Title

Essays on Empirical Market Design

Author

Arteaga Ossa, Felipe

Publication Date

2024

Essays on Empirical Market Design

by

Felipe Arteaga Ossa

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Benjamin Handel, Chair
Professor Nano Barahona
Professor Jonathan Kolstad

Spring 2024

Essays on Empirical Market Design

Abstract

Essays on Empirical Market Design

by

Felipe Arteaga Ossa

Doctor of Philosophy in Economics

University of California, Berkeley

Professor Benjamin Handel, Chair

This dissertation empirically explores different aspects of the market design in centralized school choice.[1] The first chapter studies the consequences of limited information among school choice participants, and the influence of outside options in application and enrollment decisions. 23% of the Chilean applicants who receive an offer choose to enroll elsewhere, unnecessarily blocking seats that would improve the allocation for 12% of the placed applicants and offer placement to 11% of the non-placed students. Based on a model of the joint decision of school choice and enrollment, I 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 outside options diminishes the incentive for search and lowers the cost of rejecting placement offers. The counterfactual analysis highlights the effect of different information campaigns and the inclusion of outside options in the centralized system, underscoring the importance of after-market design in centralized school choice systems. The second chapter shows that beliefs about admissions chances shape choice outcomes even when the school choice assignment mechanism is strategyproof. Data from a large-scale survey of choice participants in Chile shows that learning about schools is hard, that beliefs about admissions chances guide the decision to stop searching, and that applicants systematically underestimate non-placement risk. We then use RCT and RD research designs to evaluate scaled live feedback policies. 22% of applicants submitting applications where risks of non-placement are high respond to warnings by adding schools to their lists, reducing non-placement risk by 58%. The third chapter evaluates how new information influences families' applications and assignment outcomes in elementary school choice settings. Specifically, using a multi-country RCT based in Tacna, Peru and Manta, Ecuador, we examine the effect of providing personalized information on schooling alternatives and placement risk. We find that applicants who received feedback on placement risk and a suggestion of new schools added more schools to their applications and were more likely to include recommended schools than other alternatives.

[1]The 2nd chapter was written jointly with A. Kapor, C. Neilson, and S. Zimmerman (Arteaga et al., 2022), and the 3rd chapter with G. Elacqua, T. Krussig, C. Méndez, and C. Neilson (Arteaga et al., 2022).

Acknowledgments

I would like to express my deepest gratitude to my advisors, Nano Barahona, Ben Handel, and Chris Neilson. Their constant support, guidance, and invaluable feedback throughout all stages of my research projects have been instrumental in the completion of this work. I am also indebted to Matt Backus, Fede Echeñique, Jon Kolstad, and Chris Walters for their insightful advice and assistance whenever I needed it.

I extend my heartfelt thanks to my classmates, coauthors, and friends who have been there for me in various ways during these years. I am grateful to Pipe V. and Tati for their warm welcome to Berkeley and for guiding me through the different stages of the program. I appreciate my office mates Cris, Dami, Harry, and Pat for creating a supportive environment where we could help each other navigate our challenges.

I am also thankful to my homemates, Andrea, Anton, Matteo, and Miguel, for making Berkeley and Oakland places where I can feel at home. I will miss our pasta dinners, movie nights, Costco trips, home parties, and veggie burger BBQs. Having good friends to talk about our struggles, achievements, and changing emotions throughout the grad school roller coaster was a true blessing. As we leave our shared home, I am grateful to carry with me a little bit of Andrea's independent-mindedness and austerity, Anton's curiosity and sharpness, Matteo's insightfulness and perseverance, and Miguel's passion and genuineness.

I treasure the memories of the many tennis games and adventures I shared with friends while climbing, skiing, running, and surfing, taking advantage of the unparalleled natural amenities of the West Coast and providing a much-needed balance to the PhD life. I'm convinced that Berkeley is the best place in the world to pursue a graduate degree.

I gratefully acknowledge the financial support provided by Becas Chile and the Berkeley Economics Department, which made my studies possible.

Finally, I am deeply grateful to Vale for her boundless love, for what she teaches me with her unique perspective on life, and for being a wonderful companion in our adventures. Meeting her in the final stage of this journey was one of the greatest surprises of my life in the United States. Lastly, I will forever be grateful to my parents, Carmen Luz and Felipe, and my siblings, Maida, Consuelo, Esteban, Andrés, and Samuel, for their unconditional love and for being the ultimate driving force behind all my achievements after 35 years of affection and guidance. This dissertation is dedicated to all of you.

Contents

Contents ii

List of Figures v

List of Tables vii

1 Imperfect Information and Outside Options in Centralized School Choice 1

1.1 Introduction . . . 1
1.2 Setting . . . 5
1.3 Survey . . . 10
1.4 Model . . . 13
1.5 Bringing the Model to the Data . . . 18
1.6 Results . . . 24
1.7 Counterfactuals . . . 29
1.8 Conclusions . . . 35

2 Smart Matching Platforms and Heterogeneous Beliefs in Centralized School Choice 37

2.1 Introduction[2] . . . 37
2.2 Searching for Schools . . . 42
2.3 Setting . . . 47
2.4 Survey Findings . . . 53
2.5 Warnings, Choice Behavior, and Choice Outcomes . . . 56
2.6 Why Do Smart Platforms Work? . . . 62
2.7 Smart Matching Platforms in New Haven . . . 65
2.8 Conclusion . . . 67

[2]This chapter was published as a journal article in 2022 (Arteaga et al., 2022), and was coauthored with Adam Kapor, Christopher A. Neilson, and Seth Zimmerman.

3 Can Information on School Attributes and Placement Probabilities Direct Search and Choice? 82
3.1 Introduction[3] 82
3.2 Setting 84
3.3 Survey 88
3.4 Intervention design 91
3.5 Choice behavior and choice outcomes 95
3.6 Conclusions 99

Bibliography 108

A Appendix of Imperfect Information and Outside Options in Centralized School Choice 116
A.1 Additional Figures 116
A.2 Additional Tables 121
A.3 Simulated choice sets 127
A.4 Counterfactual simulations 130
A.5 Simulated maximum log-likelihood 134
A.6 Additional Proofs 137
A.7 Knowledge and consideration prediction functions 138
A.8 Survey Translation 140

B Appendix of Smart Matching Platforms and Heterogeneous Beliefs in Centralized School Choice 147
B.1 Model Appendix 147
B.2 Additional Figures and Tables 152
B.3 Chilean Admission Policy and Interventions 174
B.4 Datasets 191
B.5 Treatment First Stages 193
B.6 Risk Prediction 197
B.7 2020 Survey Translation 200
B.8 Additional School Quality Results 206
B.9 RDs at Multiple Cutoffs from the 2017 Pilot 210
B.10 2020 WhatsApp Treatment Arms 211
B.11 New Haven Warnings Policies and Data 212
References 223

[3]This chapter was published as a working paper in 2022 (Arteaga et al., 2022), and was coauthored with Gregory Elacqua, Thomas Krussig, Carolina Méndez, and Christopher A. Neilson.

C Appendix of Can Information on School Attributes and Placement Probabilities Direct Search and Choice? **223**
C.1 Additional Figures 223
C.2 Additional Tables 226
C.3 Outreach and Treatment Details 230
C.4 Survey Details 233

Bibliography **247**

List of Figures

1.1 Timeline of the application process 6
1.2 Compliance to placement offers and final enrollment 9
1.3 Knowledge level about schooling option 11
1.4 Correlation of enrollment and knowledge about the placed school 12
1.5 Steps to get to know a school and satisfaction with placement 13

2.1 Distribution of Placement Probabilities and Probability Predictions 68
2.2 Knowledge of and Search for Schools 69
2.3 Reasons for Stopping School Search 70
2.4 Subjective vs. Observed Application Risk 71
2.6 WhatsApp RCT Outcomes 72
2.7 2021 WhatsApp RCT with Personalized and Aggregate Information Treatments 73
2.8 Smart Warnings and Encouragement Nudges in New Haven 74
2.5 Choice Behavior and Enrollment Outcomes in the Platform Pop-Up RD 75

3.1 Main survey evidence 101
3.2 New Survey Evidence 102
3.3 Report card extracts 103
3.4 Schools added RD 103

A.1 2022 Online Application Platform Screenshots 117
A.2 Enrollment decision for non-compliers 118
A.3 Screenshot of question about knowledge level of schools 119
A.4 Knowledge level about schooling option by mother's education 120
A.5 Probability of knowledge by β 120
A.6 2020 Survey Landing Page 140

B.1 Platform Pop-Up Intervention– 2018 and 2019 159
B.2 Timeline of Feedback Interventions– 2020 160
B.3 Satisfaction with Placement Outcomes 161
B.4 Alternate Application Risk Framings 162
B.5 Balance in Platform Pop-Up RD 163
B.6 Additional Platform Pop-Up RD Figures 164

B.7 Multiple Bandwidths RD Plots of Platform-Based Pop-Up Warning Effects (Outcomes in Table 2.2) 165
B.8 Multiple Bandwidths RD plots of Platform-Based Pop-Up Warning Effects (Outcomes in Table 2.2) 166
B.9 Multiple Bandwidths RD Plots of Platform-Based Pop-Up Warning Effects (Outcomes in Table 2.3) 167
B.10 Multiple Bandwidths RD Plots of Platform-Based Pop-Up Warning Effects (Outcomes in Table 2.5) 168
B.11 Platform Pop-Up Effects over City-Years and by Market size 169
B.12 Intervention Effects on Beliefs and Preferences 170
B.13 Balance in 2021 WhatsApp RCT 171
B.14 Enrollment in Placed Rates Conditional on Placement Ranking 172
B.15 Demand for Information 173
B.16 Policy Rollout 174
B.17 Application Platform Screenshots 176
B.18 Enrollment Outcomes for Unplaced Students 179
B.19 SMSs Intervention Warning Texts – 2016 185
B.20 SMSs Intervention Warning Texts – 2017 186
B.21 Platform Pop-Ups 187
B.22 WhatsApp Intervention Warning Images – 2020 188
B.23 WhatsApp Intervention Warning Images – 2021 189
B.24 Eligibility for 2021 WhatsApp RCT 190
B.25 RD First Stages of 2018 and 2019 Interventions 195
B.26 RD First Stages in 2020 Platform Popup Intervention 196
B.27 First Stages of 2020 WhatsApp Intervention 197
B.28 2020 Survey Landing Page 200
B.29 Distributions of School Value Added and Spending by Oversubscription Status . 209
B.30 Email Communication with Parents 219
B.31 New Haven Simulator 220
B.32 Observed vs. Predicted Placement Probability in New Haven 221
B.33 Treatment Balance and First Stage Outcomes 222

C.1 Schooling options in a 2km radius from the city center 223
C.2 Knowledge About Options in Application 224
C.3 Necessary Steps for Learning about a School 224
C.4 Report card example of T_2 for Peru 2022 225
C.5 Information on the Number of Schools Nearby (Ecuador/Peru) 226
C.6 Treatments for Classifying Schools (Ecuador) 226

List of Tables

1.1 Descriptive Statistics for Choice Applicants . . . 8
1.2 School Choice Estimates . . . 26
1.3 Noise standard deviation and uncertainty penalization estimates . . . 27
1.4 Outside option estimates . . . 28
1.5 Counterfactual Results for All Applicants . . . 33

2.1 Descriptive Statistics for Chilean Choice Applicants . . . 76
2.2 RD Estimates of Platform Pop-Up Effects . . . 77
2.3 RD Estimates of Platform Pop-Up Effects on Enrolled School Outcomes . . . 78
2.4 WhatsApp RD and RCT Results – 2020 . . . 79
2.5 RD Estimates of Platform Pop-Up Effects on Subjective Beliefs . . . 80
2.6 RCT Estimates of Behavioral Nudge Effects . . . 81

3.1 Descriptive statistics for applicants . . . 104
3.2 Take-up of WhatsApp Messages and Report Card . . . 105
3.3 RCT Results: Effect of Suggestions on Application Outcomes. . . . 106
3.4 RCT Results: Effect of Suggestions on Application Outcomes by implmentation country/year. . . . 107

A.1 Composition of in-system and out-of-system schools . . . 121
A.2 Descriptive Statistics for Schools . . . 122
A.3 Counterfactual Results for Non-placed Applicants . . . 123
A.4 Counterfactual Results for Complier Placed in 2nd+ Preference . . . 124
A.5 Counterfactual Results for Complier Placed in 1st Preference . . . 125
A.6 Counterfactual Results for Non-complier . . . 126

B.1 More Descriptive Statistics for Chilean Choice Applicants . . . 152
B.2 Descriptive Statistics for Chilean Choice Applicants- Alternate Samples . . . 153
B.3 Platform Pop-Up RD Estimates of Main Outcomes (Table 2.2) with Alternate Bandwidths . . . 154
B.4 RD Estimates of Platform Pop-Up Effects on Adding Any School, by City and Year . . . 155
B.5 RD Estimates of Platform Pop-Up Effects by Market-Level Choice Experience . 156

B.6 RD Estimates of Platform Pop-Up Effects by Applicant's Socioeconomic Status 157
B.7 RD Estimates of Platform Pop-Up Effects on Enrolled School Outcomes by Applicant's Socioeconomic Status . . . 158
B.8 Aggregate Placement Results for Stable Populations by Year . . . 178
B.9 WhatsApp RD and RCT Results – 2021 . . . 184
B.10 RD Estimates of Platform Pop-Up Effects . . . 208
B.11 RD Estimates of Platform Pop-Up Effects - 2017 . . . 210
B.12 RCT Estimates Split by WhatsApp Message . . . 211
B.13 Sample Descriptives New Haven . . . 216
B.14 RD and DD Estimates of Warnings Effects in New Haven . . . 217
B.15 Treatment Balance RCT . . . 218

C.1 Summary of Information Interventions. . . . 227
C.2 RCT Results: Differences Between Treatment 2 and 3. . . . 228
C.3 RCT Results: Effect of Suggestions on Subjective Outcomes. . . . 229
C.4 Example of WhatsApp Conversation from Peru 2022 - Translation to English . . 231
C.5 Example of WhatsApp Conversation from Peru 2022 - Original Spanish 232

Chapter 1

Imperfect Information and Outside Options in Centralized School Choice

1.1 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; Abdulkadiroglu 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 1.2 provides an overview of the Chilean school admission system, setting the context for our research. In Section 1.3, we present an in-depth analysis of survey results, highlighting key insights into families' knowledge and decision-making processes. Section 1.4 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 1.5, followed by the presentation of estimation results in Section 1.6. Building upon these findings and model estimates, Section 1.7 explores a range of counterfactual scenarios, shedding light on the potential impact of policy interventions. Finally, Section 1.8 concludes the paper by summarizing our key findings and discussing their implications for policy and future research.

1.2 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.

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.

