enroll in their placement, conditional on their level of knowledge, after residualizing for enrollment satisfaction. For students placed in their 1st preference, a decrease in knowledge from our highest to the lowest level is related to a decrease of 15 percentage points. For those placed in their last ranked option, the decrease is 21 percentage points. These results suggest that knowledge at the application stage matters in enrollment decisions.

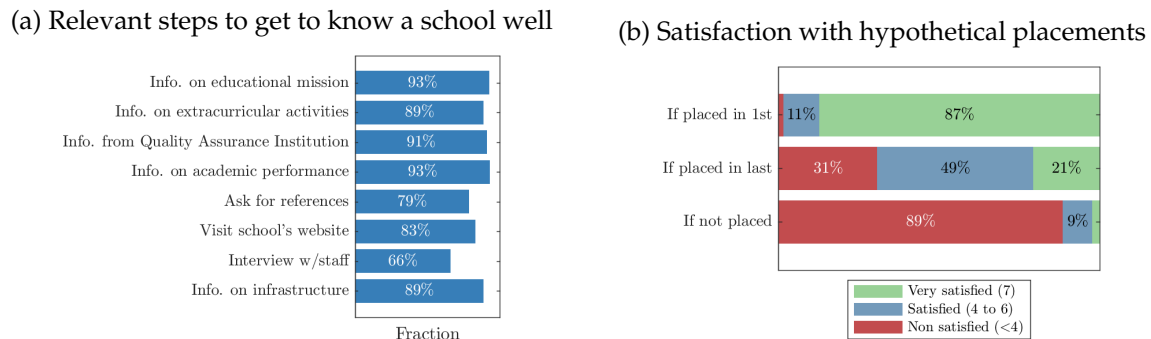
Figure IV  
Correlation of enrollment and knowledge about the placed school



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 stated satisfaction with hypothetical placement outcomes, 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 about how far down the ranking they are placed. We asked about satisfaction for three hypothetical results: being placed in their first preference, last preference, or not receiving an offer. As Panel Vb shows, almost 90% of the families told us they would be completely satisfied if they received their first option. As expected, satisfaction drops significantly if placed in their last preference; only 21% would be fully satisfied, and 31% gave a grade that is below the passing score according to Chilean standards. 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



Notes: Panel (a): answer to the survey question "When you add a school to your application, what do you consider a necessary step to know a school well before applying? (Check all that apply)". Panel (b): responses to the survey question "If [applicant name] get a seat in the following schools, from 1 to 7, how satisfied would you be?", and schools where "First preference", "Last preference", and "If you are not placed in any school". The scale is 1 to 7, the most common grading scale in Chile.

Panel A: stated satisfaction with hypothetical placement outcomes. Data are survey responses

to questions about applicant satisfaction with being placed at their first-ranked school, last-ranked school, and nonplacement. Sample: survey completers. Results reported on a 1-7 scale, with 7 being very satisfied and 1 being not at all satisfied.

#### IV. MODEL

We proceed with our analysis by introducing a dynamic model of school choice and enrollment that incorporates the possibility of noisy valuations of schools. This theoretical analysis serves two primary objectives. First, it demonstrates how uncertainty in school valuations and the availability of outside options influence students' decisions to (1) search for schools to include in their applications, and (2) enroll in schools to which they are assigned. Second, it defines a model that can be estimated and employed to evaluate the externalities of non-compliance and to estimate the effects of hypothetical changes in the market design.

Our analysis centers on an individual student who is searching for schools to add to her school choice application. She is cognizant of the fact 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 presented by [Arteaga et al. \(2022\)](#), which draws inspiration from job search models ([McCall, 1970](#)). The key distinction is that our model incorporates a dynamic component, allowing applicants to consider the outside option as an alternative to any placement outcome and to account for uncertainty regarding 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. In **stage 1**, applicants form beliefs about the utilities of the *enrollment options* at the schools they are familiar with and gather other inputs for the application decision. They decide on the ranking, search for additional schools, and submit the rank order list to the centralized platform.<sup>16</sup> In **stage 2**, each applicant receives a unique placement<sup>17</sup> and potentially learns more about this option. The applicant then decides whether to enroll in the placed school or pursue an outside option. We now describe both stages in detail.

**subsectionStage 1: Subjective Utility Formation Under Uncertainty, Search, and Application** In **stage 1**, applicants form beliefs about the utilities of the schools they are familiar with. This utility is derived from the enrollment option at each school, as 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 their expectations. Formally, the utility derived from enrollment

<sup>16</sup>Since we are modeling applications to pre-kindergarten through 12th grade, parents and students play a crucial role in the decision-making process. For simplicity, we will refer to them interchangeably as applicants, students, or families. Additionally, despite 27% of guardians filing an application for two or more students, our choice model does not directly consider the joint decision. However, we do consider the joint placement with siblings as an input for the enrollment decision.

<sup>17</sup>To be clear, non-placement is also a possible outcome. In our setting, 24% of applicants are not assigned to any of their preferred schools.

at school  $j$  for family  $i$  is  $U_{ij}$ . However, the perceived utility is composed of the true utility plus a noise term:  $U^{p1}_{ij} = U_{ij} + \eta_{ij}$ . The noise term is distributed as  $\sim N(0, \sigma_{k(i,j)})$ , and families are aware of its distribution conditional on their knowledge about the school ( $k(i, j)$ ). Nonetheless, they cannot differentiate their particular realization of noise from  $U_{ij}$ . Many aspects affect the valuation of a school, and families may overlook or value different attributes when they have imperfect knowledge. Resolving the noise implies shifts in utility in different directions; some applicants will discover positive news, while others will encounter negative news. When families incorporate this random term into their valuation, the Bernoulli utility in stage 1 is given by:

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

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

Given the uncertainty, families make choices based on a measure that is monotonically related to the expected value of  $U_{ij}^{b1}$ , which we will denote as  $EU^{s1}_{ij}$ . The curvature of the function  $f()$  reflects the effect of uncertainty on the utility. It is important to note that if families were risk neutral ( $f'' = 0$ ), the expectation of the utility at stage 1 would simply be the perceived utility ( $\mathbb{E}[U_{ij}^{b1}] = U_{ij}^{p1}$ ), since the noise term  $\eta_{ij}$  has a mean of zero.

To empirically estimate this function, we assume that  $f()$  is the constant absolute risk aversion (CARA) function with a risk parameter  $r > 0$ , which yields the following expression for the expectation:

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

It is important to note that  $U_{ij}^{p1}$  is known to the families and is therefore constant with respect to the expectation operator. The last line of equation 1 uses the fact that  $\exp(r\eta_{ij})$  follows a log-normal distribution, as  $r\eta_{ij} \sim N(0, r\sigma_{\eta_{ij}})$ . We then replace the expectation with the known analytical expression for the first moment.

The measure on which families base their choices 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, except for the term  $-\frac{r\sigma_{\eta_{ij}}^2}{2}$ . This new term indicates that uncertainty-averse families ( $r > 0$ ) perceive a lower subjective utility if their uncertainty aversion is higher ( $\frac{\partial EU_{ij}^{s1}}{\partial r} < 0$ ) or if the variance of the uncertainty is higher ( $\frac{\partial EU_{ij}^{s1}}{\partial \sigma_{\eta_{ij}}^2} < 0$ ).<sup>18</sup>

As we will describe in stage 2, families have the *option* to comply with the assigned school or choose an outside option. Therefore, what matters to them in stage 1 is the *enrollment option* utility of each school, denoted as  $w_{ij}$ . This utility is 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  $\xi_{ij}$  and  $\xi_{i0}$  are future preference shocks, known to the families only in stage 2. These shocks are assumed to be  $\overset{\text{iid}}{\sim} \text{EVI}$  and uncorrelated with  $\epsilon_{ij}$ .  $U_{i0}$  represents the observed utility of the outside option, which depends on the geographic supply of schools in the aftermarket.<sup>19</sup> The scale factor  $\lambda$  multiplying  $EU_{ij}^{s1}$  and  $U_{i0}$  allows the unobserved part of  $U_{ij}^{p1}$  (not yet introduced) and  $\xi_{ij}$  to have different variances.

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

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

<sup>18</sup>Apesteguia and Ballester (2018) notes that combining standard expected utility theory with additive unobserved utility results in non-monotonicity of choice probabilities with respect to risk preferences, an undesirable feature. However, our framework is immune to this critique since  $\epsilon$  is an additive component of  $U_{ij}^p$ , which is embedded into the Bernoulli utility function, as detailed in Section V. Subsequent algebraic manipulation of the expected utility generates a convenient additive unobserved utility component.

<sup>19</sup>We provide a detailed explanation of how we model the utility of the outside option in Section V.D.

We assume that families optimally rank schools they are familiar with based on  $w_{ij}$ , which is the dominant strategy when the allocation mechanism is Deferred Acceptance, as in our context. Let  $\Omega_i$  denote the set of known schools for family  $i$ ,  $\mathcal{C}_i \subset \Omega_i$  be the current rank order list containing  $N = |\mathcal{C}_i|$  schools, and  $p_{ij}$  represent 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 given by:<sup>20</sup>

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

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

Families will engage in a new iteration of a sequential search process if the increase in the value of the portfolio with an additional school is higher than the cost of searching for the new school. Let  $w_{is}$  denote the utility of the enrollment option for the “next school to be found,” which is an unknown object for families, and let  $\kappa_i$  represent the search cost. A new search iteration will occur if:

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

Assuming that the newly found school is added to the last position in the new portfolio<sup>21</sup>, the expected value of a new search iteration is given by:

$$\begin{aligned} & \mathbb{E}[(w_{is} - EU_{i0})p_{is} \prod_{j \leq N} R_{ij}] - \kappa_i > 0 \\ & \int (w_{is} - EU_{i0})p_{is} dF_i(EU_{is}^{s1}, p_{is}) \prod_{j \leq N} R_{ij} - \kappa_i > 0 \end{aligned}$$

The probability of a new search iteration occurring depends on (1) family  $i$ 's beliefs about the joint distribution of  $EU_{is}^{s1}$  and  $p_{is}$  ( $F_i(EU_{is}^{s1}, p_{is})$ )<sup>22</sup>, (2) the expected utility of the outside option, (3)

<sup>20</sup>Ordering schools from highest to lowest  $w_{ij}$  is the result of maximizing  $\mathcal{V}(\mathcal{C}_i)$  conditional on the choice set  $\Omega_i$  and knowing the DA rules. Any other order would shift placement probability from a preferred school to a less preferred school.

<sup>21</sup>We extend the analysis for cases where families add a school to a position other than the last in Appendix F. [Arteaga et al. \(2022\)](#) shows that among all the families who added a school to their initial portfolio, 86% added one in the last position.

<sup>22</sup>Recall that  $w_{is} = \log(\exp(\lambda EU_{is}^{s1}) + \exp(\lambda U_{i0}))$ , and the part of  $w_{is}$  that is unknown to the families is only  $EU_{is}$ .

the subjective belief about being placed in one of the already included schools ( $\prod_{j \leq N} R_{ij}$ ), and (4) the search cost ( $\kappa_i$ ).

Once the search process is complete, the family submits the rank order list  $\mathcal{C}_i$  to the centralized platform and awaits the allocation process.

Our modeling of stage 1 predicts at least five application behaviors. First, families will search more if they believe that schools they are not yet aware of are better and less congested ( $\uparrow \mathbb{E}[(w_{is}p_{is})]$ ). Second, a more attractive outside option ( $\uparrow U_{i0}$ ) makes searching less appealing.<sup>23</sup> Third, families that believe they are likely to be placed in one of the schools in  $\mathcal{C}_i$  ( $\prod_{j \leq N} R_{ij} \approx 0$ ) do not benefit significantly from additional search. Fourth, uncertainty about schools ( $\uparrow \sigma_\eta$ ) makes searching and extending the portfolio less likely by reducing  $w_{is}$ .<sup>24</sup> Fifth, if there are schools that are slightly known but not included in the ranking, there must be a “reservation utility” that justifies why families are not benefiting from the enrollment option at those schools.

#### IV.B. Stage 2: Placement Offers, Learning, and Enrollment Decision

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

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

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

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

### V. BRINGING THE MODEL TO THE DATA

Our objective is to estimate the parameters of the model that jointly describes school choice decisions and enrollment in placement offers. To accomplish this, we utilize the observed set of applicants, their submitted rank-ordered lists (ROLs), placement results, enrollment outcomes,

Hence, the expectation operator is over  $EU_{is}^{s1}$  (and  $p_{is}$ ).

<sup>23</sup>The proof and details of the effect of the valuation of the outside option on search are provided in Appendix F.

<sup>24</sup>This assumes that the search cost  $\kappa_i$  is the cost of becoming familiar with a school, but not necessarily being fully informed about it.

and survey data.<sup>25</sup>

In our model, families determine their ROLs by comparing the enrollment option utility  $w_{ij}$  among the schools in their choice set  $\Omega_i$ . Consequently, ROLs provide multiple pseudo-choices per applicant of the form  $w_{ir} > w_{ij}$ ,  $j \in \Omega_i \setminus \{1 \dots r\} \forall r \in \mathcal{C}_i$  (Train, 2009). By applying the rank-preserving transformation  $g_i(x) = \frac{1}{\lambda} \log(\exp(x) - \exp(\lambda U_{i0}))$  to each  $w_{ij}$ , we obtain a similar relation based on  $EU_{ir}^{s1}$  instead of  $w$ :  $EU_{ir}^{s1} > EU_{ij}^{s1}$ ,  $j \in \Omega_i \setminus \{1 \dots r\} \forall r \in \mathcal{C}_i$ .<sup>26</sup>

As detailed in Section IV.A,  $EU_{ij}^{s1}$  is a function of the perceived utility  $U_{ij}^{p1}$  and the uncertainty penalization term  $\frac{r\sigma_{\eta_{ij}}^2}{2}$ . The perceived utility comprises the *true* utility and a noise term  $\eta$ . Since we do not observe the noise, we model it as a random component. Following standard practice in the school choice literature (Agarwal and Somaini, 2020), we distinguish between the observed component  $V_{ij}$  and the unobserved (to us) part of the utility  $\epsilon_{ij} \stackrel{\text{iid}}{\sim} EVI$ . Combining these definitions, we arrive at the expression  $EU_{ij}^{s1} = V_{ij} + \eta_{ij} - \frac{r\sigma_{\eta_{ij}}^2}{2} + \epsilon_{ij}$ . Conditioning on the random terms yields an analytical expression for the probability of the pseudo-choices  $r \in \mathcal{C}_i$  (McFadden, 1974):

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

The likelihood of the ROL  $\mathcal{C}_i$  is the product of the probabilities of the individual pseudo-choices (Beggs et al., 1981). The second decision in our model is whether to enroll in the placement offer  $z(i)$  or opt for the outside option. Families choose to enroll in  $z(i)$  if the expected utility at stage 2,  $EU_{iz(i)}^{s2}$ ,<sup>27</sup> exceeds the utility derived from the outside option,  $\lambda U_{i0} + \xi_{i0}$ . Since both utilities have a component that follows an *EVI* distribution, conditioning on the random terms  $\epsilon_{iz(i)}, \boldsymbol{\beta}^\sigma, \boldsymbol{\eta}$  yields an analytic expression for the probability of  $EU_{iz(i)}^{s2} > \lambda U_{i0} + \xi_{i0}$ :

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

In the following subsections, we elaborate on how we map data to the various components of the enrollment option utility  $EU_{ij}^{s1}$ , the outside option utility  $U_{i0}$ , and the choice set definition  $\Omega_i$ . We conclude the section by presenting the likelihood function that we maximize to estimate our

<sup>25</sup>We will not estimate parameters related to school search and portfolio construction, as we only observe the outcome of this process (ROLs) and lack data on the sequential process described. Additionally, we did not collect survey data on beliefs about options not considered (to approximate  $F_i(EU_{is}^{s1}, p_{is})$ ) or search costs ( $\kappa_i$ ). These parameters cannot be identified without imposing strong assumptions.

<sup>26</sup>The intuition behind the effectiveness of the transformation  $g_i(\cdot)$  lies in the fact that the outside option for each choice  $j$  is the same for applicant  $i$ .

<sup>27</sup>As a reminder,  $EU_{iz(i)}^{s2}$  (stage 2) differs from  $EU_{iz(i)}^{s1}$  (stage 1) in three aspects: (1) the uncertainty penalization depends on the variance of  $\tau \times \eta_{iz(i)}$  instead of  $\eta_{iz(i)}$ , allowing for potential learning; (2) it includes a new preference shock  $\xi_{iz(i)} \sim EVI$ ; and (3) the utility, excluding  $\xi_{iz(i)}$ , is scaled by  $\lambda$  to allow  $\epsilon_{iz(i)}$  and  $\xi_{iz(i)}$  to have different variances.

model parameters.

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

Following the literature, we assume a linear functional form for the observed portion of the utility  $V_{ij} = v_i(W_j, D_{ij}, X_i)$ . Here,  $W_j$  is a vector of characteristics for school  $j$ , which includes the number of grades offered, mean cohort size, fee<sup>28</sup>, fraction of enrolled low-SES students, a dummy for charter status (private with public subsidy), math test score, and language test score.  $D_{ij}$  is a matrix of individual-specific school attributes, including distance to the school and a dummy for having a sibling enrolled.  $X_i$  is a vector of applicant  $i$ 's attributes, which includes dummies for low-SES, female, voluntary change, application year, and school level.<sup>29</sup>

$$\begin{aligned} V_{ij} &= (\gamma + \gamma_X X_i) D_{ij} + (\beta + \beta_X X_i + \beta_i^\sigma) W_j + \zeta_j \\ &= (\gamma + \gamma_X X_i) D_{ij} + (\beta_X X_i + \beta_i^\sigma) W_j + \delta_j \end{aligned}$$

To account for the fact that families may differ in their valuation of these characteristics, the functional form of  $V_{ij}$  incorporates both observable and unobservable preference heterogeneity. The vectors of parameters  $\gamma_X$  and  $\beta_X$  represent the observed preference heterogeneity. Unobserved heterogeneity is captured by the random component  $\beta_i^\sigma$ , which we assume follows a normal distribution with a diagonal variance matrix  $\Sigma_\beta$ . The term  $\delta_j$  represents the components of the indirect utility of school  $j$  that are equally perceived among applicants: the common preference over schools' observed attributes  $\beta W_j$  and the unobserved attributes summarized in  $\zeta_j$ .