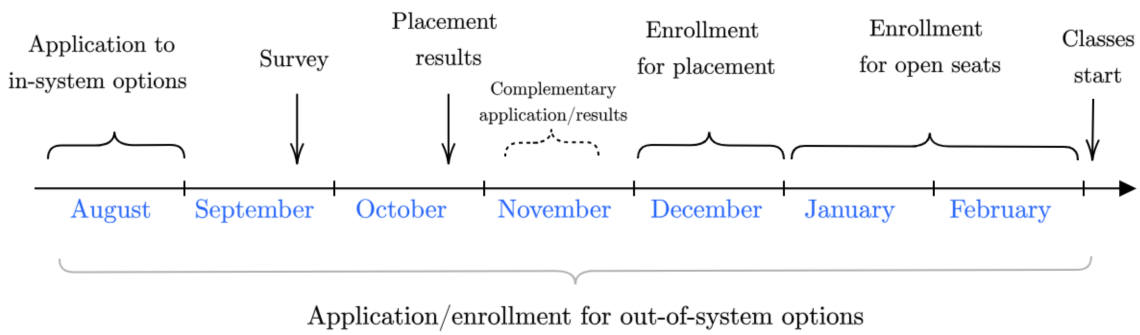
The centralized system regulates admission for most public and private-voucher schools, which we refer to as "in-system" schools. In 2022, these schools represented 87.2% of the PK-12th enrollment (a total of 3.4 million students). However, among publicly funded schools (i.e., publicly owned or private 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.[1]

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.[2]

[1] 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 A.1 in Appendix A.2 for a summary of the classification of schools and their share of the total enrollment.

[2] 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.

Figure 1.1: Timeline of the application process



The centralized application process for the "in-system" schools typically takes place during August, as shown in Figure 1.1. The government provides an online platform that not only registers applications but also offers information about the process and each school.[3] All participating 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.[4] 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).[5]

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

[3]Figure A.1 in Appendix A.1 shows screenshots of the application website from 2020.

[4]Among those placed in the main round who did not enroll in their assigned school, only 14% participated in the complementary round.

[5]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.

is a quota for low-SES students and, in a few cases, a quota for high-performing students in high-school grades.[6] 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.[7] 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.[8] 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.[9]

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 1.1 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.[10] 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 1.2a illustrates that compliance rates decrease sharply with the placement ranking, with 84% of applicants placed in their 1st preference enrolling in the offered school,

[6]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.

[7]There is a third option to remain on the waitlist for higher preferences, but this option is not commonly used.

[8]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.

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

[10]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 1.1: Descriptive Statistics for Choice Applicants

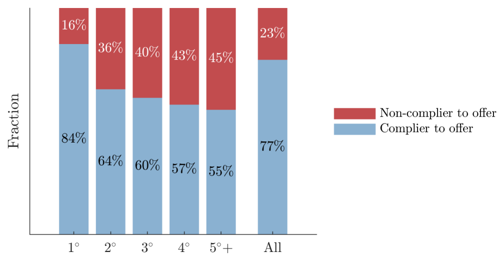
	(1) All	(2) Survey respondent	(3) Estimation sample
A. Applicants demographics			
Female	0.50	0.50	0.50
Low SES	0.54	0.48	0.42
Voluntary applicant	0.35	0.31	0.25
Santiago (main urban zone)	0.35	0.39	0.50
Rural	0.05	0.04	0.00
PK-K	0.32	0.36	0.40
1st-6th	0.29	0.27	0.27
7th-12th	0.39	0.37	0.34
2020	0.31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

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

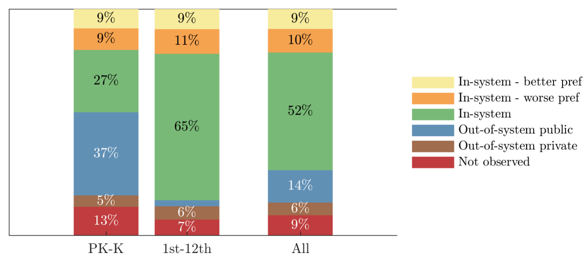
while only 55% of participants assigned to their 5th preference or lower comply with their placement.

Figure 1.2: Compliance to placement offers and final enrollment

(a) Compliance decision by placement ranking



(b) Where do non-compliers enroll?



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 of 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(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 1.2b shows the fraction of these students who enrolled in in-system and out-of-system schools, as defined at the beginning 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.[11] 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

[11]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.

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 A.2 in Appendix A.1 provides more details on these heterogeneous behaviors.

1.3 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.[12] 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 1.1.[13] 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 1.1).

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 1.3a shows the responses to the question "How well do you know the schools in your application?".[14] 30% of the families indicated that they do not know their first-

[12]See Appendix A.8 for a translated version of the survey questions.

[13]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.

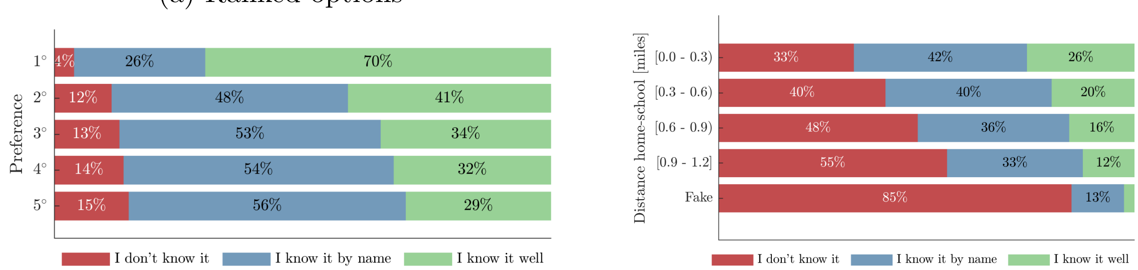
[14]A screenshot of the question from the implemented survey is shown in Figure A.3 in Appendix A.1

ranked school well, with 26% stating "I know it by name" and 4% stating "I don't know it."[15] 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 A.4a and A.4b in Appendix A.1).

Figure 1.3: Knowledge level about schooling option

(a) Ranked options

(b) Non-ranked nearby options



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 1.3b 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 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

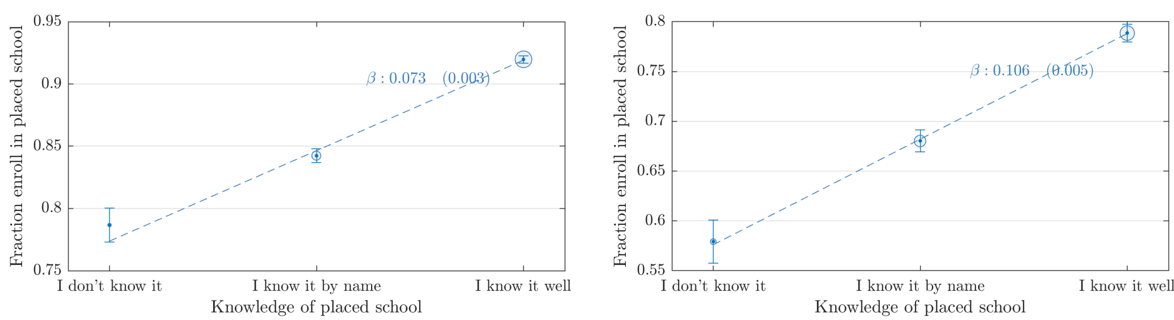
[15]Since we asked about schools they included in their ranking and some respondents said "I don't know it," we will not interpret the answers literally. Instead, we consider "I don't know it," "I know it by name" and "I know it well" as three ordinal levels of knowledge.

considered necessary to get to know a school well. Participants were allowed to choose multiple items. Figure 1.5a 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 1.4a and 1.4b 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 1.4: Correlation of enrollment and knowledge about the placed school

(a) Placed in 1st preference (b) Placed in last preference



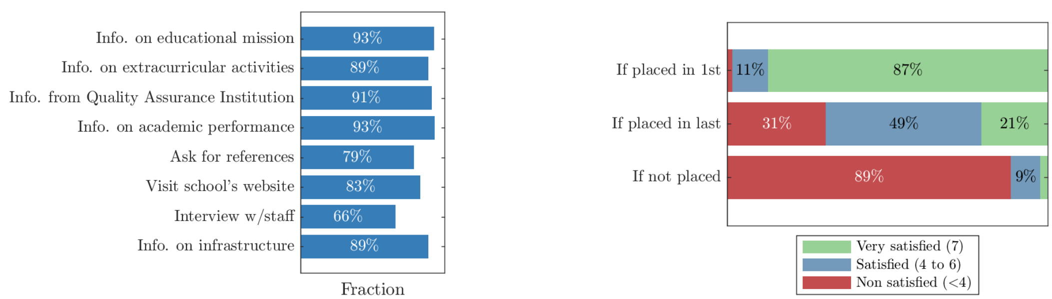
Notes: These plots show the fraction of applicants that enroll in the placement offer conditional on their answer of knowledge about the schools. Panel (a) shows for students placed on 1st preference, and panel (b) placed on last preference. Means are computed after controlling for the probability of 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 1.5b 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 1.5: Steps to get to know a school and satisfaction with placement

(a) Relevant steps to get to know a school well

(b) Satisfaction with hypothetical placements



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.

1.4 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.

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.[16] In **stage 2**, each applicant receives a unique placement[17] 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 at school j for family i is U_{ij}. However, the perceived utility is composed of the true utility plus a noise term: $U^{p1}ij = Uij + \eta_{ij}$. 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:

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

[16]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.

[17]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.

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 ($\mathbb{E}[Uij^{b1}] = U_{ij}^{p1}$), since the noise term η_{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:

$$\begin{aligned}
\mathbb{E}[U_{ij}^{b1}] &= \mathbb{E}[-\frac{1}{r}\exp(-r(U_{ij}^{p1} - \eta_{ij})] \\
&= \mathbb{E}[-\frac{1}{r}\exp(-rU_{ij}^{p1})\exp(r\eta_{ij})] \\
&= -\frac{1}{r}\exp(-rU_{ij}^{p1})\mathbb{E}[\underbrace{\exp(r\eta_{ij})}_{\sim LogN(0,(r\sigma_\eta)^2)}] \\
&= -\frac{1}{r}\exp(-rU_{ij}^{p1})\exp\left(\frac{(r\sigma_{\eta_{ij}})^2}{2}\right)
\end{aligned} \tag{1.1}$$

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 1.1 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 ($\frac{\partial EU_{ij}^{s1}}{\partial \sigma_{\eta_{ij}}^2} < 0$).[18]

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*

[18] 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

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:

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

where ξ_{ij} and ξ_{i0} are future preference shocks, known to the families only in stage 2. These shocks are assumed to be $\overset{\text{iid}}{\sim} EVI$ and uncorrelated with ϵ_{ij}. U_{i0} represents the observed utility of the outside option, which depends on the geographic supply of schools in the aftermarket.[19] The scale factor λ multiplying EU^{s1}_{ij} and U_{i0} allows the unobserved part of U^{p1}_{ij} (not yet introduced) and ξ_{ij} to have different variances.

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

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

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 Ω_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 pij 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:[20]

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

This expression requires relabeling schools for each applicant such that $w_{i1} > w_{i2} > w_{i3} > \ldots > w_{iN}$, so school 1 is the most preferred school in the choice set, but not necessarily the same school for all applicants. $R_{ij} \equiv 1 - p_{ij}$ represents the probability of not being placed

undesirable feature. However, our framework is immune to this critique since ϵ is an additive component of U^{p}_{ij}, which is embedded into the Bernoulli utility function, as detailed in Section 1.5. Subsequent algebraic manipulation of the expected utility generates a convenient additive unobserved utility component.

[19]We provide a detailed explanation of how we model the utility of the outside option in Section 1.5.

[20]Ordering schools from highest to lowest wij is the result of maximizing $\mathcal{V}(\mathcal{C}_i)$ conditional on the choice set Ω_i and knowing the DA rules. Any other order would shift placement probability from a preferred school to a less preferred school.

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 κ_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,[21] the expected value of a new search iteration is given by:

$$\mathbb{E}[(w_{is} - EU_{i0})p_{is} \prod_{j\leq N} R_{ij}] - \kappa_i > 0$$

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

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} ($F_i(EU_{is}^{s1}, p_{is})$),[22] (2) the expected utility of the outside option, (3) the subjective belief about being placed in one of the already included schools ($\prod_{j\leq N} R_{ij}$), and (4) the search cost (κ_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.[23] Third, families that believe they are likely to be placed in one of the schools in $\mathcal{C}i$ ($\prod j \leq NR_{ij} \approx 0$) 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}.[24] 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.

[21] We extend the analysis for cases where families add a school to a position other than the last in Appendix A.6. 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.

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

[23] The proof and details of the effect of the valuation of the outside option on search are provided in Appendix A.6.

[24] This assumes that the search cost κ_i is the cost of becoming familiar with a school, but not necessarily being fully informed about it.

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 τ_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 ξ_{i0} and decide to attend the offered school if $EUiz(i)^{s2} > \lambda Ui0 + \xi_{i0}$, 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.

1.5 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, and survey data.[25]

In our model, families determine their ROLs by comparing the enrollment option utility w_{ij} among the schools in their choice set Ω_i. Consequently, ROLs provide multiple pseudo-choices per applicant of the form $w_{ir} > w_{ij}$, $j \in \Omega_i \backslash \{1 \dots r\} \, \forall r \in \mathcal{C}_i$ (Train, 2009). 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^{s1} instead of w: $EU_{ir}^{s1} > EU_{ij}^{s1}$, $j \in \Omega_i \backslash \{1 \dots r\} \, \forall r \in \mathcal{C}_i$.[26]

As detailed in Section 1.4, 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 η. 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

[25] 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.

[26] The intuition behind the effectiveness of the transformation $g_i()$ lies in the fact that the outside option for each choice j is the same for applicant i.

between the observed component V_{ij} and the unobserved (to us) part of the utility $\epsilon_{ij} \overset{\text{iid}}{\sim} EVI$. 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):

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

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, $EU_{iz(i)}^{s2}$,[27] 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 Ω_i. We conclude the section by presenting the likelihood function that we maximize to estimate our model parameters.

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,[28] 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

[27] 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 λ to allow $\epsilon_{iz(i)}$ and $\xi_{iz(i)}$ to have different variances.

[28] 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 A.2 in Appendix A.2.

i's attributes, which includes dummies for low-SES, female, voluntary change, application year, and school level.[29]

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

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

Noise term (η_{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 η_{ij} in stage 1 is as follows:

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

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 ρ_1^{s1} and ρ_2^{s1}, which represent the uncertainty-penalization term:

[29]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.

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

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

Choice Sets (Ω_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 Ω_i, but only the subset they applied to ($\mathcal{C}_i \subset \Omega_i$). A common approach in the school choice literature is to assume that choice sets are composed of all schools available within a limited geographic zone where applicants 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).[30] 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).[31] 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.[32] 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 A.3.[33]

---

[30] 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)."

[31] 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.

[32] 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.

[33] 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.

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.[34] 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.[35] 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).

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:[36]

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

[34]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.

[35]We define an applicant as voluntary if they are currently enrolled in a school that offers the next grade.

[36]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 $EU^{s2}_{iz(i)} < U_{i0} + \xi_{i0}$. For applicants who are not placed ($z(i) = 0$), the likelihood is simply $P(\mathcal{C}_i)$.

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

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

The log-likelihood function is defined as $ll = \sum_{i=1}^{I} \log(L_i)$. We estimate the parameters via simulated maximum likelihood (Train, 2009), following the standard procedure except for the generation of draws for $\epsilon_{iz(i)}$. Given that the Deferred Acceptance algorithm places each student in 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.[37] This is because schools with higher unobserved utility ϵ_{ij} are more likely to be ranked at the top of the list. To generate approximate draws from the distribution of $\epsilon_{iz(i)} | \boldsymbol{\beta}^{\sigma}, \boldsymbol{\eta}$, we follow a two-step procedure that is described in detail in Appendix A.5.

1.6 Results

We now turn to describe the estimated parameters of the model detailed in Section 1.4. We use the Simulated Maximum Likelihood procedure over the function defined at the end of Section 1.5. 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.

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 1.2. Following Abdulkadiroğlu et al. (2017), all estimates are relative to the effect of 1 mile for the $\boldsymbol{X}_i = \boldsymbol{0}$ student (male, mid-high-SES,

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

non-voluntary, applying to elementary school in 2020), so they can be interpreted as the willingness to travel. For example, the $\boldsymbol{X}_i = \mathbf{0}$ student values having a sibling in the school as much as having the school 7.39 miles closer. The model in Abdulkadiroğlu et al. (2017) assumes a common distaste for distance, while we allow for observed preference heterogeneity, which makes interpretation less direct. As an example, in our case, an extra mile for the $\boldsymbol{X}_i = \mathbf{0}$ student affects the valuation of a school $-1/(-1+0.57) = 2.3$ times more than for his equivalent who is applying to high school (HighSch = 1).[38]

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.[39] 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 1.2. 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.

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

[39] 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 1.2: School Choice Estimates

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

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

Noise, uncertainty penalization and learning

Table 1.3 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).

Table 1.3: Noise standard deviation and uncertainty penalization estimates

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

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

Families' beliefs about the noise variance are coherent with our estimates of the actual variance of the noise. Recall that $\rho^{s1}_1 = \frac{\sigma^2_{\eta 1}}{2}$ 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 τ is 0.71, suggesting that the dispersion of the noise is reduced by 30% from stage 1 to stage 2, which we interpret as learning. At the same time, we also observe that the penalization term, if any, increases from 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 σ_η do not update at the same rate as the shrinkage of the true variance of η. 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 σ_η are correct.

Valuation of the Outside Option

The compliance model compares the updated expected utility of enrolling in the assigned school $EU^{s2}_{iz(i)}$ 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 τ, and preference shocks, represented by $\xi_{iz(i)}$.

Table 1.4: Outside option estimates

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

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

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 1.4 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.

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

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 (ϵ_{ij}). With those inputs, we calculate the expected utility (EU_{ij}^{s1}) of each school and build the ROLs ($\mathcal{C}_i$).

After obtaining the ROLs, we proceed to run the Deferred Acceptance algorithm using the real school capacities and recover the placements $z(i)$ of the students. We then construct the utility of the outside option U_{i0}, simulate the preference shocks realized in stage 2 ($\xi_{iz(i)}$ and ξ_{i0}), and generate compliance decisions ($EV_{iz(i)} > U_{i0} + \xi_{i0}$?).[40]

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.[41]

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.

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 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.

[40]We run the entire process 100 times. We provide details of the construction of the indirect utility and overall simulation in Appendix A.4. Details of the function that predicts the level of knowledge can be found in Appendix A.7.

[41]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.

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: $\epsilon_{iq(i)}$ s.t. $EU^{s2}_{iq(i)} > \lambda U_{i0} + \xi_{i0}$. 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, β, 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 ($EU^{s1}_{iq(i)}$). Conversely, when $\beta = 0$, families have the lowest level of information in our model ($k(i, q(i)) = 1$). Any value of β between 0 and 1 represents a probability distribution over the different levels of information. As β approaches 0, the probability of $k(i, q(i)) = 1$ increases. When β is close to 0.5, the probability of $k(i, q(i)) = 2$ is highest, and as β approaches 1, the probability of $k(i, q(i)) = 3$ becomes dominant. Figure A.5 in Appendix A.1 provides a precise mapping of probabilities for different values of β.

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.

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 1.5: Counterfactual Results for All Applicants

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

Notes. This table shows the changes in placement (columns 1 to 3) and 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 1.5. 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 A.5 to A.6 in Appendix A.2.

Examining Panel A of Table 1.5, 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 A.4) and 11% (Table A.3), respectively. Interestingly, 9% of non-compliers (Table A.6) 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 1.5. 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.[42]

Panel D of Table 1.5 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%.

1.8 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 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%.

[42] 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.

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.

Chapter 2

Smart Matching Platforms and Heterogeneous Beliefs in Centralized School Choice

2.1 Introduction[1]

Many school systems around the world use centralized mechanisms to assign students to schools. An important contribution economists have made to the design of centralized school choice is to guide policymakers towards centralized mechanisms that are strategically simple for participants to use. In cities including New York, Boston, New Haven, and Santiago, economists have helped design "strategyproof" choice systems where applicants' dominant strategy is to list schools they like, in the order that they like them (Abdulkadiroğlu et al., 2005,?; Correa et al., 2019; Akbarpour et al., 2020). A central point in the case for strategyproof approaches is that knowledge of admissions chances— which may be costly to acquire or unequally distributed— is not required for optimal play.

The conclusion that strategyproof centralized mechanisms relieve choice participants of the need to know about their admissions chances follows from the maintained assumptions of the canonical "school choice problem" (Abdulkadiroğlu and Sönmez, 2003) that applicants know which schools are available to them and which they like. But what if learning about schools is costly, and families do not know about all of their options? This paper examines how costly search interacts with beliefs about admissions chances to shape what families know about schools, how much this matters for choice outcomes, and what policymakers can do about it. We take an empirical approach, drawing on large-scale surveys and policy variation in the Chilean and New Haven school choice systems.

We make two main points. The first point is that costly search for schools is central to

[1]This chapter was published as a journal article in 2022 (Arteaga et al., 2022), and was coauthored with Adam Kapor, Christopher A. Neilson, and Seth Zimmerman.

the way families experience choice, and that this places beliefs about assignment chances back in a key role even when the assignment mechanism is strategyproof. Many participants stop searching for schools because they think they will be admitted to a school already on their application. Systematic over-optimism about admissions chances leads participants to submit applications with high non-placement risk.

The second point is that a new kind of intervention, which we call a "smart matching platform," can help families navigate costly search more effectively. Smart matching platforms aggregate data on the distribution of choice applications to provide live feedback on admissions chances to platform users. We use experimental and quasi-experimental research designs to test smart platforms at scale, and find that they change application behavior, raise placement rates, and cause students to enroll in higher-quality schools. We conclude that reducing the burden of choice requires not just strategyproofness inside the centralized system, but also support during the search process that precedes it.

We begin by developing a simple model of school search in a strategyproof assignment mechanism. We draw on models of job search such as McCall (1970), with the key difference being that individuals add schools they find to an application portfolio, rather than making one-time decisions to accept or decline an offer. Applicants engage in costly, sequential search for schools to add to their choice application. Once applicants decide to stop searching, they submit the application to a strategyproof assignment mechanism.

The key insight of the model is that the value of search depends on how likely the applicant thinks she is to be placed in the school she finds. We use the model to derive two main results. First, over-optimism about admissions chances can reduce search and increase the risk of non-placement. Second, information interventions implemented after individuals have decided to stop search weakly raise the probability that applicants search for and find additional schools to add to their applications. Applicants who respond to the intervention by adding schools to their applications are "compliers" with the intervention policy (Angrist et al., 1996).

The theory of costly school search has testable predictions. People should not know about all the available schools. People should report that the activities involved in search are challenging, and that one reason they stopped searching is that they thought they would be placed. And, if in addition people tend to be over-optimistic, they should respond to information about admissions chances by searching more and adding schools to their applications.

We test these predictions using data from two school choice systems. The first is the national centralized choice system in Chile. Chile implemented centralized primary and secondary school choice in 2016. All cities in Chile use the same choice platform to implement a strategyproof deferred acceptance (DA) assignment mechanism. We use data from the entire system for the years 2018–2020. Our second setting is the school choice system in New Haven, Connecticut in 2020. New Haven uses a "truncated" DA mechanism in which applicants can list only a limited number of schools. Truncated DA mechanisms are not strategyproof, but they are less manipulable than the common alternative of Immediate Acceptance (Haeringer and Klijn, 2009; Pathak and Sönmez, 2013). Studying the Chilean and New Haven settings together allows us to consider the role of search under different implementations of DA, within different choice platforms, and in different cultural contexts.

We supplement our administrative records with extensive survey data on choice participants in Chile. As part of the 2020 Chilean choice process, the Chilean government surveyed families submitting applications to the choice process about their search for schools, their preferences over schools, and their beliefs about their placement chances. These surveys were administered online, after the submission of applications but before results were known. 48,929 applicants completed the choice survey. The combination of a very large sample size and novel questions about both the choice application and the search process allow us to construct a detailed picture of the way families navigate choice.

Survey findings provide strong evidence that strategic, costly search for schools is one of the central challenges applicants face in the choice process, and that our stylized model captures important elements of the way students use potentially inaccurate beliefs to build their application portfolios. We have four main survey findings.

Our first survey finding is that search is, in fact, costly, and that applicants have limited information about relevant schools. When asked about what steps they need to take to know a school, large majorities of respondents give a long list of attributes and activities, including academic performance, extracurriculars, and interviews with staff. Obtaining this information would typically require both internet research and in-person visits or phone calls. Only 17% of respondents report that they know a randomly-chosen nearby school well, compared to 64% who report knowing their first choice well.

Our second survey finding is that the choice to terminate search is a strategic one to which beliefs about admissions chances are an important input. When we ask applicants why they did not add more schools to their list, the modal response is that they think they will be placed at one of the schools on the list already. 35% of respondents give this answer, compared to 30% who say they stopped adding schools because there are no more schools around. Applicants reporting higher subjective placement probabilities are much more likely to say they stopped search because they thought they would be placed.

Our third survey finding is that applicants are over-optimistic about their admissions chances. On average, respondents submitting applications with non-zero risk of non-placement overestimate their admissions chances by 32 percentage points. Applicants with true placement chances close to zero report average subjective placement beliefs of nearly 70%. Beliefs matter for search, but they are often wrong.

Our fourth survey finding is that the welfare stakes are large. Only 12% of applicants report that they would be at least somewhat satisfied with an outcome of no placement, compared to 69% who report they would be satisfied with the last-ranked school on their application. Differences in satisfaction manifest in enrollment choices. 93% of applicants placed at a school where they say they would be very satisfied go on to enroll in the school, compared to 40% of students placed at schools where they say they would be unsatisfied.

The survey results suggest that access to information about admissions chances would be helpful to applicants searching for schools. However, providing this information presents a logistical challenge. Placement chances are attributes of applications, not schools, and depend not just on individual submissions but on the distribution of submissions in the market.

We develop a new approach to address these challenges, which we call a "smart matching platform." The smart matching platform aggregates back-end data on the distribution of submitted applications to produce live, personalized predictions about application risk for platform users on the front end. This approach combines several features that past research has shown to be critical for successful information interventions, including the reduction of computational burdens, timely provision, and provision from a trusted source (Mani et al., 2013; Fernandes et al., 2014; Dynarski et al., 2021).

We evaluate smart matching platforms in Chile and New Haven. In both cases, the platform warned applicants submitting applications with high non-placement risk. In Chile, these warnings consisted of a pop-up in the application platform, as well as off-platform text messages. In New Haven, warnings came via email and directed applicants to an application simulator, which they could use to view placement chances for hypothetical applications. To assess risk in advance of application deadlines, policymakers combined data from previous years with data on applications already submitted in the current cycle. These policies were implemented nationwide in Chile starting in 2017, and in New Haven starting in 2020.

Because choice administrators need to choose some cutoff for what makes an application "risky," risk warnings lend themselves naturally to a regression discontinuity research design. In the face of quantity limits on messaging, choice administrators in Chile also randomized the provision of off-platform messages on the intensive margin. That is, all risky applicants received a text message but some received an additional, earlier message including an image. This allows us to employ RCT research designs as well. These experimental and quasi-experimental approaches allow us to evaluate our theoretical model without restricting access to information or reducing policy efficacy.

Warning applicants about their risky applications leads to lengthened applications, reduced risk of non-placement, and enrollment in better schools. We focus first on Chile, where sample sizes are much larger. Policymakers designated all applications with at least a 30% predicted chance of non-placement as risky. All applications above that cutoff received the live notification on the choice platform.

Receiving a warning caused 21.6% of students (SE=1.0) to add schools to their applications, corresponding to the complier group in our model. This is an extremely large effect for a light-touch policy: only three of 241 such policies analyzed in DellaVigna and Linos (ming) generated higher take-up rates. Consistent with model predictions, essentially all of the application changes we observe in response to treatment are additions of new schools.

Students complying with the intervention reduced their non-placement risk by an average of 15.5 percentage points (SE=1.3), or 58% of mean ex post application risk. Most of the additional placements are at schools with slack capacity, suggesting that the intervention reduces market-level congestion. Applicants receiving the intervention are no less likely to enroll in their placed schools. This is consistent with the idea that the intervention does not cause students to match to schools they like less.

Applicants who receive risk warnings enroll in higher quality schools. Compliers with the intervention enroll in schools where test score value added is 0.10σ higher. These schools pay teachers 12% more, enroll 40% more students per grade, and are 54% more likely to charge students a monthly fee. The intervention helps families avoid the fringe of small, low quality schools that characterizes some voucher systems (Abdulkadiroğlu et al., 2018; Neilson, 2021).

Smart matching platforms are effective in a wide variety of choice settings. We find large treatment effects across cities and years, and effects do not vary with market-level school choice experience, suggesting a limited role for "learning-by-doing" about admissions chances. We observe treatment effects in markets of all sizes, but applicants with more schooling options nearby tend to add more schools to their applications. Effects are large for both high- and low-SES students. Results from the text-message RCT show that our findings are not local to the cutoff, and that warnings matter on the intensive margin.

The "smart" part of the smart platforms— the personalized risk information— is critical for their effectiveness. We present four pieces of evidence on this point. First, we show that people who receive warnings change their beliefs about placement risk, consistent with the key causal channel in our model. Second, we present results from a series of "behavioral nudge" RCTs that encouraged people to add schools to their application, but did not include risk information. These nudges did not change behavior. Third, we show that personalized smart platforms outperform warnings about *aggregate* risk. Fourth, we show that "coercive nudges" that require students to add schools to their application but do not explain *why* adding schools is important lead to low rates of enrollment in placed schools. This contrasts with what we see in our smart platforms intervention, and is consistent with the hypothesis that smart platforms work because they motivate applicants to engage in meaningful search.

In the last part of the paper, we present results from a smart platform intervention in New Haven, Connecticut. While the broad structure of the New Haven intervention paralleled the approach in Chile, the cultural context, choice institutions, and intervention details differed substantially. Nevertheless, the intervention had similar effects. 13.8% of applicants near the risk cutoff comply with the intervention policy by lengthening their application; these applicants reduce their application risk by 42%. Also as in Chile, a randomly assigned nudge without risk information had no effect on choice behavior.

We contribute to three strands of literature. First, we show that strategyproofness within the school choice problem does not correspond to strategyproofness in the broader choice *process*, and that the divergence between the two places substantial information demands on participants. Many papers consider how students make choices under different assignment mechanisms (Abdulkadiroğlu et al., 2011; Pathak and Sönmez, 2013; De Haan et al., 2023; Agarwal and Somaini, 2018; Calsamiglia et al., 2020; Kapor et al., 2020). These papers analyze behavior in the choice problem, and typically ignore deviations from optimal behavior in strategyproof settings. We show that these deviations are empirically important and provide an economic rationale for why they occur.