#### V.B. Noise term ( $\eta_{ij}$ )

In our model, families cannot distinguish between the noise and the *true* utility. They have a sense of the magnitude of this noise, reflected in their belief about the second moment of its distribution. As researchers, we also do not observe the noise, so from an econometric perspective, it is treated as 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  $k(i, j) \in \{1 \dots 3\}$  as the answer of applicant  $i$  about school  $j$ , we assume that the distribution of  $\eta_{ij}$  in stage 1 is as follows:

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

<sup>28</sup>Only 21% of the schools participating in the centralized admission system charge a fee, and they represent 25% of the enrollment among participating schools in 2022. Details are provided in Table B.II in Appendix B.

<sup>29</sup>In the Chilean context, school levels do not perfectly map to the U.S. system of elementary, middle, and high school. Chile has a system of pre-básica (ages 4 to 5), básica (ages 6 to 14), and media (ages 15 to 18). We use the divisions of ages 4 to 5, 6 to 12, and 13 to 18.

This specification implies that families who answered “I know it well” face zero noise for that particular school, and thus, their perceived utility is equivalent to the true utility.

The uncertainty enters the expected utility in stage 1 through the uncertainty-penalization term  $\frac{r\sigma_{\eta_{ij}}^2}{2}$ , where  $\sigma_{\eta_{ij}}$  is family  $i$ ’s belief about the second moment of the noise distribution, and  $r$  is the risk parameter of the CARA Bernoulli utility function. We are unable to separately identify  $r$  and  $\sigma_{\eta_{ij}}$ , so we will estimate the parameters  $\rho_1^{s1}$  and  $\rho_2^{s1}$ , which represent the uncertainty-penalization term:

$$\frac{r\sigma_{\eta_{ij}}^2}{2} = \rho_{k(i,j)}^{s1} = \begin{cases} \rho_1^{s1} = \frac{r\sigma_{\eta_1}^2}{2} & \text{if } k(i,j) = 1 : \text{“I don’t know it”} \\ \rho_2^{s1} = \frac{r\sigma_{\eta_2}^2}{2} & \text{if } k(i,j) = 2 : \text{“I know it by name”} \\ \rho_3^{s1} = 0 & \text{if } k(i,j) = 3 : \text{“I know it well”} \end{cases}$$

Since families’ beliefs about the variance of the noise in stage 2 may change, we will estimate  $\rho_1^{s2}$  and  $\rho_2^{s2}$  as the uncertainty-penalization terms included in the enrollment decision ( $\rho_3^{s2} = 0$ ).

#### *V.C. Choice Sets ( $\Omega_i$ )*

Identification of the parameters that define the indirect utility function relies on comparing attributes of chosen options with those of other considered alternatives, which requires a choice set definition (Agarwal and Somaini, 2020). Unfortunately, we do not observe applicants’ complete choice sets  $\Omega_i$ , 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 reside (Abdulkadiroglu et al., 2020; Ainsworth et al., 2023; Bodéré, 2023, for example). However, based on the limited awareness about options that families declared, this approach will likely result in biased estimates, as we would be considering irrelevant alternatives (McFadden, 1978). We are in the realm 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 preferences, arguing that

it is a more robust assumption.

As attempts to overcome the problem of unobserved choice sets, we see two ways to estimate our model with two choice set definitions. First, as our main specification, we define the choice set as the schools on the ROLs ( $\Omega_i = \mathcal{C}_i$ ). This approach guarantees that we are inferring preferences from real trade-offs that families make. The downside is that we can only use for the estimation 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 breaks the independence of irrelevant alternatives (IIA) property of plain Logits (Guevara and Ben-Akiva, 2013).<sup>30</sup> This empirical approach of using subsets of the true choice sets has been labeled “differencing out” (Crawford et al., 2021) and was pioneered by McFadden (1978).

The second approach is based on the alternative specific consideration (ASC) model introduced by Manski (1977).<sup>31</sup> 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 chosen alternatives, becoming computationally infeasible with many options (Abaluck and Adams-Prassl, 2021; Crawford et al., 2021). For settings like ours, they suggest following a simulated choice sets approach implemented by Sovinsky Goeree (2008), who estimated a demand model for home PCs in a universe with 2,112 options and unobserved choice sets. Our approach is similar to theirs.

The method uses simulation to approximate the integration over all potential choice sets. The procedure starts by calculating a consideration probability  $\hat{p}_{ij}^c$  for each potential option  $j \in \{1 \dots J_i\}$  of applicant  $i$ . In each simulation  $s$ , we draw  $J_i$  uniform random variables  $uijs$  for all  $i$ . The inclusion of alternative  $j$  in the *simulated* choice set of  $i$  in simulation  $s$  is defined by the Bernoulli variable  $b_{ijs} = \mathbb{1}(\hat{p}_{ij}^c > uijs)$ . Our approach differs from Sovinsky Goeree (2008) in how we calculate  $\hat{p}_{ij}^c$ . In Sovinsky Goeree (2008), the probabilities are calculated endogenously using advertisement measures as consideration shifters that don’t affect choice probabilities.<sup>32</sup> 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.<sup>33</sup>

<sup>30</sup>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 models, there seems to be no exact procedures to estimate consistent parameters from subsets of true choice sets. As Crawford et al. (2021) state, “To the best of our knowledge, in the context of cross-sectional data, 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).”

<sup>31</sup>The method is also labeled as the “integrating over approach” in Crawford et al. (2021) or “ARC” in Barseghyan et al. (2021). Abaluck and Adams-Prassl (2021) and Barseghyan et al. (2021) describe it and derive identification results.

<sup>32</sup>Abaluck and Adams-Prassl (2021) prove 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.

<sup>33</sup>The results using this second approach are not included in this dissertation but will be available in future versions of the working paper linked to this chapter.



### 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 applicant's area of interest, which we define as the centroid of the schools the student is applying to. We consider four types of schools, counting only alternatives that offer education for the student's grade and gender. First, we measure the density of schools that participated in the centralized platform as an indicator of the richness of the process, which we expect to be negatively related to the value of the outside option. Second, we count the number of publicly funded schools that are not part of the centralized system, including schools that only offer elementary education (PK to K) and schools with ad-hoc admission processes, such as those focused on arts or sports. Third, we group the schools that provide education for students with special language needs.<sup>34</sup> Fourth, we count the number of fully private schools around the centroid.

We allow for observable preference heterogeneity by interacting the availability of each type of alternative with dummies that reflect whether the student has low socioeconomic status (SES) and whether the student is applying voluntarily.<sup>35</sup> Low SES families and applicants who are voluntarily changing schools may assess the outside options differently.

Additionally, we include two attributes in the outside option that are inherently characteristics 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 sibling's rank-ordered list (ROL) had any overlap with  $\mathcal{C}_i$ . We then checked if the applicant was placed in the same school as the sibling. If families prefer to have their siblings in the same school, being placed in different schools should reduce the likelihood of enrolling in the placed school. As a second attribute, we add a dummy for the placement ranking to account 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 who enrolls in their placement offer takes the following form:<sup>36</sup>

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

<sup>34</sup>Schools that offer a curriculum for students with language problems have a different admission process that allows them to screen 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 can easily obtain a medical certificate that allows them to apply.

<sup>35</sup>We define an applicant as voluntary if they are currently enrolled in a school that offers the next grade.

<sup>36</sup>The likelihood for applicants with placement but who decide not to enroll is very similar, but with  $s(i) \neq z(i)$  instead of  $s(i) = z(i)$ , which implies  $EU_{iz(i)}^{s2} < U_{i0} + \xi_{i0}$ . For applicants who are not placed ( $z(i) = 0$ ), the likelihood is simply  $P(\mathcal{C}_i)$ .