An emerging literature considers the search aspect of school choice directly. Several recent papers use theoretical and laboratory approaches to study the equilibrium implications of costly (but rational) search in matching markets (Chen and He, 2020, 2021; Immorlica et al., 2020; Hakimov et al., 2021). Son (2020) and Ajayi and Sidibe (2020) use application data from centralized choice systems to estimate empirical models that allow for limited consideration sets and belief errors.[2] Our empirical contributions here are to provide survey evidence that the frictions these papers build into their models are important in practice, to provide credible tests of model predictions that shocks to beliefs affect search, and to demonstrate that smart matching platforms are an effective policy response. From the theory side, our contribution is to unpack the way systematic belief errors affect search from the perspective of the individual applicant. Our work fits into a broader set of studies that consider how strategic actions taken prior to participation in centralized mechanisms affect assignments within the mechanism, for example in spectrum auctions (Doraszelski et al., 2017; Milgrom and Segal, 2020).

Our second contribution is to illustrate the importance of information interventions that target search *strategy*, as opposed to fixed product attributes. Research on both education and product markets explores the effect of providing consumers with information on choice attributes (e.g. Jin and Leslie, 2003; Hastings and Weinstein, 2008; Allende et al., 2019). Findings are mixed, with some interventions changing choices and others finding precise zeros (e.g. Gurantz et al., 2021). Though our intervention is conceptually quite different, our findings can help rationalize null results in some attribute-focused studies. If applicants are confident they will be admitted to a school they like, they may not think it is worth it to conduct the additional due diligence required to add a new option to their portfolio, even when prompted with appealing (but limited) information about that option. On the measurement side, we innovate by linking scaled policy evaluation with participant surveys. Direct evidence on how people approach the economic challenges of market participation is crucial for designing interventions on strategy and understanding why they work.

Our third contribution is to show the power of combining market design principles, which limit the need for strategic sophistication, with “prediction machines” (Agrawal et al., 2018) and “choice engines” (Thaler and Tucker, 2013), which distill complex datasets into the information people need to make the strategic decisions that remain. We bring narrow AI into a matching setting where it aggregates information on market-level outcomes and identifies the part of that information relevant for specific participants. This contrasts with previous work focusing on attribute comparisons in product markets (Gruber et al., 2020).

2.2 Searching for Schools

Model Overview

We guide our empirical analysis using a model of search for schools with imperfect information about admissions chances. The theoretical analysis has two goals. The first is to show how beliefs about admissions chances affect students’ decisions to search for schools to add

[2] In addition, Grenet et al. (2021) model information acquisition in college choice. Bobba and Frisancho (2020) and Tincani et al. (2021) consider how college applicants learn about their own *abilities*.

to their applications. The second is to show how interventions that reduce optimism about placement can cause students to search more, discover more schools, and reduce application risk.

Our analysis takes the perspective of an individual student searching for schools to add to her school choice application. The approach is similar to models of job search (McCall, 1970), with the key difference being that agents in our model add schools they find to a multi-school application portfolio, from which placement outcomes are determined by a centralized assignment mechanism. This contrasts with the standard approach to job search models, in which agents decide whether to take jobs as they arrive, and search terminates once the agent accepts an offer. It also contrasts with models of the school choice problem that focus on market equilibria as the main outcomes of interest. Our model highlights the strategic challenges facing individuals even when the centralized assignment mechanism is strategyproof, and allows us to draw out the role of beliefs about admissions chances.

Model Setup

Consider an applicant to a strategyproof centralized assignment mechanism with limited information about what schools are available to her. The applicant is endowed at time zero with consideration set $\mathcal{C}_0 \equiv \{1, 2, 3, \ldots, N_0\} \subseteq \mathcal{J}$, where $\mathcal{J}$ is the set of all schools. The applicant receives utility u_j from placement at school j. Without loss of generality suppose $u_1 > u_2 > \ldots > u_{N_0} > 0$, and that utilities are measured relative to the outside option of non-placement, which yields utility zero. For each $j \in \mathcal{C}_0$, the individual knows their utility from placement at j, $u_j \in \mathbb{R}$, and has subjective beliefs about admissions chances $p_j \in [0, 1]$, which they believe to be independent across j.[3]

Individuals may choose to pay a cost κ, known to them, to add a school to their consideration set. If so, this school's subjective placement probability $p \in [0, 1]$ and utility $u \in \mathbb{R}$ are drawn from a distribution $F_{p,u}(p, u)$ with marginal distribution of utilities $F_u(u)$ and conditional distribution $F_p(p|u)$, where $F_u(0) = 0$ without loss.[4] We emphasize that although $F_{p,u}(\cdot)$ is the distribution from which new schools are drawn, the initial consideration set $\mathcal{C}_0$ need not be drawn from this distribution. Individuals have accurate beliefs about the distribution of utilities at schools outside their consideration set, $F_u(u)$, and potentially inaccurate beliefs about the distribution of admissions chances $F_p(p|u)$ that may depend

[3] In the empirical settings that we consider, admissions outcomes are determined by lotteries which are independent across schools. In principle, additional uncertainty about the general number of seats or level of demand might induce correlation in beliefs within a portfolio. For instance, rejection by school j might indicate that demand for some other school k was higher than the student had believed. In practice, school choice applicants seem to exhibit "correlation neglect" (Enke and Zimmermann, 2019; Rees-Jones et al., 2020).

[4] Suppose that, at constant cost cost $\tilde{\kappa} > 0$, students may discover a new school with utility distributed according to $\tilde{F}_u(\cdot)$ where $\tilde{F}_u(0) > 0$. Because search costs are sunk, if the expected benefit of finding a new school exceeds the cost at consideration set $\tilde{C}_0$ and cost $\tilde{\kappa}$, but the school that was found is unacceptable, it is worthwhile to search again. In expectation, the applicant will have to conduct $1/(1 - \tilde{F}_u(0))$ searches to discover a school with positive utility. Define $F_u(u) = \tilde{F}_u(u|u > 0)$ and $\kappa = \tilde{\kappa}/(1 - \tilde{F}_u(0))$.

on their value of being placed at the school. Search costs differ across individuals and are distributed according to $\Phi(\kappa)$, which we assume is differentiable with pdf ϕ.

This setup captures the idea that students need to know what a school is like before they apply to it. We think of κ as reflecting the cost of achieving this level of familiarity. As in the canonical school choice model, we assume that students know the utilities of the schools that they are considering. We also assume that students have accurate beliefs about the distribution of utilities of schools they have not yet discovered. These assumptions let us focus on the novel aspect of our contribution, which is to analyze the effects of erroneous beliefs about admissions chances.

The Value of Learning about a School

Define $R_j = 1 - p_j$ as the subjective risk of non-placement at school j. The value of the optimal portfolio given consideration set $\mathcal{C}_0$ is given by:

$$V(\mathcal{C}_0) = p_1 u_1 + p_2 u_2 R_1 + \ldots + p_{N_0} u_{N_0} \prod_{j<N_0} R_j. \tag{2.1}$$

Now consider the set $\mathcal{C} = \mathcal{C}_0 \cup \{s\}$, where school s has utility u_s and "chance" p_s. Let $r = \min\{j \in \mathcal{C}_0 : u_j < u_s\}$ be the best school in the original consideration set that is dispreferred to s if such a school exists. If there is no such school, let $r = N_0 + 1$. We have

$$V(\mathcal{C}) = \sum_{j=1}^{r-1} p_j u_j \prod_{j'<j} R_{j'} + p_s u_s \prod_{j'<r} R_{j'} + \sum_{j=r}^{N_0} p_j u_j R_s \prod_{j'<j} R_{j'},$$

and

$$V(\mathcal{C}) - V(\mathcal{C}_0) = p_s(u_s - \Gamma_r) \prod_{j<r} R_j, \tag{2.2}$$

where

$$\Gamma_r = \sum_{j=r}^{N_0} p_j u_j \prod_{j'=r}^{j-1} R_{j'}$$

is the expected value of the application portfolio conditional on not receiving a placement at schools ranked better than r.[5]

Optimism and the Value of Finding a School

We assume a simple, multiplicative structure for belief errors. Let $R_j = (1-a)R_j^*$ for all j, where R_j^* is the true risk. Similarly, let $p_j^* = 1 - R_j^*$ denote the true chance of being admitted to j. The parameter a measures optimism: as a grows, people believe risk is smaller. Assume

[5]We adopt the convention that, when $h > l$, we have $\prod_{j=h}^{l} x_j = 1$ and $\sum_{j=h}^{l} x_j = 0$ for any x_j.

$a < 1$ so that people do not rule out all application risk, and assume $0 < R_j < 1$ for all $j \in \mathcal{J}$. Taking the log of $V(\mathcal{C}) - V(\mathcal{C}_0)$ and then taking the derivative with respect to a yields

$$\frac{d\log(V(\mathcal{C}) - V(\mathcal{C}_0))}{da} = \frac{1-r}{1-a} + \frac{R_s^*}{1 - R_s^*(1-a)} + \frac{d\Gamma_r}{da}\frac{1}{\Gamma_r - u_s}. \tag{2.3}$$

See Appendix B.1 for details.

The effect of optimism on the value of adding new schools operates through three channels. The first channel is that more optimism reduces the value of adding school s by increasing applicants' confidence they will be placed in a school they prefer to s. This is the first term in the sum. It is equal to zero if $r = 1$ (i.e., if added school s is first-ranked on the new application) and negative for $r > 1$. It will tend to be bigger as optimism grows.

Second, increased optimism raises the value of adding a school to the portfolio because applicants think they are more likely to be admitted to that school. The second term of the sum captures this effect. It is positive for all values of a.

Third, increasing optimism reduces the value of adding school s by raising the expected value of falling below s on the application. The third term of the sum is negative whenever s is not the last school on the application, in which case it is equal to zero. $\frac{d\Gamma_r}{da} > 0$, because optimism shifts students towards believing they will be placed at higher-ranked schools given that they have fallen below s. We have $\frac{1}{\Gamma_r - u_s} < 0$ because the value of a placement at s is larger than the expected value of possible placement at schools with lower utility than s.

These channels combine to affect the subjective value of adding school s to the application.

Proposition 1. Let $\mathcal{C}_0$ contain $N_0 \geq 1$ schools, and let school $s \notin \mathcal{C}_0$ have $0 < u_s < u_{N_0}$. Then, letting $r = N_0 + 1$, we have $\frac{r-1}{rR_s^*} > 0$, and the value of adding s to the application is decreasing in a whenever $a > 1 - \frac{r-1}{rR_s^*}$.

Proof. See Appendix B.1.

This proposition shows that for sufficiently high levels of baseline optimism, additional increases in optimism reduce the value of adding schools to the bottom of the application. As we discuss below, this case—optimistic students adding schools to the bottom of their applications—is the modal one in our setting. More broadly, this analysis shows that information on admissions chances can be important to choice strategy even if it does not affect the applications students submit given their consideration set.

Information Interventions and Search Behavior

The expected value of search $U[\text{Search}|\mathcal{C}_0, a]$ is given by integrating the value of adding a newly discovered school s over the distribution of utilities and subjective admissions chances:

$$U[\text{Search}|\mathcal{C}_0, a] = \int\int \left(V(\mathcal{C}_0 \cup \{s\}) - V(\mathcal{C}_0)\right) dF_{p,u}(p_s, u_s),$$

where s has utility u_s and subjective placement chance p_s. At the optimal strategy given applicants' beliefs, we have $U[Search|\mathcal{C}_0, a] \leq \kappa$; otherwise applicants would not have stopped searching.

Taking this optimal portfolio as a starting point, consider how a decrease in optimism, $-\Delta_a$ for $\Delta_a > 0$, alters search behavior. Individuals for whom this change reduces the value of search cannot "unsearch," so their search behavior does not change. Individuals for whom changing optimism raises the value of search, such as those identified in Proposition 1, increase search if their decision to stop was marginal.

Proposition 2. Consider an applicant with optimism a who has searched optimally given this level of optimism. The effect of a surprise reduction in optimism by Δ_a is to weakly raise the probability of further search, and to raise the probability of adding at least one school to the choice application by an equal amount.

Proof. See Appendix B.1.

Applicants who add at least one school to their application in response to the information treatment Δ_a are *compliers* with the intervention policy. In our model, this set is identical to the set of people who engage in additional search.

Adding schools to the application reduces non-placement risk. Compliers' true non-placement risk falls by at least the expected amount induced by adding one school. Define non-placement risk prior to the change in a as $RISK_0 = \prod_{j \leq N_0} R_j^*$. Then, the change in placement risk after adding a given school s to the application is

$$RISK - RISK_0 = R_s^* \times \prod_{j \leq N_0} R_j^* - \prod_{j \leq N_0} R_j^* = -RISK_0 \times p_s^*.$$

Integrating over schools s that an individual may add to his application, it follows that compliers' risk falls by at least $RISK_0 \times E(p^*)$. In sum, we expect information interventions implemented at the conclusion of search to raise search rates, cause individuals to lengthen their applications, and reduce non-placement risk.

Enrollment and Welfare

Appendix B.1 extends our baseline model to include applicants' decisions about whether to enroll in the school where they are placed. The insight this extension delivers is that individual utility from an information intervention increases in proportion to placement rate, except to the extent it is offset by declines in enrollment conditional on placement. Enrollment is a common measure of satisfaction in market design research (Abdulkadiroğlu et al., 2017; Kapor et al., 2020). Section 2.4 presents evidence that it applies in our setting as well.

Discussion

Our goal is to study the impacts of interventions that provide accurate information about placement chances in settings in which applicants tend to be optimistic. One might extend our model to relax the assumptions that applicants know their utilities, know the distribution of utilities of schools they have not considered, and can discover acceptable new schools at a constant cost. These assumptions are not essential, and are not imposed in our empirical work. In addition, it is unlikely that our empirical findings are driven by violations of these assumptions, as the channels that our simplified model rules out would tend to push the impacts of our interventions toward zero. See Appendix B.1 for further discussion.

2.3 Setting

Centralized Choice in Chile

We study the importance of costly search using nationwide survey and administrative data from Chile and district-level data from New Haven, Connecticut. We focus first on Chile, where sample sizes are several orders of magnitude larger. This section describes school choice institutions in Chile and interactions between policymakers and choice applicants that help us understand the role of search. We return to the New Haven setting in section 2.7.

Chile introduced nationwide, voucher-based school choice in 1981 (Hsieh and Urquiola, 2006). Students receive vouchers they can spend at schools, and schools may charge limited additional fees. For the first 35 years, school choice in Chile was *decentralized*. Families applied to each school separately. In 2016, policymakers adopted centralized assignment with the goal of making the school choice process more transparent and equitable (Gobierno de Chile Ministerio de Educación, 2017). The centralized choice system was rolled out on a region-by-region basis, with adoption in all cities by 2019 and all grades by 2020. The centralized process includes 93% of primary school matriculation in the country, covering almost all public schools and private schools that accept school vouchers.[6] 450,000 applicants participated in 2020.

All cities in Chile use the same choice platform, which assigns students to schools using a deferred acceptance (DA) assignment mechanism (Correa et al., 2019). To ration seats in oversubscribed schools, the mechanism combines coarse sibling, school employee, and alumni priorities with lottery-based tiebreakers.[7] Applicants may list as many schools as they want on their choice application.[8] This means that the mechanism is *strategyproof*. The approach Chile takes to centralized assignment is similar to that used in major US districts such as New York and Boston (Abdulkadiroğlu et al., 2005,?).

[6]The remaining 7% of PK-12 students enroll in expensive private schools that do not accept vouchers or in schools where the highest grade is Kindergarten. These schools do not participate in the centralized process.

[7]Alumni priorities are for students who want to return to a school they previously attended. Schools also use quotas for vulnerable students and, in a very small number of cases, for high-performing students.

[8]Students applying in zones with more than one option who are either entering the schooling system from outside or enrolled in a school that does not offer the next grade must list a *minimum* of two schools.

The centralized school choice platform opens in August of each year. Applicants have access to the platform for roughly one month, during which time they may view, submit, and edit their applications. The application deadline falls in early September, and students are notified of their placements in late October. Applicants who receive a placement can choose to turn down that placement if they want. Applicants who reject their placement, who are not placed, or who did not participate in the main round can join a secondary application process in late November that lasts one week. Between early January and the beginning of the school year in March, students who still do not have a placement and placed students who decide to decline their placements may enroll in undersubscribed schools, outside of the centralized system. We focus our analysis on the first placement round, which accounts for more than 90% of placements over the period we study. See Appendix B.3 for further discussion of school choice institutions and enrollment outcomes for unplaced students.

We analyze the choice process using data on all applicants to the centralized platform between 2018 and 2020. We describe the applicant population in Table 2.1.[9] The platform received just under 1.2 million applications (defined at the student-year level) over this period. 49% of these applications came from students identified by the Chilean Ministry of Education (Mineduc) as "economically vulnerable," a classification based primarily on income and benefits receipt. 95% of applicants come from urban areas, as defined by the 2017 Census.[10]

Many applicants interact more than once with the application platform between the time it opens and the application deadline. Panel B of Table 2.1 describes these interactions. The first portfolio an applicant submits contains an average of 2.8 schools. Following their initial submission, applicants are free to revisit their submission and change, add, or subtract schools at any time before the deadline. At the deadline, the average portfolio length rises to 3.1 schools. The average applicant submits 1.4 distinct portfolios to the centralized platform before the deadline. 25% of applicants submit a final application that differs from their initial application. The most common change is to add a new school to the application: 21% of all applicants have a school on their final application that was not on their initial application. Most people who add schools add them to the bottom of their portfolio– 18% make such an addition– but 3% add a new school to the middle of their application (i.e., above some but not all previously-ranked schools) and 2% add a school to the top (above all previously-ranked schools).[11] Columns 2 and 3 of Table 2.1 show that lower-income students tend to have shorter applications and are less likely to change their applications.

Most but not all students receive a placement through the centralized process. As reported in Panel C of Table 2.1, 79% of applicants receive a placement at some school on their first-round application. 54% of students are placed in their first-ranked school, 13% in their second, and 6% in their third. 5% of students place at a school lower than third.

[9]See Appendix B.4 for a discussion of our data sources.

[10]The Census definition of urban areas includes (primarily) all settlements with more than 2000 inhabitants. We define applicants' geographic zone based on the location of their first-choice school. Individual geocoding is unreliable for a large portion of applicants, while school locations are known precisely.

[11]See Appendix Table B.1 for details on the changes students make to their applications.

Placement rates are *higher* for lower-income students despite their shorter applications. 84% of low-income students receive a placement, compared to 74% of higher-income students. 9% of students who participate in the first round go on to participate in the second centralized round, and 7% receive a second-round placement.[12]

Non-placement occurs despite slack capacity. Panel D of Table 2.1 displays the (applicant-weighted) average share of seats in a market that are unfilled after the first placement round. On average, participants apply in markets where 42% of seats are unfilled; the share of unfilled seats in schools that are free to students is even higher. These values exceed the share of students placed in the second placement round, indicating that follow-on attempts to fill slack capacity do not fully succeed.

Most students who are placed in a school enroll in that school. As reported in Panel E of Table 2.1, nearly all (97%) students enroll in *some* school. 62% of students enroll in a school where they were placed through the centralized process, reflecting a compliance rate of 78% for the 79% of students who receive a placement.

We describe the schools students attend using school-by-year outcome and input data from Neilson (2021). Our main measure of quality is test score value added (VA). The scale is student-level standard deviations, with the mean normalized to zero in 2016. We measure VA using fourth grade scores, which are available for most primary schools but few schools serving grades nine and up. We focus our VA analysis on students in grades eight and below. 77% of these students enroll in schools with a VA estimate. See Appendix B.4 for details.

Students who enroll through the centralized process enroll in better schools. Mean value added for students who enroll at their placed school is 0.11, compared to 0.04 for other students. This gap is larger (0.09 SDs) for economically vulnerable students than for other students (0.04 SDs). Low-SES students enroll at schools with lower average monthly fees than high-SES students and with higher shares of low-SES peers.

Intervention Design

Heading into the 2017 process, non-placement risk was a major concern for education policymakers in Chile. Our research team worked with Mineduc to evaluate the causes of non-placement risk and formulate a policy response. Preliminary descriptive and qualitative evidence suggested that some families had inaccurate, overly optimistic beliefs about their chances of being assigned to schools. Based on this evidence, we helped Mineduc design a set of information interventions alerting applicants to non-placement risk. These interventions identified applicants whose submissions placed them at risk of non-placement, and notified them of this risk prior to the close of the application deadline.

The key feature enabling these interventions is the ability to interact with both application data and applicants themselves in real time over the course of the application process, to compute and communicate risk. The technical and logistical demands of implementing live feedback at scale led one member of the research team (Neilson) to found an NGO, Con-

[12] Applicants who do not participate in the first round are not included in our analysis.

siliumBots, specializing in school choice services. The NGO partnered with Mineduc to run the interventions from 2018 on. See our Disclosure Statement for details on the relationship between the research team and the implementing partner NGO.

Mineduc conducted two kinds of information interventions over the period we study. We summarize them here with additional detail in Appendix B.3.

The first intervention was an interactive pop-up embedded in the application platform, which we label the *platform pop-up*. This intervention computed a predicted risk value for each application submitted through the platform. Applications identified as "risky"– defined as having a non-placement risk greater than 30%– received a pop-up warning about their application immediately after they clicked submit. The warning stated that many families were applying to the same schools, and not enough seats were available for all applicants. It encouraged students to add more schools to their applications, while also offering them the option to continue and submit the application as-is. Appendix Figure B.1 displays the pop-up, with key text translated to English.

Mineduc implemented this intervention throughout the choice system. In 2018 and 2019, Mineduc activated the pop-up functionality one to two days after the date that applications opened. This delay reflected a combination of implementation difficulties and a desire to collect data on early applications for use in demand predictions. Our empirical analysis of pop-up effects in 2018 and 2019 excludes the students who submitted their first application attempt before the pop-up came online. These students made up 39% of applicants in these years. In 2020, the pop-up was available over the full application window for most applicants.[13]

Column 4 of Table 2.1 describes the 73% of applicants who submitted applications at times and in markets where the pop-up was available. We label this group "pop-up eligible" because members received a warning if their application was deemed risky. Pop-up eligible applicants resemble the full population in their demographic characteristics and application behavior.

The second kind of intervention consisted of supplemental "reminders" to risky students. These reminders were delivered via text message or the messaging service WhatsApp, and contained information similar to the pop-up.

Our analysis of reminder interventions focuses on the 2020 application cycle, when Mineduc sent a sequence of up to three messages to applicants who submitted risky applications. As in previous years, these interactions began with the pop-up intervention on students' initial application submission. All applicants who had submitted risky applications as of day 20 of the application cycle received a text message from Mineduc. Mineduc sent another text message to risky applicants on day 27 (the day before application close) repeating this information and providing a link to the student's choice application.

On day 25 of the application cycle, between the two text messages to all risky applicants,

[13]Demand predictions for early applicants in 2020 relied on data from the previous year. We did not have previous year demand data for students applying to non-entry grades in the Metropolitan Region, hence the pop-up was activated later for them (9% of total 2020 applicants). See below and Appendix B.6.

Mineduc and the NGO conducted an RCT evaluation of a WhatsApp intervention. We call this the *WhatsApp RCT*. The NGO chose a random subset of ten thousand risky applicants and sent them a WhatsApp message with an image containing a personalized risk warning.[14] The warning stated that their risk of non-placement was high, and suggested that students add schools to their applications to address this risk. Two factors motivated the WhatsApp RCT. The first was the idea that an image sent through the popular messaging service might be an effective supplement to the other interventions. The second was a constraint placed by the WhatsApp messaging contractor, which capped the number of messages that could be sent. Appendix Figure B.2 outlines the time path of interactions with risky applicants in 2020, and presents images of each intervention.

The set of reminders implemented in 2020 built on a more limited reminder policy implemented in 2018. In 2018, Mineduc sent a single SMS message to all risky students four days before the application deadline. Mineduc did not send any reminder messages in 2019.

We evaluate the platform pop-up using a regression discontinuity design around the 30% risk cutoff. In 2019, the RD estimates capture the effect of the pop-up for applicants near the cutoff. In 2018 and 2020, the RD estimates capture the effect the pop-up and its interaction with the subsequent reminder interventions. Our goal in the RD analysis is to provide proof of concept that smart platform information interventions affect search behavior and placement outcomes, not to unpack the differential effects of interventions by medium and timing. In what follows, we present RD estimates separately by year. Readers who are interested in understanding the effects of pop-up absent their interactions with subsequent reminders can focus on the 2019 implementation year.[15]

We evaluate the WhatsApp reminder in a standard RCT framework. Because treatment and control in the WhatsApp RCT are drawn from the set of students who still have risky applications after receiving previous reminders, the RCT evaluation tells us about intensive margin treatment effects in a group that is negatively selected on its response to previous similar treatments. It also provides information on the distribution effects both close to the risk cutoff and higher in the distribution of application risk. Putting the RCT together with the RD yields a rich picture of how information on admissions chances shapes outcomes for students at different points in the risk distribution and at different points in the choice process.

In addition to our main analyses of the 2018-2020 platform pop-up and the WhatsApp RCT, we present some supplemental results from the 2016, 2017, and 2021 choice processes. The process in these years was similar to 2018-2020. We note relevant cross-year differences in the text as needed, with details in Appendix B.3.

[14] In addition to high application risk, the NGO imposed other restrictions on the sample universe for RCT randomization. To be RCT-eligible, applicants needed to be a) early-grade applicants in b) urban zones who c) did not have access to sibling priority. In addition, they d) had to have declined engagement with previous Mineduc outreach attempts (unrelated to application risk) sent via email. See Appendix B.3.

[15] See Appendix B.5 for a detailed discussion of interactions between treatment types.

Application Risk and Risk Predictions

Predicted application risk is a critical input to the interventions we study. The NGO computed application risk in each market-year as follows. They first obtained the vector of reported school capacities for the current year, a projected number of applicants N, and a dataset of applications and student types (i.e. priorities). For the first few days of each market-year, these data consisted of the previous year's joint distribution of applications and priorities. For the remaining days, these data consisted of submissions thus far in the current process.

The NGO's algorithm resampled N (application list, student priority type) tuples from this dataset, drew lottery numbers, and simulated the matching process. Repeating this process 500 times, the NGO computed the probability of non-placement within each school-grade-priority group. This procedure is related to the resampling approach introduced by Agarwal and Somaini (2018) for calculating placement probabilities.

The NGO then developed a web service that used the calculated probabilities to predict the risk of non-placement for any individual application. These are equal to the probability of not being assigned to any of the schools in the list, for the specific grade and priority of the applicant. For more details on simulation and demand prediction see Appendix B.6.

Risk predictions closely track applicants' true non-placement risk. Panel A of Figure 2.1 describes the distribution of predicted placement probabilities across different values of true, ex-post placement probability. The ex post placement probability is constructed identically to the placement prediction, but using realized rather than predicted applications. Predicted values cluster around true placement probabilities across the distribution. The slope of the predicted value in the true value is 0.81, with deviations from one driven by slight but systematic underprediction of risk among the most risky applications. Our assessment is that the predicted risk measure provides a reasonable guide to true risk, particularly in comparison to applicants' risk beliefs, which we discuss in detail below.

Many applicants submit risky initial applications. Panel F of Table 2.1 describes ex post (or "true") risk on the initial application attempt. Mean non-placement risk on the initial application is 24%. A majority– 59%– of applicants are almost sure to be placed. We classify individuals as facing zero risk if their nonassignment probability is less than 0.01. At the same time, many applicants submit very risky applications. 30% submit initial applications with non-placement risk above 30%. Median risk for students submitting applications with non-zero risk is 62%, and 25% of such applicants have non-placement risk of 92% or higher. Panel B of Figure 2.1 plots the histogram of the risk distribution for the first and final application attempts. Mass stacks on the edges of the distribution, at very high and low risk levels. Mass shifts slightly towards lower-risk applications between the initial and final submissions.

As reported in column 5 of Table 2.1, 20% of all applicants—233,678 students over the three years—are classified as risky by the choice platform based on their initial application. Risky applicants are less likely to be economically vulnerable than other applicants and more likely to come from urban areas. They submit shorter initial applications than the sample population as a whole, but longer final applications, and are more likely to change their applications between their initial submission and the deadline. 45% end up being placed at one of their preferences in the first round, while 11% receive a second-round placement.

Appendix Table B.2 describes the sample of students critical to our analysis of the effects of application warnings. Applicants near the cutoff for receiving a pop-up warning (defined here as having non-placement risk between 0.1 and 0.5) have slightly higher socioeconomic status, slightly longer applications, and similar rates of application changes to the full sample. Like the broader sample of risky applicants, the sample of risky 2020 applicants in the text message RCT is relatively high-income and characterized by longer choice applications and more frequent engagement with the choice process than the population as a whole.

Survey Design

To learn more about how families engaged with the choice process, the NGO helped Mineduc conduct a survey of choice participants in 2020. The survey asked questions about several school choice topics. It included modules about preferences, beliefs, and search designed to provide context for the interventions we study here. The survey innovates over past surveys of choice participants (De Haan et al., 2023; Kapor et al., 2020; Wang and Zhou, 2020) by recruiting a larger sample and by asking about search in addition to preferences and beliefs. See Appendix B.7 for survey text.

Mineduc contacted students using an email message sent from the official school choice email account. Mineduc sent the message following the application deadline, but before the release of placement outcomes. They chose this time to maximize applicants' recall of their school choice experience while ruling out the possibility that the survey might affect applicants' portfolios. In total, Mineduc contacted 373,710 families. 48,929, or 13%, completed the survey. Appendix Table B.2 describes survey respondents. They are slightly less likely to be economically vulnerable and rural than the population as a whole, but closely resemble the broader population in terms of application behavior.