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

$$\begin{aligned}
L_i &= \int P(w_{ir} > w_{ij}, \forall j \in \Omega_i \setminus \{1 \dots r\}, \forall r \in \mathcal{C}_i | \beta^\sigma, \eta) \times \\
&\quad \left( \int P(EU_{iz(i)}^{s2} > U_{i0} + \xi_{i0} | \epsilon_{iz(i)}, \beta^\sigma, \eta) dF(\epsilon_{iz(i)} | \beta^\sigma, \eta) \right) dF(\beta^\sigma, \eta) \\
&= \int \left( \prod_{r \in \mathcal{C}_i} \frac{\exp(V_{ir} + \eta_{k(i,r)} - \rho_{k(i,r)}^{s1})}{\sum_{j \in \Omega_i \setminus \{1 \dots r-1\}} \exp(V_{ij} + \eta_{k(i,j)} - \rho_{k(i,j)}^{s1})} \times \right. \\
&\quad \left. \int \frac{1}{1 + \exp(\lambda V_{i0} - \lambda (V_{iz(i)} + \tau \times \eta_{k(i,z(i))} - \rho_{k(i,z(i))}^{s2} + \epsilon_{iz(i)}))} dF(\epsilon_{iz(i)} | \beta^\sigma, \eta) \right) dF(\beta^\sigma, \eta)
\end{aligned}$$

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 their most preferred school where their lottery number is above the cutoff, it is very likely that the unobserved part of the placed school's utility,  $\epsilon_{iz(i)}$ , is not *iid* EVI.<sup>37</sup> This is because schools with higher unobserved utility  $\epsilon_{ij}$  are more likely to be ranked at the top of the list. To generate approximate draws from the distribution of  $\epsilon_{iz(i)} | \beta^\sigma, \eta$ , we follow a two-step procedure that is 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 present the results of the choice process that describe the weights on school characteristics defining the indirect utility function of school enrollment. Then, we discuss 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 are relative to the effect of 1 mile for the  $X_i = \mathbf{0}$  student (male, mid-high-SES, non-voluntary, applying to elementary school in 2020), so they can be interpreted as the willingness to travel. For example, the  $X_i = \mathbf{0}$  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

<sup>37</sup>As Abdulkadiroğlu et al. (2017) point out, the assumption on the relationship between ranking and utilities restricts the values of the unobserved terms.

distance, while we allow for observed preference heterogeneity, which makes interpretation less direct. As an example, in our case, an extra mile for the  $X_i = 0$  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 ( $\text{HighSch} = 1$ ).<sup>38</sup>

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 to distance compared to younger applicants. Female applicants penalize distance marginally more than male applicants. Also, applicants in 2021 and 2022 are less willing to travel than those in 2020.<sup>39</sup> The second row shows that families strongly prefer schools with siblings, but the relevance decreases in upper grades.

Applicants demonstrate heterogeneous preferences for school attributes. This diversity is evident in Panel B of Table II. Female students put more weight on math test scores and less on language test scores compared to males. Low SES applicants have a higher taste intensity for charter schools and schools with larger enrollment, and they pay less attention to language test scores, but the same attention to math scores than higher SES students. 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 school students, the school size is more relevant – both the number of grades and enrollment – and they care less about the SES of the student body or math test scores. Additionally, high school applicants are more inclined toward charter schools than applicants to lower grades.

---

<sup>38</sup>As an example, the additional willingness to travel to attend a charter school for the last described high school applicant is  $-(0.184)/(-1 + 0.57) = 0.3$ .

<sup>39</sup>As a response to the COVID-19 pandemic, classes in Chile were fully remote from mid-March 2020 to June 2021. Applicants in 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)
	$\gamma$	Observable heterogeneity ( $\gamma_X$ )						
		Female	Low SES	Voluntary	MidSch	HighSch	2021	2022
Distance (1 mile)	-1.000 (0.026)	-0.060 (0.020)	0.091 (0.021)	-0.051 (0.024)	0.011 (0.032)	0.564 (0.026)	-0.119 (0.025)	-0.081 (0.024)
Sibling	7.390 (0.167)	-0.097 (0.139)	-0.836 (0.140)	0.019 (0.210)	-1.780 (0.182)	-3.451 (0.176)	0.359 (0.176)	-0.572 (0.164)
	$\sigma_\beta$	Observable heterogeneity ( $\beta_X$ )						
		Female	Low SES	Voluntary	MidSch	HighSch	2021	2022
# of grades offered		-0.008 (0.040)	0.047 (0.040)	0.133 (0.047)	-0.047 (0.076)	0.524 (0.104)	0.076 (0.048)	0.133 (0.046)
Fee		-0.006 (0.029)	-0.044 (0.030)	-0.155 (0.034)	0.000 (0.039)	0.059 (0.039)	0.062 (0.035)	0.019 (0.033)
Share of low SES		-0.087 (0.050)	0.191 (0.053)	-0.138 (0.061)	0.151 (0.069)	0.438 (0.071)	-0.061 (0.062)	-0.136 (0.058)
Charter		-0.057 (0.041)	0.135 (0.042)	-0.030 (0.050)	0.123 (0.059)	0.184 (0.067)	-0.053 (0.050)	-0.127 (0.047)
Enrollment per grade		0.006 (0.027)	0.063 (0.026)	0.037 (0.029)	-0.015 (0.052)	0.182 (0.050)	0.049 (0.031)	-0.035 (0.031)
Math test score	1.578 (0.044)	0.189 (0.073)	-0.007 (0.074)	0.020 (0.086)	0.086 (0.104)	-0.481 (0.112)	-0.097 (0.087)	-0.006 (0.083)
Language test score		-0.139 (0.062)	-0.179 (0.063)	0.123 (0.073)	-0.245 (0.091)	-0.089 (0.096)	-0.027 (0.075)	0.013 (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 who answered the middle level of knowledge ( $k(i, j) = 2$ ). Estimates suggest that the noise is substantial compared to the people who declared that they know the school well ( $k(i, j) = 3$ ). A noise realization from the 20th percentile of the distribution is equivalent to moving the school 1.7 miles further from home (or closer if the noise comes from the 80th percentile).

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

Table III  
Noise standard deviation and uncertainty penalization estimates

Noise	Stage 1 $\boldsymbol{\eta}$	Stage 2 $\tau \times \boldsymbol{\eta}$
<i>A. Standard deviation of noise</i>		
$\sigma_1$	-5.768 (0.146)	4.404
$\sigma_2$	-4.251 (0.087)	3.246
<i>B. Uncertainty-penalization</i>		
$-\rho_1$	-8.169 (0.087)	-10.375 (0.444)
$-\rho_2$	-5.226 (0.055)	-6.097 (0.273)
<i>C. Other parameters</i>		
$\lambda$		0.528 (0.019)
$\tau$		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  $\lambda$  and  $\tau$ .  $\lambda$  is the ratio of standard deviations of the unobserved portion of  $EV_{iz(i)}^{s1}$  and the preference shock realized in stage 2 ( $\xi_{iz(i)}$ ).  $\tau$  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  $\tau$ .

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

The estimated shrink parameter  $\tau$  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 stage 1 to stage 2. Our model predicts it should shrink at a  $\tau^2 = .55$  rate. We provide two hypotheses for this result. First, it could be that families' beliefs about  $\sigma_\eta$  do not update at the same rate as the shrinkage of the true variance of  $\eta$ . Second, the stakes in stage 2 are higher since it involves the enrollment decision instead of the application decision (or potential 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  $\sigma_\eta$  are correct.

### 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 the updated perceived enrollment utility, which includes the potential learning, manifested in the shrinkage of application noise at rate  $\tau$ , and preference shocks, represented by  $\xi_{iz(i)}$ .

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 that families facing more availability of schools in the centralized application platform (in-system schools) 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 who live around private schools that don't receive vouchers value the outside option more. 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 participating in the centralized system are responsible for two or more students who have at least one school in common between applications. We observe that applicants placed without other family members are less likely to enroll in their 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 non-compliance behavior. We start with simple scenarios in which we run the assignment algorithm after dropping the preferences that non-compliers will not enroll in. We continue with model-based counterfactuals, where we

Table IV  
Outside option estimates

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



evaluate the changes in allocation when an information campaign is implemented, varying its effectiveness and including out-of-system alternatives in the centralized process.

#### VII.A. Baseline scenario

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 rank-ordered lists (ROLs). We construct the observed portion of expected utility ( $V_{ij}$ ) using the model's estimated parameters, schools, and applicants' characteristics ( $W_j$  and  $X_i$ ). Once we predict the level of knowledge for each school in the choice set  $k(i, j)$ , we can map to a specific distribution of noise ( $\eta_{k(i, j)}$ ) and uncertainty penalization term ( $\rho_{k(i, j)}$ ). We simulate the unobserved portions of the utility ( $\epsilon_{ij}$ ). With those inputs, we calculate the expected utility ( $EU_{ij}^{s1}$ ) of each school and build the ROLs ( $C_i$ ).

After obtaining the ROLs, we proceed to run the Deferred Acceptance algorithm using the real school capacities and 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  $\xi_{i0}$ ), and generate compliance decisions ( $EV_{iz(i)} > U_{i0} + \xi_{i0}$ ).<sup>40</sup>

Our simulated baseline scenario closely resembles the real scenario in two crucial aspects. First, the percentage of students assigned to their preferences or left unassigned is nearly identical. Second, the compliance rate is comparable, with only a 1 percentage point (pp) difference. This good fit persists when we disaggregate the measures at the urban zone level.<sup>41</sup>

In the following section, we describe the counterfactuals that we will compare with the baseline scenario. We begin with a simple exercise of calculating the allocation assuming non-compliers do not apply to the school they were placed in. We then proceed to model-based counterfactuals, where we simulate the effects of an information campaign and include out-of-system options in the centralized process.