2.4 Survey Findings

Placement, Enrollment, and Student Welfare

The main focus of our analysis is whether students receive any placement through the centralized mechanism. Evidence from our applicant survey supports the idea that placement vs. non-placement is a critical margin from a welfare perspective. The survey asked respondents to report how satisfied they would feel if they were placed at the first-ranked school on their application, if they were placed at the last-ranked school, or if they were unplaced. At the time of the survey, applicants had submitted their applications but not received results, so responses reflect certainty over what the schools in question were, but not ex post rationalization of known outcomes. Appendix Figure B.3 reports two findings, which we summarize here. First, most (69%) of applicants would be satisfied with a placement at their last-ranked school, while nearly all (89%) would be *unsatisfied* with nonplacement. Second, the choice to enroll in the placed school tracks measures of preference for the school. 93% of students placed in schools they give the highest satisfaction rating choose to enroll, compared to 40% at schools with the lowest rating.

Search Costs and Search Strategies

We now turn to the question of how applicants search for schools. Our first result here is that getting to know a school well requires a lot of information, some of which may be costly to obtain. Our survey asked respondents what they needed to know about a school to feel that they knew it well. Respondents could select multiple options from a list of possibilities. As reported in Panel A of Figure 2.2, large majorities gave a long list of attributes. Some of these attributes are relatively easy to learn about from public sources. 83% said they would need to visit a school's website, and 93% said they would need to learn about a school's academic performance, which is also available online. Information on others, like extracurricular activities or school infrastructure, could likely be obtained upon a short visit. However, some kinds of information that respondents value would likely be hard to find. For example, 66% of respondents said they needed to interview school staff. 79% said they required references from current families.

Our second result is that applicants do not feel that they know many schools well. We asked each respondent how well they knew a randomly-selected nearby school, a nearby school that was high-performing and expensive, and a nearby school that was low-performing and free.[16] We also asked respondents about a "fake" school– i.e., a school that did not exist. Panel B of Figure 2.2 reports the share of students that claim to know each school well. Only 17% of students report knowing the random nearby school and the high-performing, expensive school well. 14% report that they know the low-performing, free school well. Encouragingly, only 3% report knowing the fake school well. Search is costly enough that at the end of the choice process, most families do not feel well-informed about many nearby schools.

Consistent with the idea that applicants learn about schools before applying to them, respondents claim to know the schools on their applications better than they know randomly chosen nearby schools. Panel C of Figure 2.2 displays applicants' responses to a question asking how much they knew about the schools on their submitted application. 64% of students claim to know their first-listed school well and 48% claimed to know the second-listed school well. Knowledge declines with application rank, but 30% of students who

[16]Schools in this question were selected from the alternatives within 2km from the residential location of the student that were not included in her application. We used the performance classification of the "Agencia de la Calidad de la Educación," an institution that classifies schools in 4 tiers using standardized test scores, after taking into account socioeconomic status of the student body. We classify a school as "high-performing" if it is in the top two tiers and "low-performing" if it is in the worst tier. "Expensive schools" are those that charge a monthly copayment of at least $35 USD on top of the voucher.

submit applications including at least five schools claimed to know the fifth school well. This is nearly twice the share claiming to know a randomly-chosen school well.

We now turn to the role of beliefs about admissions chances in search. Proposition 1 in our model provides conditions under which applicants who think they will be admitted to a school in their existing portfolio will be less likely to engage in additional search. Two survey findings suggest that this kind of behavior is widespread.

First, we asked applicants directly why they stopped adding schools to their application. Respondents could choose from four options: (1) there were no more schools to around to add, (2) there were schools around but they would rather not attend these schools, (3) it is hard to find more schools, and (4) they think they will be placed at one of the schools already on their application.

The most common reason applicants give for stopping search is that they think they will be placed in a school already on their list. As reported in Panel A of Figure 2.3, 35% of respondents chose this option. Another 17% said they stopped adding schools because additional schools were hard to find, a response that also invokes costly search. Together, these two search-related responses account for a majority (52%) of all responses. We interpret this as a likely lower bound on the share of respondents for whom costly search affected choice, since costly search might also have played a meaningful but not primary role for applicants giving other responses. The remaining 48% of respondents gave answers more in line with the traditional school choice problem, in which applicants list all available schools ("no more schools around") or list schools preferable to an outside option ("I'd rather not be placed at remaining schools").

Second, applicants who thought their chances of being placed were high were the most likely to say they stopped search because they thought they would placed. Our survey asked respondents what they thought their chances were of being placed at any school on their submitted portfolio. Panel B of Figure 2.3 plots the share of students saying they stopped search because they thought they would be placed at one of their submitted options at each quintile of the distribution of subjective placement chances. Respondents become much more likely to give this reason for stopping search as their subjective placement beliefs increase. 51% of respondents in the top quintile of the subjective belief distribution said they stopped search because they were confident in their placement chances. In contrast, only 9% of respondents in the bottom quintile gave this reason for stopping search.

Optimism and Search

Our first set of survey findings shows that search for schools is hard, and that beliefs about placement chances are a critical input to search strategy. Our second set of findings shows that these beliefs are wrong. We do so by comparing respondents' reported beliefs about placement chances to our calculations of objective placement chances.

Panel A of Figure 2.4 shows the distribution of subjective and true placement chances for applicants with non-zero risk of non-placement. Applicants far overrate their placement chances. The mean subjective placement probability is 76%, 32 percentage points above the mean true placement probability of 44%. The graph shows a mass of subjective beliefs piling up around a placement probability of one. The densest part of the distribution of true placement chances for these students is near zero, with no corresponding mass in subjective beliefs. Panel B shows the distribution of optimism, defined as the difference between subjective and true placement chances. This distribution is shifted far to the right of zero. Many respondents overestimate their placement chances by fifty percentage points or more.

In a mechanical sense, the source of this optimism is that many applicants with low true placement chances think they are likely to receive a placement. Panel C of Figure 2.4 plots the distribution of subjective placement beliefs, binned into groups by true placement probability. If beliefs were accurate on average, they would follow the 45 degree line. We instead observe a weak positive relationship with a large upward shift. The mean subjective belief for applicants with true admissions chances near zero is close to 70%.

For comparison, we also plot the distribution of the NGO's predicted risk measure, as computed at the time of the application for the set of survey respondents. As in the full sample, risk predictions do not precisely track the final risk values. However, it is clear that predictions are much closer to true placement probabilities than are subjective beliefs.

Several pieces of evidence indicate that our belief measures are credible. We have already shown that beliefs are related to stated reasons for stopping search. Additional results in Appendix Figure B.4 show that our findings on the distribution of beliefs are consistent whether we frame the question in terms of placement chances or in terms of non-placement risk, and also that respondents' overall assessments of application risk are closely related to the level of application risk implied by their beliefs about school-specific placement chances.

2.5 Warnings, Choice Behavior, and Choice Outcomes

The Platform Pop-Up

Our survey findings show that many applicants strategize on the basis of overly-optimistic beliefs about admissions chances. Together with our theoretical analysis, this suggests that applicants should respond to warnings about non-placement risk by adding more schools to their portfolios. We test this proposition using experimental and quasi-experimental research designs implemented in the Chilean and New Haven choice systems.

We focus first on the platform pop-up administered to Chilean students inside the choice system. Because all students with at least a 30% chance of non-placement received this warning, we evaluate it using a regression discontinuity design. In our visual analysis of RD outcomes, we display binned means together with global polynomial fits, to provide a sense of broad patterns in the data and how they relate to observed discontinuities. When computing estimates of RD effects, we use local linear specifications with a triangular kernel and a bandwidth of 0.1. This bandwidth approximates that given by optimal bandwidth calculations (Calonico et al., 2014).[17]

[17] We report estimates obtained with Calonico et al. (2014) bandwidth selection in Appendix Table B.3

We first show that applicants' observable characteristics are unrelated to which side of the 30% cutoff they fall on. Panel A of Table 2.2 shows how the share of students from rural areas and the share of low-income students vary by position relative to the cutoff for the full sample and for each choice year. Cross-threshold differences in these attributes are small in economic terms. Because our sample size is quite large– roughly 41,000 applicants within the local bandwidth– our estimates are very precise, and some economically small effects are marginally statistically significant. Appendix Figure B.5 shows that there is no visual evidence of discontinuities in predetermined covariates or in the density of the running variable. These findings are consistent with the observation that the 30% cutoff had no significance for applicants prior to policy implementation.

Choice Behavior

Panels A through C of Figure 2.5 and Panel B of Table 2.2 show how receiving the platform pop-up changed choice behavior. Receiving a warning caused 21.4% of applicants to alter their submissions. Essentially all of these changes are additions to the application. Receiving a warning caused 21.6% of applicants to add at least one school to their application.[18] Students add an average of 0.34 schools, and ex post risk of non-placement falls by 3.3 percentage points, 13% of the below-threshold mean. These effects are stable across years.

The effects of the warning on choice behavior are extremely large in the context of light-touch policies. DellaVigna and Linos (ming) describe the results of 241 randomized evaluations of light-touch interventions implemented as public policy. The average effect of these interventions on take-up rates for the desired action is 1.4 percentage points, roughly 6% of the 21.6 percentage point effect we observe. Only three of the 241 policies had take-up effects of 20 percentage points or more.

As discussed in proposition 2, the 21.6% of students who add a school to their application in response to the intervention are compliers with the warnings policy. The second column of Table 2.2 displays instrumental variables estimates in which adding at least one school to the application is the endogenous regressor. The resulting effect estimates can be interpreted as LATEs for the complier group. Compliers add an average of 1.6 schools to their application list, and reduce their ex post non-placement risk by 15.5 percentage points, equal to 58% of the below-threshold mean. The share of compliers with the intervention policy is large, and the risk reduction within this group is substantial.

The changes applicants make to their applications are consistent with the idea that the intervention leads to additional search. As reported in Panel B of Table 2.2, most but not all students who change their applications do so by adding schools to the end. Receiving the warning raises the chance a student will add a school to the end of their application

and Figures B.7-B.10. Alternate approaches to RD estimation do not change our findings.

[18]These calculations compare students across the RD threshold. Hence, although applicants who add a school are a subset of those who alter an application, treatment effects need not be ordered in this way. The estimated share induced to add a school is slightly larger than the share induced to make any change because a slightly larger fraction of "control" students change their applications without adding schools.

by 20.5 percentage points, about 95% of the share of students adding any school to their application. The frequency with which students add schools to the end of their application indicates that Proposition 1's focus on students adding schools to the bottom of their rank list is empirically relevant. However, receiving a warning also causes 7.8% of policy compliers to add schools to the middle of their list. This suggests that at least some applicants are learning about new schools, and not just adding known schools to the bottom of their rank lists. Very few students add schools to the top of their rank list, indicating that for the most part students identify their top-choice schools early in the search process.

The platform pop-up does not cause students to drop schools from their rank lists. This is consistent with our model, in which students who find additional schools add them to their portfolio and do not "un-search." We find some evidence that a small share (1%) of students re-order the existing schools on their application in response to the intervention, although the visual evidence here is not as compelling.[19] The warning may prompt some students to revise their applications as their preferences change over time (Narita, 2018). However, any such effect is second-order compared to the share of students adding schools.

Enrollment and Welfare

Changes in application behavior translate to changes in choice outcomes. Panel C of Table 2.2 reports the effect of receiving the warning on placement outcomes. Students receiving a warning are 3.8 percentage points more likely to be placed in one of their listed schools. As expected, this closely tracks the reduction in non-placement risk (within one standard error).

Warnings do not produce lower quality placements. Overall rates of enrollment rise in proportion to changes in placement across the cutoff, and the rate at which students enroll in school conditional on placement is continuous through the cutoff value. Panel D of Figure 2.5 displays the RD plot for this outcome, which shows no evidence of a discontinuity.

In Appendix B.1, we show that the effect of the intervention on individual welfare is proportional to the change in placement rates, except as offset by declines in enrollment conditional on placement. Our findings suggest that there are no offsetting enrollment effects. The implication is that receiving the warnings intervention raises welfare (excluding search costs) for compliers with the information intervention by 21% (= 0.15/0.74)– the per-complier change in placement rate as a percentage of the below-threshold mean.[20]

[19] Appendix Figure B.6 displays this plot and plots for other outcomes not shown in the main text.

[20] 0.15 is the IV estimate of the Δ risk effect reported in Table 2.2. 0.74 is the mean placement rate at the risk cutoff, computed as the intercept term from our main RD specifications with placement probability as the outcome. Note that placement chances at the risk cutoff are slightly above 0.7. This is because the running variable in the RD is the *simulated* risk of the *initial* application, while the placement chances outcome is the *true* risk of the *final* submitted application.

Decongestion vs. Reshuffling

The goal of our analysis is to understand how beliefs about admissions chances affect school search and placement outcomes from the perspective of individuals. However, it is also useful to think about how information on placement chances affects market-level congestion. If the warnings policy causes applicants to place in undersubscribed schools, the individual gains we observe may come "for free," in the sense that other students are not displaced. Congestion effects are important to consider because beliefs interventions do not guide applicants towards specific schools. This contrasts with the provision of information on school attributes, which may point applicants towards oversubscribed schools.

We assess the congestion effects of the platform pop-up by looking at how receiving a warning affects placement rates at schools with excess capacity. As reported in Panel D of Table 2.2, receiving a warning raises the chances that students add at least one undersubscribed school to their application by 7.3 percentage points. Put another way, roughly one third of applicants adding at least one school add an undersubscribed school.

Most of the decline in application risk from receiving a warning comes from increased chances of placement in an undersubscribed school. Receiving a warning raises applicants' chances of placing at an undersubscribed school by 1.9 percentage points. This corresponds to an 8.8 percentage point increase for compliers with the warnings policy, 57% of the overall risk reduction of 15.5 percentage points reported in Panel B of Table 2.2. The warnings policy helps reduce congestion, a core goal of centralized choice (Abdulkadiroğlu et al., 2017).

School Quality and Characteristics

In addition to affecting whether applicants place in schools they like, receiving risk warnings shapes the characteristics of the schools students attend. Table 2.3 and Panels E through H of Figure 2.5 report results from RD specifications with characteristics of the schools where students enroll as the outcomes of interest. Panel A of Table 2.3 reports that nearly all students on both sides of the cutoff enroll in some school (inside or outside the centralized process). The share of applicants for whom value-added measures are available is also stable across the cutoff. Differential censoring is not a concern.

Our headline finding here is that receiving a risk warning improves school quality. Value added at the schools where students enroll rises by 0.022 student-level SDs across the cutoff. This corresponds to a 0.103 SD increase for compliers. This is a large effect. For example, it is roughly comparable to a one standard deviation improvement in teacher quality (Chetty et al., 2014) or one half to one quarter of the gains from attending a high-performing charter school (Abdulkadiroğlu et al., 2011). In our context, it is roughly equal to the difference in school quality between the schools that low-SES and high-SES students attend.

Measures of market demand, input intensity, and peer social status rise along with value added. Focusing first on demand measures, both price and quantity shift upward across the cutoff. Compliers with the warnings treatment are 15.2 percentage points (54%) more likely to enroll in schools that require some copayment, with the average monthly copayment rising by \$9 USD on a base of \$22. Total enrollment per grade (i.e., quantity) rises by 39.5 students on a base of 99 students. Turning to the input side, mean teacher compensation rises by \$3,700 USD, or 12%. Interestingly, spending per student is flat, suggesting that the high value added, highly demanded schools that students shift towards do not necessarily spend more per student, but do spend more efficiently. Finally, on the peer attributes side, compliers with risk warnings attend schools where the share of economically vulnerable peers is 2.9 percentage points (5.1%) lower. Distance from home to school does not change.

These findings support our revealed preference argument that warnings increase individual welfare. By facilitating search, the warnings treatment gives families the opportunity to make larger investments in their own education and avoid the small, low-price, low-quality schools associated with poor performance in voucher systems in Chile and elsewhere, including the US (Abdulkadiroğlu et al., 2018; Neilson, 2021). Appendix B.8 further explores the shifts in enrollment patterns that drive the observed increases in school quality.

Replication and Heterogeneity

The platform pop-up intervention increases search across markets. Appendix Figure B.11 describes the distribution of estimated effects over all markets (defined by city-year) and split by measures of market size. Looking across markets, modal values of treatment effects on adding any school and number of schools added are similar to the reported overall effects of 0.22 and 0.34, respectively. The cross-market IQR of the estimated effect of treatment on adding any school is (0.11,0.29), and, as reported in Appendix Table B.4, effects in the three largest markets are each close to the nationwide average. Splits by market size, as measured both by total number of available choice options and number of schools geographically close to an individual applicant, show that treatment effects on the add any school outcome are similar across different-sized markets, but that treatment effects on the count of schools added are larger in larger markets.

The effects of risk warnings on beliefs might diminish as market participants gain experience with choice. To test this, Appendix Table B.5 repeats Table 2.2, but splits the sample by the number of years a city-by-grade combination has used the centralized platform. Shifts in search and risk are similar for city-grade-years with one, two, or three or more years experience using the centralized platform. We see no evidence that the effects of the platform pop-up intervention decline as experience with choice rises. This is consistent with results in Kapor et al. (2020) showing large belief errors in a setting with a long history of choice.

Smart matching platforms affect both high- and low-SES applicants. Appendix Table B.6 repeats the analysis of Table 2.2, splitting by economic vulnerability. Rates of application modification and risk reduction are slightly larger for economically vulnerable applicants. As reported in Appendix Table B.7, attributes of enrolled schools shift for high- and low-SES students. Gains in VA are large for high-SES students and small (but noisily estimated) for others. Gains in teacher pay, enrollment per grade, and copayment fees, as well as declines in low-SES peer share, are all larger for low-SES applicants.

Warnings across the Risk Distribution

We use the random assignment of reminder message interventions to study how the effects of warnings about risky applications vary away from the 30% risk cutoff and on the intensive margin. In the 2020 choice process, randomly selected risky applicants received a WhatsApp text warning three days before the application deadline. 44 hours after that, on the day before the deadline, all risky applicants received the same warning through an SMS. In this context, what random assignment does is raise the number of warnings to which risky applicants are exposed between the time they first fill out their application and the application deadline. For non-risky students (below the 30% risk cutoff) treatment and control status are randomly assigned, but the "treated" group does not receive a risk warning.

Figure 2.6 presents the effects of the RCT by plotting outcomes for the treatment and control groups by application risk at the time of randomization into the text message treatment. Panel A shows that the number of warnings students receive (summing over all interventions) rises across the cutoff for both treatment and control groups, but rises more for the treatment group, which receives the additional WhatsApp message. The 0.48 difference in messages viewed for treatment relative to control among risky applicants reflects the share of applicants who opened WhatsApp and viewed the image.

Panels B and C of Figure 2.6 describe application behavior in the 44 hour window between the randomized message to the WhatsApp treatment group and the message to all risky students. Risky students randomly assigned to the WhatsApp treatment are more likely to add schools to their application and reduce their application risk than untreated students. On average, assignment to the treatment group causes 3.3% of risky students to add at least one school to their application. This corresponds to a 6.8 percentage point effect for each student that views the treatment image. These changes cause application risk to fall by 1.0 percentage points, or 2.1 percentage points per message view. The implied risk reduction for applicants who comply with the WhatsApp intervention by adding schools is 29.7 percentage points, equal to 49% of mean risk in the RCT sample.

Search and risk reduction outcomes in the treatment group outpace those in the control group over the full distribution of risk values above 30%. To facilitate comparison between RCT and RD estimates, panels B and C of Figure 2.6 display RD estimates for the WhatsApp treatment calculated across the risk cutoff within the treatment group. RD estimates are smaller than RCT estimates. We see little evidence that students near the 30% risk cutoff respond more to information interventions than applicants higher in the risk distribution.

Panels D and E repeat the analysis from Panels B and C, but now look at *all* application changes between the randomized warning and the application deadline. These measures include the effects of the final text reminder sent to all risky students. Despite the text message followup, gaps between treatment and control expand over time. As for the 44-hour outcomes, treatment-control comparisons span the full distribution of risk above the risk cutoff, except perhaps the very top. Average treatment effects in the RCT are larger for endline outcomes than for the 44-hour outcomes. 4.4 percent of students add a school to their application, and the mean risk reduction for these students is 30 percentage points.

Table 2.4 summarizes findings from the RCT and RD analysis of the WhatsApp intervention. Treatment and control groups are balanced on observable characteristics. For choice outcomes, we present both ITT effects reported in Figure 2.6 and IV estimates that take adding at least one school as the endogenous regressor of interest.

Overall, compared to the platform pop-up, the share of compliers with the WhatsApp RCT is smaller. This makes sense given that the RCT population is negatively selected on the response to previous interventions. However, the percent reduction in risk per complier is similar, and the percentage point reduction in risk is larger.

We draw two conclusions from this analysis. The first is that the effects of warnings persist across the risk distribution. Appendix B.9 provides additional evidence in support of this point from a 2017 pilot of the platform pop-up that included warnings cutoffs at 30%, 50%, and 70% risk levels. The second is that there may be benefits to providing the same person with information multiple times. The effects of information provision tend to be largest near the time of choice (Madrian, 2014). Providing multiple reminders may raise the chances that one is received around the time applicants need it.

2.6 Why Do Smart Platforms Work?

Smart platforms work. But why? Thus far, we have focused on the idea that the information intervention shifts students' beliefs about admissions chances, which in turn leads them to engage in costly search for new schools. Our survey analysis showed that inaccurate beliefs and costly search are key features of applicants' choice experiences. This section provides direct evidence that a) the intervention operates by shifting beliefs and b) interventions that do not include personalized information are not as effective.

Smart Platforms Change Beliefs, Not Preferences

Because our survey of placement beliefs took place after applications closed but before results were revealed, we can test the theoretical prediction that risk warnings shift beliefs by placing survey belief measures on the left side of our main RD specifications. Table 2.5 reports results from this test. Panels A and B show that receiving a risk warning does not affect the probability that applicants respond to our questions about subjective beliefs, and that respondents' behavioral responses to the risk intervention are broadly similar to those in the population.

Panel C of Table 2.5 reports how the intervention shaped beliefs. Applicants' average subjective nonplacement risk rises by 3.6 percentage points (22%) across the cutoff. Because applicants who receive the risk warning add schools to their lists, the estimated mean effect here understates the true belief shift, holding the application fixed. Applicants' subjective beliefs about placement in their first choice school fall by 4.9 percentage points (6.5%) at the cutoff. Because treatment does not cause applicants to alter their first choices and because admissions chances at the first choice do not depend on other features of the application, this estimate is closer to a "pure" beliefs effect. Both of these effects are visually apparent in standard RD plots. See Appendix Figure B.12.

Our survey also provides evidence that the treatment does *not* change preferences. In principle, applicants might draw inferences about the quality of schools on their choice applications from information about demand for those schools. Panel D of Table 2.5 places survey levels of stated satisfaction with (hypothetical) placement at the first-listed school on the left side of the RD specifications. We see no evidence that preference for the first-choice school changes, even as beliefs about admissions chances decline. While it is not possible to prove the null that the intervention had no effect on preferences, we view these results as a strong indicator that the intervention acts mainly by changing beliefs, rather than preferences.

Behavioral Nudges, Costly Shoving, and Impersonal Information

Testing "Behavioral" Nudges.

Nudge policies that encourage students to raise their placement chances by applying to more schools but do not include information about risk produce much smaller effects than smart matching platforms. In 2016, we worked with Mineduc to test a variety of behavioral nudges aimed at getting students to apply to more schools. These interventions were similar in format and timing to our later smart platform interventions, but did not contain any risk information. Our goal was to test whether approaches from the behavioral nudge toolkit could shift students towards less-risky applications. Appendix B.3 reports implementation details.

We considered three kinds of nudges. The "more schools, higher chances" nudge gave applicants guidance that applying to more schools increases one's chance of being placed. The "range heuristic" nudge told applicants that listing between five and ten schools increased one's chances of being placed. And the "role model" nudge told applicants that families who have submitted "good" applications typically listed five or more schools. Each of these options aimed to reduce the complexity of choice by providing guidance about how many schools to list. Because the interventions came from the choice authority (via SMS), they conveyed official approval for long lists. The second intervention adds to the first by providing a decision-making heuristic (Tversky and Kahneman, 1974). The third augments the second with a social pressure/social identity message (Lavecchia et al., 2016).

None of these approaches worked. Table 2.6 reports results pooling all of the behavioral nudge interventions and separately by treatment arm. In the full sample, the average effect was to raise the chance applicants added at least one school to their application by a statistically insignificant and economically small 1.5 percentage points. This is an order of magnitude smaller than the effects on the same outcome we observe in the smart platform interventions from 2017 and later. We observe similarly small effects when the sample is restricted to applicants with non-trivial application risk and when we look at each branch separately.

These findings provide further evidence that the information that smart platforms provide is a key reason that they are effective. In fact, it was the failure of these initial behavioral nudges that motivated us to pilot smart platforms in the 2017 cycle.

Impersonal Information.

The second type of alternate policy we consider is the provision of impersonalized information on application risk. The evidence we have presented so far shows that smart platforms work and that they shift applicants' beliefs about their admissions chances. However, it does not show that smart platforms are the *only* way to shift beliefs. It may be possible to obtain similar effects using approaches that do not require personalized messages, such as providing information about aggregate nonplacement rates. To the extent that misperceptions of own application risk are rooted in misperceptions of average risk, our theoretical and empirical analyses thus far predict that aggregate information interventions could also be effective.

To test the value of personalized relative to aggregate risk warnings, we conducted a supplemental WhatsApp RCT in the 2021 application process. Randomly selected applicants above the 0.3 risk cutoff received a personalized risk warning with text similar to the platform pop-up intervention. The key addition in the 2021 RCT is that randomly selected applicants in the 0.2 to 0.3 risk range received a message identical to the main treatment but with a warning about aggregate as opposed to personal risk.[21] We evaluate the effects of the aggregate information and smart platform treatments by comparing treatment and control groups within the relevant risk bins, and assess the effect of personalized relative to aggregate information by looking at the discontinuity at the cutoff within the treated group. A point of contrast with the 2020 RCT is that applicants in the 2021 RCT sample universe were selected from a subgroup that did not receive the platform pop-up, so the WhatsApp message was their first risk warning. See Appendix B.3 for details. Appendix Figure B.13 shows that predetermined covariates are balanced with respect to treatment.

Figure 2.7 reports three key results. First, we replicate the 2020 WhatsApp RCT finding that the smart platform warning causes applicants to lengthen their lists.[22] Second, we show that aggregate information also causes students to lengthen their applications, but that the effect is about half as big as the smart platform effect. Third, the RD comparison of aggregate to personalized information interventions at the 0.3 risk cutoff confirms that the aggregate information effect is about half the size of the smart platform effect. Findings from both interventions support the central claim that risk information shapes application choices, and the comparison between the two shows that personalization matters for policy efficacy.

[21] i.e., we targeted the aggregate treatment using personalized information. The goal was to avoid scaring low-risk applicants. Improved targeting is a benefit of smart platforms that we abstract from here.

[22] The behavioral effects of smart platforms in 2021 were roughly twice as big as in 2020, consistent with the ideas that a) the first risk warning changes behavior more than subsequent warnings, and b) messaging interventions can have large effects on choice behavior when the messages are well-formulated.

Costly Shoving.

The third type of alternate policy we consider is coercive nudges or "shoves" towards longer applications. These policies require students to list a certain number of schools on their application before they submit. Our costly search/limited information model predicts that shoves will produce low-quality matches. Applicants who are forced to add schools but believe they will be placed in a higher-ranked school may list schools they don't know much about, take up spots in those schools, and then decline their placements. This contrasts with smart platform interventions that make clear to students that added schools are welfare-relevant.

The distinction between coercive and search-inducing nudges is important. As described in section 2.3, the Chilean application system required many applicants to list at least two schools. Appendix Figure B.14 compares enrollment rates for students who applied to two schools on their initial application to enrollment rates for students who initially applied to one and were forced to add a second school. Conditional on being placed to the first-choice school, enrollment rates for the two groups are similar. However, students who were forced to add a second school are 17% (10pp) less likely to enroll in that school (if placed there) than students who added the second school voluntarily (and are placed in that school).[23]

These results contrast with findings from the smart platform intervention, where we see no difference in enrollment rates conditional on placement. The contrast supports our theoretical argument about the mechanisms underlying the smart platform intervention. Further, because declined placements can produce market congestion, these findings also provide an argument for the efficacy of smart platform policies relative to plausible alternatives.

2.7 Smart Matching Platforms in New Haven

In addition to our work in Chile, we partnered with the NGO and the New Haven, Connecticut school district to implement a warnings intervention during the 2020 choice process. The New Haven implementation of the smart platforms policy involved much smaller sample sizes than the Chilean implementation, but incorporated both smart platform and encouragement-focused nudge arms. It provides additional evidence on the cross-setting generalizability of smart platforms, and on the comparison between smart platform and behavioral nudges.

New Haven is a medium-sized school district that has used centralized choice since the mid-1990s. Starting in 2019, New Haven adopted a truncated deferred acceptance assignment mechanism. See Kapor et al. (2020) and Akbarpour et al. (2020) for institutional details.

The warnings policy in New Haven was similar in broad strokes to the policies implemented in Chile. The application window opened at the end of January, with a deadline of March 2. Seven days before the deadline, the district identified applications with a non-placement risk of greater than 50% as risky. Application risk was computed using data from

[23] "Placeholder" schools show up in other choice contexts. In Ghana, 20% of students add repeat or non-existent programs to satisfy length requirements (Ajayi & Sidibe 2021; correspondence with Modibo Sidibe).

the previous year.[24] All applicants identified as risky received an email stating they were at risk of non-placement. The email included a link to a website where they could input hypothetical applications and view the chances of admission at each school, again based on the previous year's data.[25]

The New Haven policy differed from the Chile policy in two important ways. The first is that, in addition to warning all risky applicants, the district selected a randomly chosen fifty percent of non-risky applicants to receive an email that provided a recommendation to learn more about admissions chances by visiting the same application simulator website. This encouragement nudge intervention did not include information on application risk. The second contrast is sample size: in Chile, 233,768 students received a warning about a risky application. In New Haven, the number was 740. This reduces statistical precision substantially.

Figure 2.8 presents a visual summary of our findings in New Haven. These graphs plot the rate at which students make different kinds of application changes in each ten percentage point bin of the predicted risk distribution, with additional bins for risk values of zero and one. We display these statistics for 2020 applicants, who received a warning email when predicted risk was 50% or higher, and for a comparison group of 2019 applicants, who did not receive a warning regardless of risk score.[26] For non-risky applicants in 2020, the graphs split out the set of applicants who received the encouragement prompt from those who were not contacted.