#### VII.B. Mechanical Counterfactuals

Families that do not comply with the placement offer must apply to a school without the convenience of the centralized system. This decision may be optimal given the new information acquired and/or the presence of out-of-system schools. However, non-compliance also generates a negative externality on other families who would have preferred the school assigned to a non-complier over their own placement.

We aim to determine the hypothetical placement if assigned applicants who do not enroll in

<sup>40</sup>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.

<sup>41</sup>For the counterfactuals, we consider 70 urban zones, omitting only very small geographic areas from all urban areas in Chile, that are also not included in the estimation sample of our main model.

the offered school had not applied to that school in the first place. Since we observe preferences, placement, and compliance, we can replicate the assignment using the allocation rules of the implemented version of the Deferred Acceptance algorithm in Chile. This allows us to evaluate any Rank Ordered Lists (ROLs) changes 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 to, and (2) when they do not apply to the assigned school or any preference ranked below it. These counterfactuals allow us to quantify the externalities that non-compliers impose on the rest of the applicants. Our group of interest is students with room for improvement, which includes applicants who were placed in their second or lower preference and complied with the offer and applicants who were not placed in any of their preferences. Those two groups represent 46% of all applicants.

### VII.C. Information Campaign

Ideally, we would like to simulate the effect of policies aimed at ensuring families are more informed about application and enrollment decisions. However, 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 leverage the fact that we observe the outcome of the search process that non-compliers undertake: the school where they ultimately enroll after dismissing the centralized placement offer. If applicants had applied to the school they ended up enrolling in from the beginning, compliance would have been less of a problem.

We introduce a counterfactual policy in which an information campaign aims 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, as if it could perfectly predict 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 for whom we correctly predict the enrollment decision. For the  $1 - \alpha$  fraction of applicants, we provide a “naive recommendation”: the most popular feasible school within 2 miles that was not included in the ranking.

When a school is recommended, families form beliefs about it and decide whether to add it to their ranking. When the recommended school is the school the student ends up enrolling in ( $q(i) \equiv s(i)$ ), 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 ( $s(i)$ ) must be greater than 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:  $\epsilon_{iq(i)} \text{ s.t. } EU_{iq(i)}^{s2} > \lambda U_{i0} + \xi_{i0}$ . When the suggested school is the naive recommendation, we construct the observed utility using 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 recommended by the policy. We introduce a parameter,  $\beta$ , to quantify this variation. When  $\beta = 1$ , families possess comprehensive knowledge about school  $q(i)$ , expressed as  $k(i, q(i)) = 3$ , eliminating any uncertainty penalty in the expected utility ( $EU_{iq(i)}^{s1}$ ). Conversely, when  $\beta = 0$ , families have the lowest level of information in our model ( $k(i, q(i)) = 1$ ). Any value of  $\beta$  between 0 and 1 represents a probability distribution over the different levels of information. As  $\beta$  approaches 0, the probability of  $k(i, q(i)) = 1$  increases. When  $\beta$  is close to 0.5, the probability of  $k(i, q(i)) = 2$  is highest, and as  $\beta$  approaches 1, the probability of  $k(i, q(i)) = 3$  becomes dominant. Figure B.V in Appendix A provides a precise mapping of probabilities for different values of  $\beta$ .

#### VII.D. Out-of-System Outside Options Available in Centralized Choice

In our final counterfactual, we simulate the inclusion of out-of-system publicly funded schools within the centralized platform, effectively making them “in-system” and allowing our information campaign to inform families about them. We will refer to this new set of on-platform schools as “included schools.” This counterfactual is motivated by the observation that 20% of non-compliers enroll in an out-of-system school, and among those, 70% enroll in a school that receives public funding but is exempt from participating in the centralized system.

To proceed with this counterfactual, we must make assumptions about school capacities and preferences over these newly included schools. In-system schools must declare their available seats to the centralized authorities, as this information is a key input for the allocation mechanism. However, out-of-system schools are not obligated to provide this information, and as a result, we do not observe their capacity. As an approximation, we assume that the capacity of the included schools is equal to their observed enrollment, which serves as a lower bound for the actual number of seats.

We assume that only applicants who are informed about the included schools through our campaign will add them to their preference list. By design, the information campaign is targeted solely to non-compliers, making this a restrictive assumption, as the presence of new schools in the centralized system will potentially affect the rank-ordered lists (ROLs) of all students. We argue that our results regarding the reduction in congestion in this counterfactual will reflect a lower bound since lifting the assumption and making the included schools available to all applicants will further reduce the pressure on the original set of in-system schools. The combination of the capacity and preference assumptions results in every applicant to the included schools being guaranteed a seat with certainty.

### VII.E. Counterfactual Results

Our objective is to evaluate the externalities produced by families placed by the centralized mechanism in a school but ultimately enrolling in a different institution, as well as the effect of including out-of-system schools in the centralized mechanism. The model allows us to evaluate the effect on overall placement while also considering the enrollment decision following the assignment. This is crucial because the final outcome that matters is enrollment. For example, improving a student’s placement will not be welfare-relevant if they choose the outside option regardless. We begin by presenting the changes in placement and enrollment decisions for the overall population of applicants. Throughout the analysis, we will refer to several groups of particular interest, defined by their placement and enrollment decisions at baseline: compliers assigned to their 1st preference (i.e., no room for improvement), compliers assigned to their 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
<i>A. Mechanical counterfactuals</i>						
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
<i>B. Model-based counterfactuals: oracle information campaign</i>						
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
<i>C. Model-based counterfactuals: naive information campaign</i>						
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
<i>D. Model-based counterfactuals: including out-of-system options in centralized platform</i>						
Oracle recommendation + 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 enrollment (columns 4 to 6), comparing counterfactuals to the baseline scenario for all applicants. The classification (Better, Same, or Worse) is based on the utility derived by the placed or enrolled school. 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.

The results are summarized in Table V. Columns 1 to 3 show the fraction of applicants placed in a better, same, or worse preference in each counterfactual scenario. Columns 4 to 6 reflect the changes in outcome after the enrollment decision. These two results could differ if, for example, an applicant is assigned to a better preference than the baseline but, in both cases, does not comply with the offer. In this situation, they would be classified as having a “Better” placement but

the “Same” enrollment. Additionally, we will discuss the results for the four groups of students mentioned above, with the detailed results displayed in tables B.V to B.VI in Appendix B.

Examining Panel A of Table V, we observe that if non-compliers do not apply to their offers, 5% of all applicants will experience an improvement in their allocation. If, in addition, non-compliers do not include any preference below their baseline placement, then 7% of applicants will be assigned to a better preference. These results are primarily driven by two groups of applicants: those who, at baseline, were assigned to their 2nd or lower preference, and those who were not assigned to any preference. Among these groups, the improvements are 12% (Table B.IV) and 11% (Table B.III), respectively. Interestingly, 9% of non-compliers (Table B.VI) will also experience an improvement in their placement.

When focusing on enrollment decisions, the results are more modest. Only 5% of all applicants are better off when non-compliers don’t apply to their assigned school or any school ranked below it. For the specific group of non-compliers, the fraction enrolled in a better preference than the baseline is only 4%, which is less than half of the proportion assigned to a better school (9%). This decline is due to the fact that non-compliers are likely to reject any placement offer. For students assigned to their 2nd or lower preference, the fraction of students in better positions remains the same when comparing enrollment to placement. This can be explained by the fact that they were already complying with a lower preference, so a better placement should also lead to compliance. The proportion of better-off non-placed students decreases from 11% to 8% when transitioning from placement to enrollment.

In the counterfactual scenario where we drop the placement offer from the ranking of non-compliers, we observe that 15% of total applicants have a worse placement than the baseline. This result is almost entirely explained by the non-compliant group, who are not assigned to any school after dropping their baseline offer from the rank order list. However, when examining enrollment results, only 1% of the students are in a worse position under this mechanical counterfactual. This can be attributed to the fact that non-compliers rejected the baseline offer anyway, so placement followed by rejection is equivalent to non-placement for them.

The results for the counterfactual that simulates the oracle information campaign are depicted in Panel B of Table V. This scenario mimics non-compliers applying to the school they will enroll in the future, and assumes they have full knowledge about these schools ( $\beta = 1$ ). As a result, 8% of the total students will be placed in a better preference. Most of the gain comes from better results for the group of non-compliers, with 40% benefiting from it. Still, 5% of the group of students with room for improvement are better placed due to reduced congestion in schools of their rankings. Interestingly, we observe that less than 1% of applicants end up in a worse school, which is a noteworthy result given that the oracle recommendation is for in-system schools and could potentially have displaced other applicants. When we analyze enrollment, we also observe that 8% of the total students are better off, the same proportion as better placed applicants. This happens because, by revealed preferences, the recommended school is better than the outside option since we observe them enrolled in it.

When we set the knowledge level of the suggested school to the lowest ( $\beta = 0$ ), the overall percentage of winners reduces from 8% to 6%. 30% of the target population of the policy, the non-compliers, are placed in a better school, instead of 40%. Since the congestion alleviation is less pronounced, the fraction of compliers who are better off reduces to 3%. In this case, the placement benefits do not translate into enrollment; the proportion of better-off applicants after the enrollment decision is only 2%. The reason behind this drop is that most of the new placements are rejected by the baseline non-compliers due to the penalization component in the utility of the new school, which originates from the uncertainty about it.

Examining the results of the naive recommendation policy in Panel C, which involves informing applicants about the most popular school not included in their ranking, we observe that the fraction of applicants who are better placed is higher than with the oracle campaign (11% vs 8%). This is expected, as the campaign reaches every applicant, not just non-compliers. However, the proportion of students who are worse off is more than seven times greater with the naive policy, resulting in a probable lower net benefit.<sup>42</sup>

Panel D of Table V describes the effect when we include out-of-system publicly funded schools in the centralized platform, effectively making them “in-system,” and implement the oracle campaign. We observe that 14% of applicants in the system would experience an improvement in placement, and 13% would secure a better enrollment. Notably, there is no increase in the proportion of applicants who are worse off. Including out-of-system publicly funded schools increases the effectiveness of the information campaign by 72%.

## 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 key findings. First, we observed a significant lack of information among families about neighborhood schools and the ones they apply to. Second, non-compliance emerges as a substantial issue, with 23% of applicants not enrolling in their assigned school and 70% of these non-compliers enrolling in schools they initially bypassed. Third, 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, which accounts for imperfect knowledge and uncertainty aversion, leads to two crucial insights. First, it demonstrates that uncertainty about schools adversely affects their perceived value, thereby decreasing compliance with placement offers. Second, the presence of 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, benefits targeted students by suggesting

---

<sup>42</sup>We are unable to provide precise overall welfare results, as we have not introduced any cardinal welfare measure or a method to aggregate it among applicants. We hope to make progress on this front in future versions of the working paper associated with this chapter.

new schools that would have not included otherwise and non-targeted students by reducing congestion externalities from non-compliers.

Furthermore, we evaluate the effectiveness of information campaigns based on the depth of information provided. Our findings suggest that a campaign with superficial information has an overall effect on improving enrollment of one-third of the magnitude compared to comprehensive information provision. This highlights the importance of the quality of the information campaign in a context with uncertainty aversion.

We also show that incorporating out-of-system schools in the centralized assignment could further improve enrollment outcomes. An information campaign that can also suggest schools that are publicly funded but do not participate in the centralized application would increase the effectiveness of the campaign by 72%.

Our research presents key challenges for policymakers. The success of information campaigns promoting new schools for application critically depends on the depth and quality of the information provided. Moreover, including out-of-system options in centralized applications could help 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.
- Abdulkadiroğlu, 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.
- Abdulkadiroğlu, A., Pathak, P. A., Schellenberg, J., and Walters, C. R. (2020). Do Parents Value School Effectiveness? *American Economic Review*, 110(5):1502–39.
- Abdulkadiroğlu, 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.
- Barahona, N., Otero, S., and Dobbin, C. (2021). Affirmative Action in Centralized College Admission Systems: Evidence from Brazil | Stanford Institute for Economic Policy Research (SIEPR). Technical report, Stanford Institute for Economic Policy Research (SIEPR).
- 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

May 9, 2024

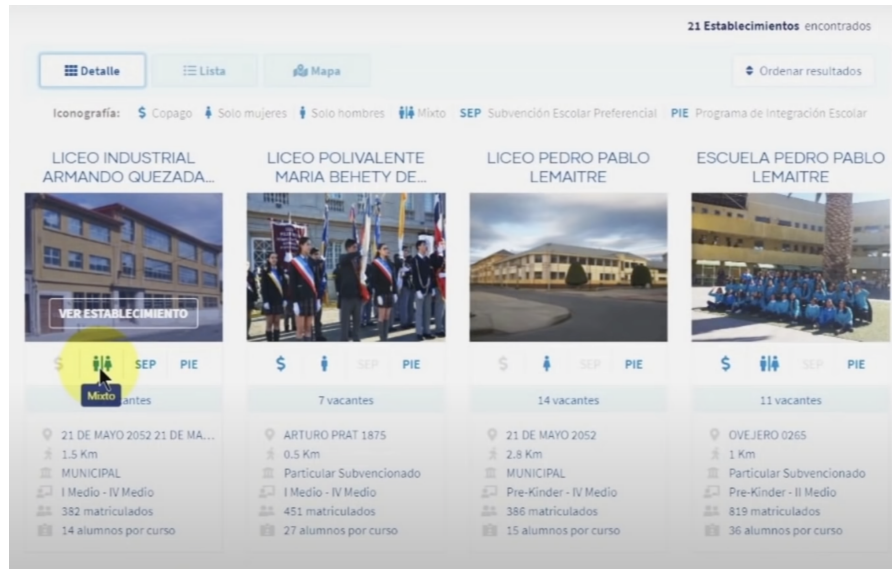
## CONTENTS

<b>A Additional Figures</b>	<b>3</b>
<b>B Additional Tables</b>	<b>7</b>
<b>C Simulated choice sets</b>	<b>13</b>
C.1 Consideration probability of unranked school . . . . .	13
<b>D Counterfactual simulations</b>	<b>16</b>
D.1 Inputs for counterfactuals . . . . .	16
D.2 Simulation procedure . . . . .	16
D.3 Home location imputation procedure . . . . .	19
<b>E Simulated maximum log-likelihood</b>	<b>20</b>
E.1 Two-step procedure . . . . .	21
<b>F Additional Proofs</b>	<b>23</b>
F.1 Better outside option reduces the value of search . . . . .	23
<b>G Knowledge and consideration prediction functions</b>	<b>24</b>
G.1 Knowledge prediction function for ranked school . . . . .	24
G.2 Knowledge prediction function for unranked school . . . . .	25
<b>H Survey Translation</b>	<b>26</b>

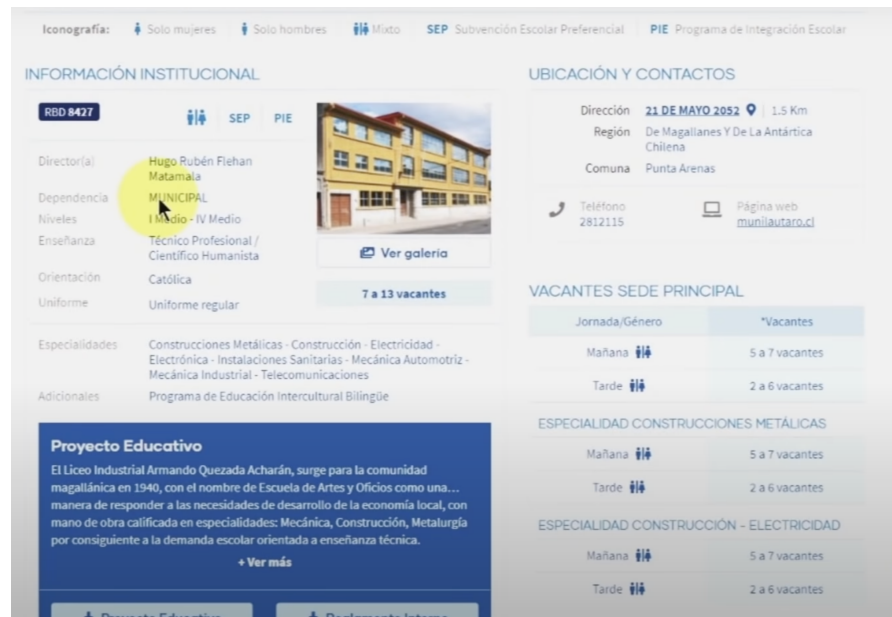
## A. ADDITIONAL FIGURES

Figure B.I  
2022 Online Application Platform Screenshots

(a) Gallery of schools

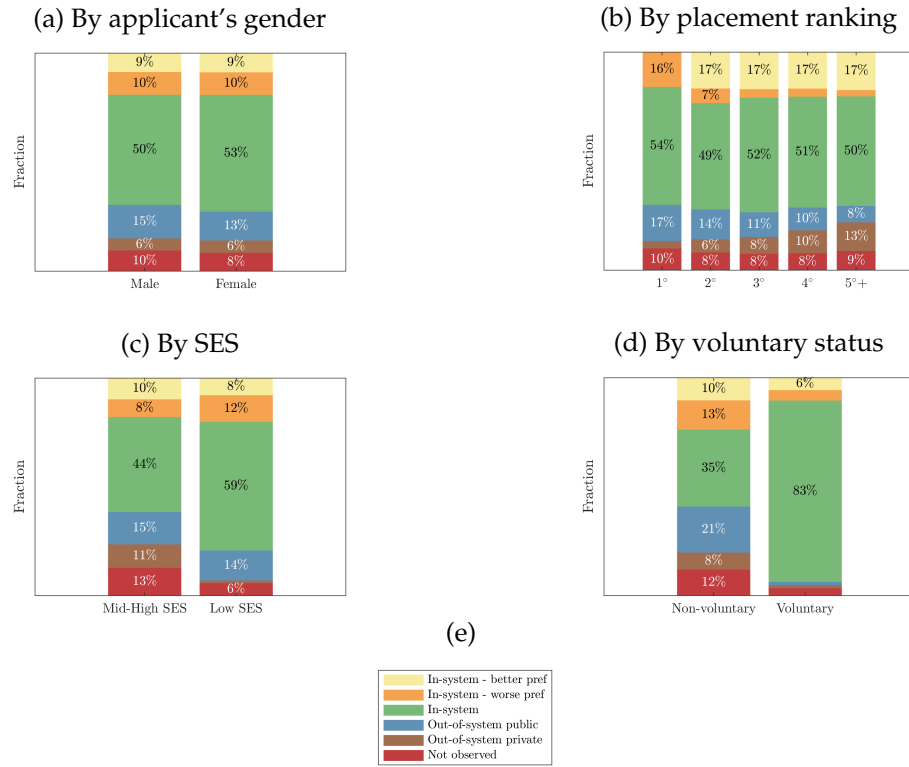


(b) Detailed Information of a School



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



Notes: "In-system - better(worse) pref" reflects a school that was in the ranking on a better(worse) preference than the placement offer. "In-system" represents a school not in the student's ranking but available to apply. "Out-of-system public(private)" is a school with an application process outside the centralized system and does(not) receive public funding.



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

(a) Translated version to English

**How well do you know the schools in your application?**

	I don't know it	Only by name	I know it well
1° Condorito School	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2° Neruda School	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3° El Litre School	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

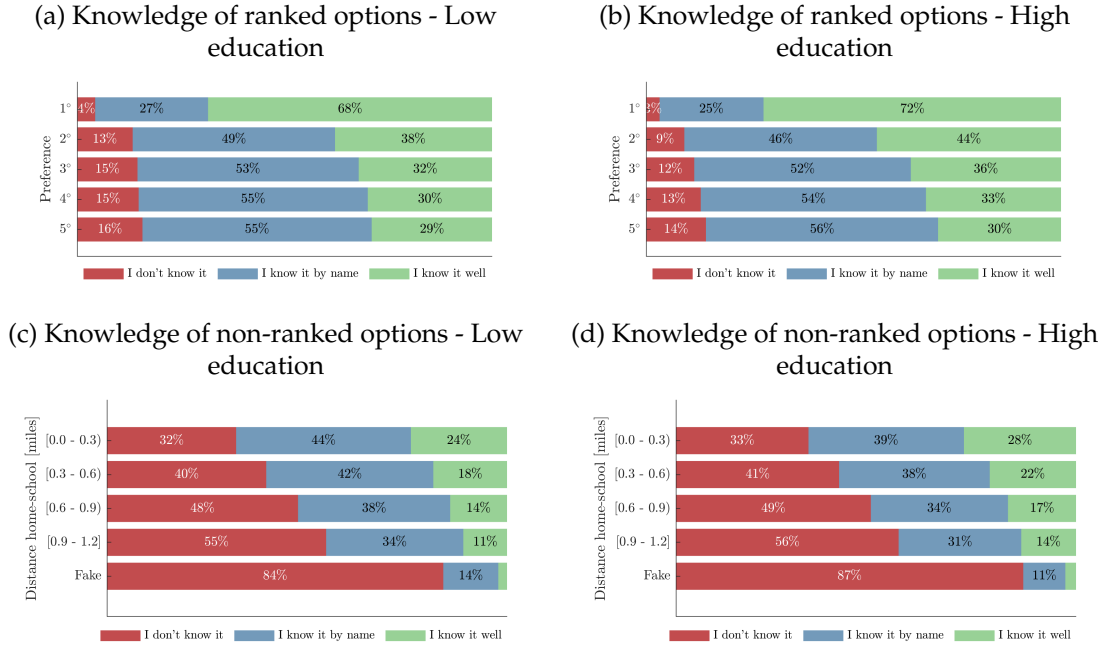
(b) Original version in Spanish

**¿Qué tan bien conoces los establecimientos de tu postulación?**

	No lo conozco	Solo de nombre	Lo conozco bien
1° Colegio Condorito	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2° Colegio Neruda	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3° Colegio El Litre	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

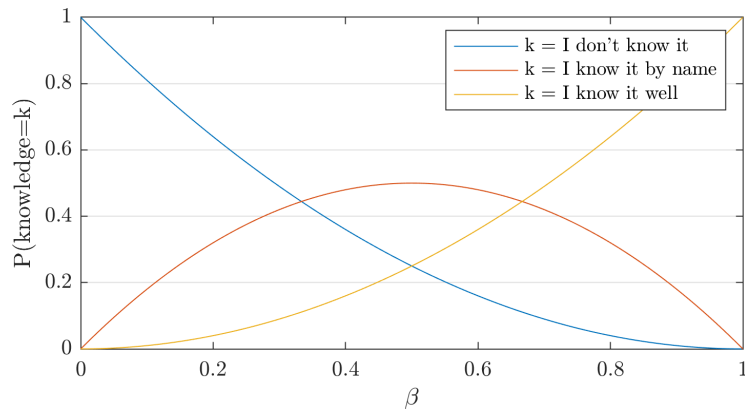
*Notes:* Figure shows how survey respondents saw the knowledge question. The survey was implemented in the Qualtrics platform.

Figure B.IV  
Knowledge level about schooling option by mother's 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). Panels (a) and (b) show the responses to the question "How well do you know the schools in your application?" by position on the rank-order list. Panels (c) and (d) show the answers to the question "Here are five schools. How well do you think you know these schools?" about schools not included in the rank-order list but within 1.2 miles of the applicant's home address. The last school is a made-up institution to check responses quality.

Figure B.V  
Probability of knowledge by  $\beta$



## 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: 2022 enrollment data, 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	
<i>A. Unweighted</i>					
# 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
<i>B. Weighted by enrollment</i>					
# 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
<i>A. Mechanical counterfactuals</i>						
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
<i>B. Model-based counterfactuals: oracle information campaign</i>						
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
<i>C. Model-based counterfactuals: naive information campaign</i>						
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
<i>D. Model-based counterfactuals: including out-of-system options in centralized platform</i>						
Oracle recommendation + internalizing out-of-system ( $\alpha = 1, \beta = 1$ )	0.088	0.912	0.000	0.065	0.935	0.000

*Notes.* This table shows the changes in placement (columns 1 to 3) and enrollment (columns 4 to 6), comparing counterfactuals to the baseline scenario for applicants who were not placed in any preference at baseline (26% of total). The classification (Better, Same, or Worse) is based on the utility derived by the placed or enrolled school. 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.

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
<i>A. Mechanical counterfactuals</i>						
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
<i>B. Model-based counterfactuals: oracle information campaign</i>						
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
<i>C. Model-based counterfactuals: naive information campaign</i>						
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
<i>D. Model-based counterfactuals: including out-of-system options in centralized platform</i>						
Oracle recommendation + internalizing out-of-system ( $\alpha = 1, \beta = 1$ )	0.089	0.902	0.008	0.088	0.901	0.011

*Notes.* This table shows the changes in placement (columns 1 to 3) and enrollment (columns 4 to 6), comparing counterfactuals to the baseline scenario for compliant applicants who were placed in a worse preference than the first at baseline so they could potentially get a better placement (20% of total). The classification (Better, Same, or Worse) is based on the utility derived by the placed or enrolled school. 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.

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
<i>A. Mechanical counterfactuals</i>						
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
<i>B. Model-based counterfactuals: oracle information campaign</i>						
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
<i>C. Model-based counterfactuals: naive information campaign</i>						
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
<i>D. Model-based counterfactuals: including out-of-system options in centralized platform</i>						
Oracle recommendation + internalizing out-of-system ( $\alpha = 1, \beta = 1$ )	0.000	0.994	0.006	0.000	0.993	0.007

*Notes.* This table shows the changes in placement (columns 1 to 3) and enrollment (columns 4 to 6), comparing counterfactuals to the baseline scenario for compliant applicants who were placed in first preference at baseline (39% of total). The classification (Better, Same, or Worse) is based on the utility derived by the placed or enrolled school. 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.

Table B.VI  
Counterfactual Results for Non-complier

	(1)	(2)	(3)	(4)	(5)	(6)
	Placement			Enrollment		
	Better	Same	Worse	Better	Same	Worse
<i>A. Mechanical counterfactuals</i>						
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
<i>B. Model-based counterfactuals: oracle information campaign</i>						
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
<i>C. Model-based counterfactuals: naive information campaign</i>						
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
<i>D. Model-based counterfactuals: including out-of-system options in centralized platform</i>						
Oracle recommendation + internalizing out-of-system ( $\alpha = 1, \beta = 1$ )	0.682	0.304	0.014	0.649	0.351	0.000

*Notes.* This table shows the changes in placement (columns 1 to 3) and enrollment (columns 4 to 6), comparing counterfactuals to the baseline scenario for non-compliant applicants (17% of total). The classification (Better, Same, or Worse) is based on the utility derived by the placed or enrolled school. 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.



### 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.<sup>1</sup> 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  $\hat{p}_{ij}^c$  with the function described on section C.1 for each potential option  $j \in \{1 \dots J_i\}$ .
3. For each simulation  $s \in \{1 \dots S\}$  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  $b_{ijs} = \mathbb{1}(\hat{p}_{ij}^c > u_{ijs})$ .

#### C.1. Consideration probability of unranked school

Considered schools for applicant  $i$  (or schools on her choice set  $\Omega_i$ ) 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

---

<sup>1</sup>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:

$$C_{ij} = \alpha_1 \times \text{distance}_{ij} + \alpha_2 \times \text{distance}_{ij}^2 + \beta \times \text{connection}_{ij} + \delta_{f(i)j} + \epsilon_{ij}$$

$\text{distance}_{ij}$  is the euclidean distance between home of applicant  $i$  and school  $j$ .  $\text{connection}_{ij}$  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  $\text{female}_i$  and  $\text{LowSES}_i$  ( $2 \times 2$ ).  $\epsilon_{ij}$  is an unobserved (to us) portion of  $C_{ij}$  and is assumed *IID Logistic*(0, 1).

We will assume that there is a threshold  $\kappa$ , 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  $\epsilon_{ij} \sim \text{Logistic}(0, 1)$ , the probability  $P(c(i, j) = a | \theta)$  has a simple analytical form once we condition on the vector  $\theta$ , that contains the parameters that define  $C_{ij}$  and the threshold.<sup>2</sup> The

---

<sup>2</sup>See Train (2009) for details.

log-likelihood function of observing  $c_{ij}$  for each school  $j \in \mathcal{S}_i$ <sup>3</sup> and survey respondent  $i \in \{1 \dots I\}$  is the following:

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

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  $\{\boldsymbol{\theta}_z\}_{z \in \{1 \dots 70\}}$ .

---

<sup>3</sup> $\mathcal{S}_i$  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,<sup>4</sup> 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 II 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 IV 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 ( $\lambda$  and  $\tau$ ), displayed in Table III.

resUncertainty

**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 are in Appendix G.1.

### *D.2. Simulation procedure*

We borrow the home location<sup>5</sup>, 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:  $\{\hat{p}_{ij}^1, \hat{p}_{ij}^2, \hat{p}_{ij}^3\}$ . Then draw a knowledge level  $\hat{k}(i, j)$  randomly from levels 1, 2 or 3 with probabilities  $\{\hat{p}_{ij}^1, \hat{p}_{ij}^2, \hat{p}_{ij}^3\}$ .

---

<sup>4</sup>This section is based on the explanation structure of Barahona et al. (2021).

<sup>5</sup>Home location for applicants with unreliable geocoding are imputed with the method detailed in Section D.3 of this appendix.

3. Simulate the noise component  $\eta_{\hat{k}(i,j)} \sim N(0, \sigma_{\hat{k}(i,j)})$
  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  $\hat{EU}_{ij} = \hat{V}_{ij} + \hat{\eta}_{\hat{k}(i,j)} - \hat{\rho}_{\hat{k}(i,j)}^{s1} + \hat{\epsilon}_{ij}$
- 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  $\hat{EU}_{iz(i)}^{s2} = \lambda(\hat{V}_{iz(i)} + \hat{\tau} \times \hat{\eta}_{k(i,z(i))} - \hat{\rho}_{\hat{k}(i,j)}^{s2} + \hat{\epsilon}_{iz(i)}) + \hat{\xi}_{iz(i)}$
    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

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

### **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.<sup>6</sup>
  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 ( $z(i)$ ). In practice, we draw the unobserved portion of the expected utility of the enrolled schools constrained to the utility inequality ( $\epsilon_{iq(i)} st. EU_{iq(i)}^{s1} > EU_{iz(i)}^{s1}$ ). 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 recommendation, we calculate the observed utility based on model's estimated 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  $\alpha$  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,  $\beta$ , to quantify this variation. At

---

<sup>6</sup>All simulated non-complier applicants have the same geolocation of at least one real applicant, but not necessarily a non-complier.

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 ( $EU_{iq(i)}^{s1}$ ). 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<sup>7</sup>, 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.

---

<sup>7</sup>We consider at most the first 3 schools in the ranking.

## E. SIMULATED MAXIMUM LOG-LIKELIHOOD

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

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

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:

$$\begin{aligned} \beta_i^\sigma &= \phi_i^\beta \cdot \sigma^\beta & \phi_i^\beta &\sim N(0, I_{|\beta^\sigma|}) \\ \eta_1 &= \phi_i^\eta \sigma_1^\eta & \phi_i^\eta &\sim N(0, 1) \\ \eta_2 &= \phi_i^\eta \sigma_2^\eta & \phi_i^\eta &\sim N(0, 1) \end{aligned}$$

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

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

Then, we search for the vector  $\hat{\theta}$  that maximizes the sum of the logarithm of  $\hat{L}_i$ :<sup>8</sup>

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

The vector  $\hat{\theta}$  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-

---

<sup>8</sup>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. Two-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 ( $\beta^\sigma$  and  $\eta$ ) 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  $[\beta^\sigma, \eta]$ , and then recover draws from the distribution of  $\epsilon_{ij}$ , imposing a “coherence constraint” between  $\epsilon_{ij}$  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 ( $\beta^\sigma$ ) and the noise term ( $\eta$ ). 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:

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

- With the maximum likelihood estimates of the rank choice parameters in hand, for each applicant  $i \in \{1 \dots I\}$  and simulation  $s \in \{1 \dots S\}$  (i.e. conditional on  $\beta_{is}^\sigma$  and  $\eta_{ijs}$ ):
  1. We predict the observed part of the expected utility  $V_{ijs} + \eta_{k(i,j)s} - \rho_{k(i,j)}^{s1}$  using the estimated parameters and the generated draws ( $\beta_s^\sigma$  and  $\eta_s$ ) for every school in the ranking.
  2. We generate  $T$  approximate draws from  $F(\epsilon_{ij} | \beta_s^\sigma, \eta_s)$ , 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).<sup>9</sup>

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

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  $\beta^\sigma$  and  $\eta$ , and the approximate draws of  $\hat{\epsilon}_{iz(i)}$  generated on the step 1.

---

<sup>9</sup>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  $\hat{\epsilon}_{is}$ . The most coherent is defined as the vector which minimizes the "incoherent distance" between adjacent ranked schools, defined as  $\sum_{r=1}^{|C_i|-1} \max\{0, EU_{r+1} - EU_r\}$  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:

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

Assuming that the support of  $EU_{is}^{s1}$  is positive<sup>10</sup>, 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:

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

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.

---

<sup>10</sup>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 ordered 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:

$$K_{ij} = \alpha_1 \times \text{distance}_{ij} + \alpha_2 \times \text{distance}_{ij}^2 + \beta \times \text{connection}_{ij} + \phi \times \text{rank}_{ij} + \delta_{f(i)j} + \epsilon_{ij}$$

$\text{distance}_{ij}$  is the euclidean distance between home of applicant  $i$  and school  $j$ .  $\text{connection}_{ij}$  is a vector that includes dummies representing a familiar connection with the school (a currently enrolled sibling, employed parent, or alumni).  $\text{rank}_{ij}$  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  $\text{female}_i$  and  $\text{LowSES}_i$  ( $2 \times 2$ ).  $\epsilon_{ij}$  is an unobserved (to us) portion of  $K_{ij}$  and is assumed *IID Logistic*(0,1).

We will assume that there are thresholds  $\kappa_1$  and  $\kappa_2$  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  $\epsilon_{ij} \sim \text{Logistic}(0, 1)$ , the probability  $P(k(i, j) = a|\theta)$  has a simple analytical form once we condition on the vector  $\theta$ , that contains the parameters that define  $K_{ij}$  and the thresholds.<sup>11</sup> The log-likelihood function of observing responses  $a_{ij}$  for each school  $j \in \mathcal{C}_i$ <sup>12</sup> and survey respondent  $i \in \{1 \dots I\}$  is the following:

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

The estimate of the vector of parameters  $\theta = [\alpha, \beta, \phi, \delta, \kappa]$  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  $\{\theta_z\}_{z \in \{1 \dots 70\}}$ .

### 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*<sup>2</sup> 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.

---

<sup>11</sup>See Train (2009) for details.

<sup>12</sup>In the surveys 2020 and 2021, we asked for at most five schools; in 2022, at most seven.

## H. SURVEY TRANSLATION

Figure B.VI  
2020 Survey Landing Page



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](http://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 placed 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) Standardized 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 Step-mother) 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*]