Panel A of Figure 2.8 shows results for application modification. Rates of application modification for low-risk applicants were similar in 2019 and 2020. In 2020, we observe a large jump in rates of modification at the 50% cutoff for the warning treatment, with no similar increase for the 2019 comparison group. As shown in Panel B, almost all of these changes involve lengthening the application. As shown in Panel C, the effect of these additions is to reduce application risk. RD estimates reported in Appendix B.11 show that crossing the threshold causes 13.8% of applicants to add at least one school to their application, and that compliers with the warnings policy reduce their application risk by 23.2 percentage points, or 42% of below-threshold mean ex post risk. The encouragement nudge does not affect search in any panel: the nudge and no contact series track each other at all tested values of risk.

These findings add to our findings from Chile in two ways. First, they show that information on admissions chances is an important input to choice behavior in a variety of contexts. Second, they provide further evidence that the "smart" part of smart matching platforms is important to their efficacy at expanding search.

[24]The district focused on major choice grades, where choice probabilities are relatively stable across years. Two schools opened in 2020. Risk scores were not computed for applicants to these schools.

[25]See Appendix B.11 for a detailed description of the intervention procedures in New Haven, the distribution of application risk, and the relationship between our risk simulations and realized application risk.

[26]We compute predicted risk for 2019 applicants using a snapshot of predicted risk status as of seven days prior to the admissions deadline. This procedure parallels our approach to identifying risky applicants in 2020.

To conclude our discussion of generalizability across settings we highlight a simple statistic. In both Chile and New Haven, we conducted surveys asking applicants what information would be helpful in filling out their applications. Roughly 90% of respondents in both settings said they needed information on admissions chances. See Appendix Figure B.15 for details.

2.8 Conclusion

This paper shows that beliefs about admissions chances are a key determinant of the way applicants search for schools in centralized choice systems, that optimism about school placement chances leads applicants to search too little, and that the smart matching platforms that build live feedback on application risk into the choice system increase search, reduce non-placement risk, and help students enroll in better schools.

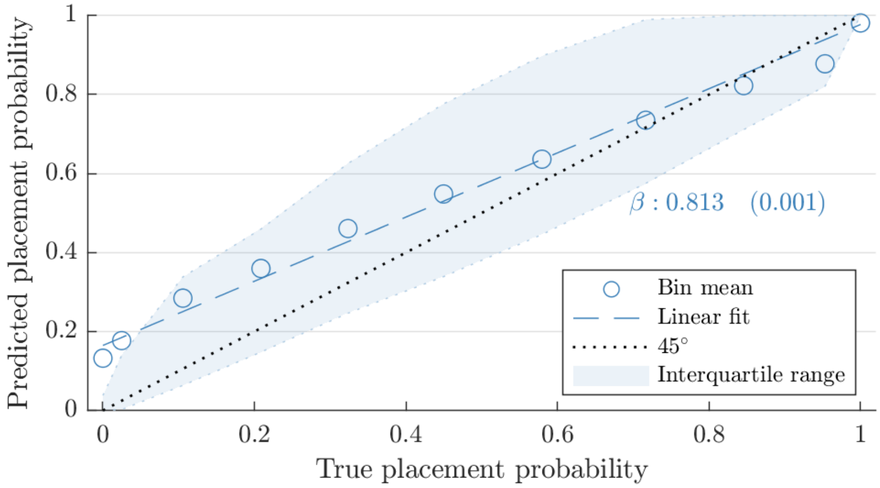
The main implication of our findings is that policymakers seeking to reduce the burden school choice places on participants need *both* to choose a strategyproof assignment mechanism *and* to provide choice supports that aid with search for schools. The strategic challenges posed by school search are central to families' experiences of school choice even when the centralized assignment mechanism is strategyproof.

The smart matching platforms we propose and evaluate in this paper are an effective and generalizable approach to reducing the burden of school search. Critically, smart platforms are not researcher-driven proofs-of-concept, which often decline in effectiveness when taken to scale (DellaVigna and Linos, ming). They are *products* already at scale. At the time of this writing, policymakers in Brazil, Peru, and Ecuador are in the process of implementing the techniques we discuss in this paper. The close collaboration between researchers, policymakers, and implementation partners that made this work possible may be a useful approach for conducting scalable interventions in other domains.

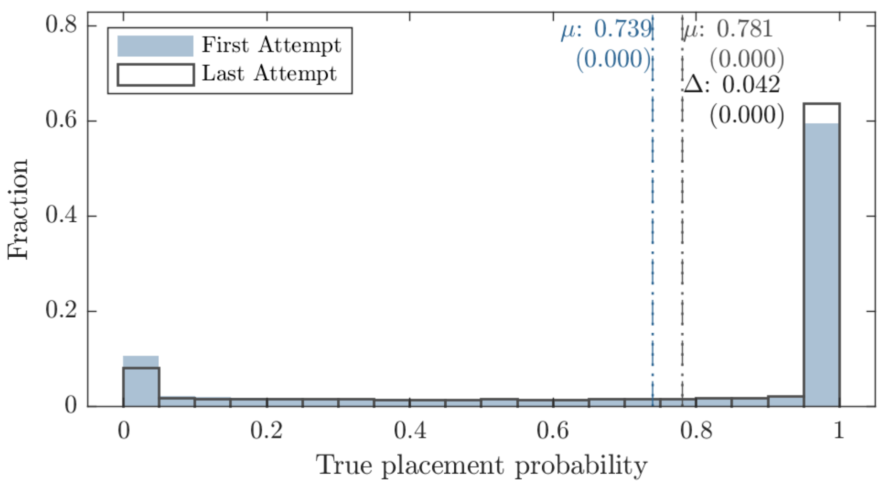
Figures and Tables

Figure 2.1: Distribution of Placement Probabilities and Probability Predictions

(a) Predicted vs. True Placement Probabilities



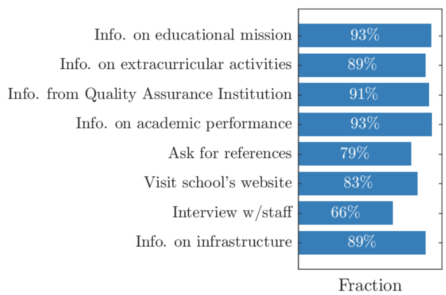
(b) Histogram of True Placement Probabilities



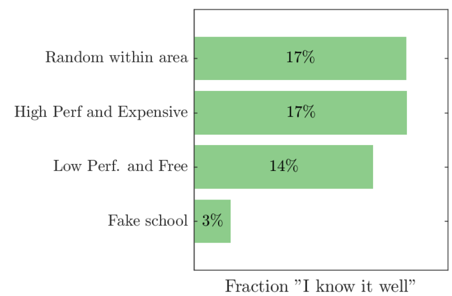
Notes. Panel A: binned means, linear fit and interquartile range of predicted placement probability by true placement probability. Points are centered means of 10 quantile-spaced bins of the support of the true placement probability $\in [0.00; 0.99]$. The last point at the right represents the mean of predicted placement probability for observations with true probability greater than 0.99. Placement predictions in Panel A combine observed applications at the time an individual submits her application with historical projections. See section 2.3 for details. Panel B: histogram of true placement probability for initial application attempt and final application submission. Vertical lines display means.

Figure 2.2: Knowledge of and Search for Schools

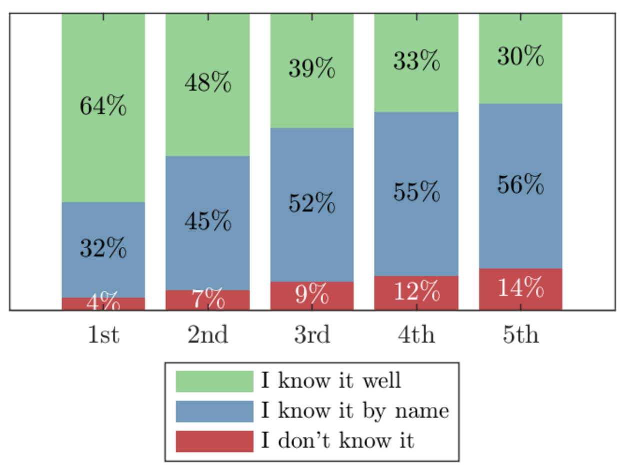
(a) Relevant Steps to Know a School



(b) Knowledge of Schools Not on Application List



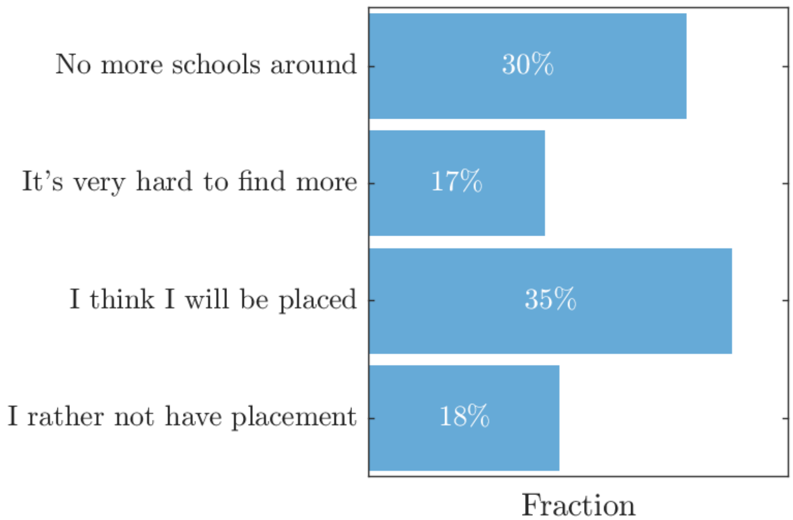
(c) Knowledge of Schools on Application List



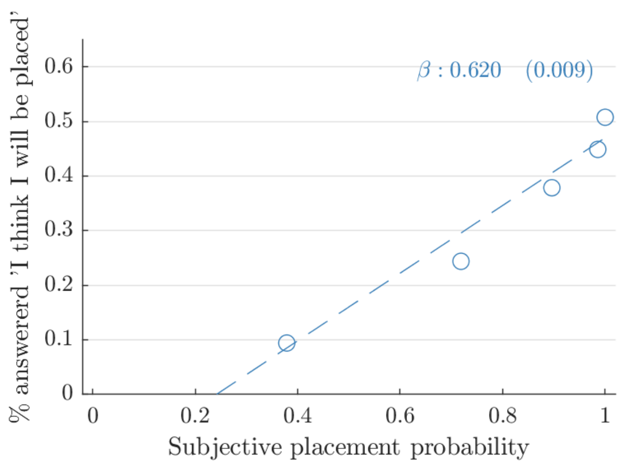
Notes. Panel A: share of survey respondents stating an understanding of listed attribute was relevant for "know[ing] a school well." "Info on educational mission" is refers to qualitative information on a school's educational goals and approach; schools are required to report this information to Mineduc and it is posted online. "Info from Quality Assurance Institution" is information on academic performance and other indicators not related to standardized tests from education regulators in charge of the evaluation of schools. Panel B: share of students stating that they "know well" schools not listed on their application, for schools of type listed on horizontal axis. All schools are within an applicant's local area, defined as 2km from student's location (home address reported on platform, replaced with centroid of application if geocoding was unreliable). "High performing and expensive schools" are those classified in 2 best tiers of performance (out of 4) by the Quality Assurance Institution, with a monthly copayment of $35 USD or more. "Low performing and free" schools are defined as schools within the worst tier of performance, with no copayment. "Fake schools" are schools that do not exist in the student's local area. Panel C: stated knowledge of schools on application list, by rank. See section 2.4 for details.

Figure 2.3: Reasons for Stopping School Search

(a) Stated Reason for Not Adding More Schools



(b) Stated Reason is "I Will Be Placed" vs. Declared Risk



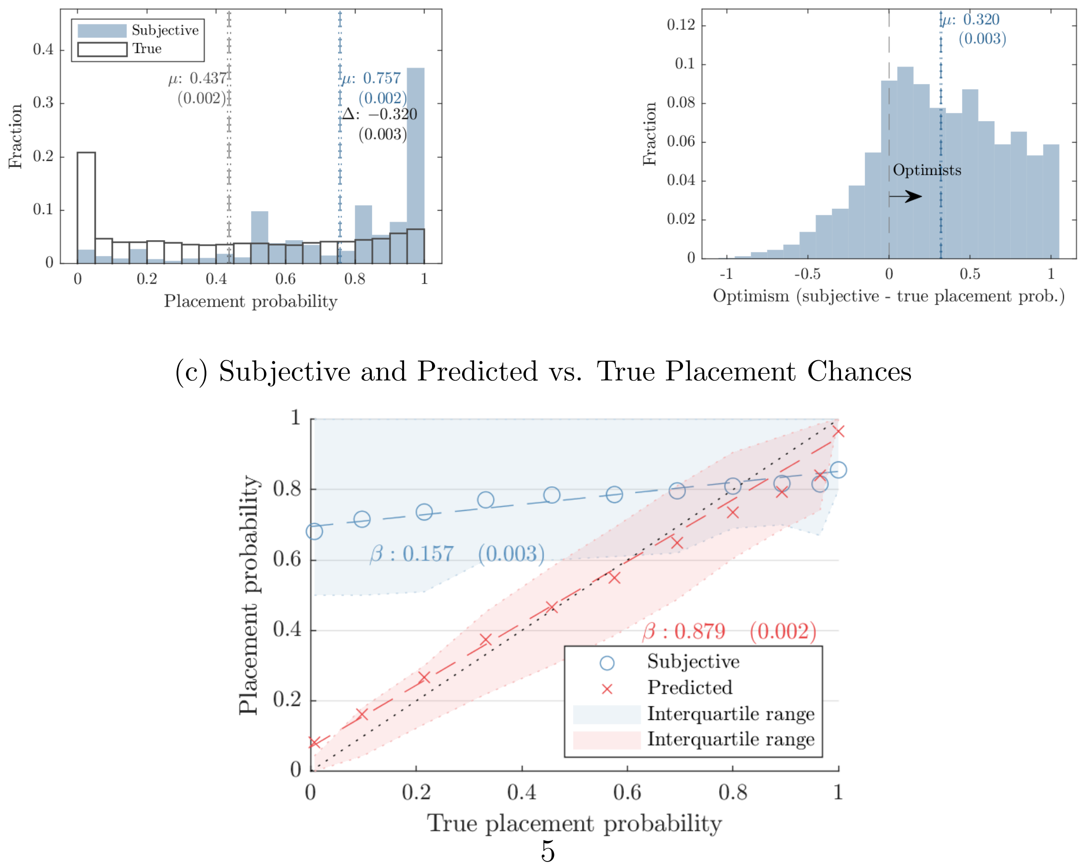
Notes. Panel A: survey reports of reason for not adding more schools to the choice application. Panel B: share of survey completers stating that they stopped search because they think they will be placed, by survey report of subjective placement probability. Sample in both panels: survey completers.

Figure 2.4: Subjective vs. Observed Application Risk

(a) Distributions of Placement Chances

(b) Optimism

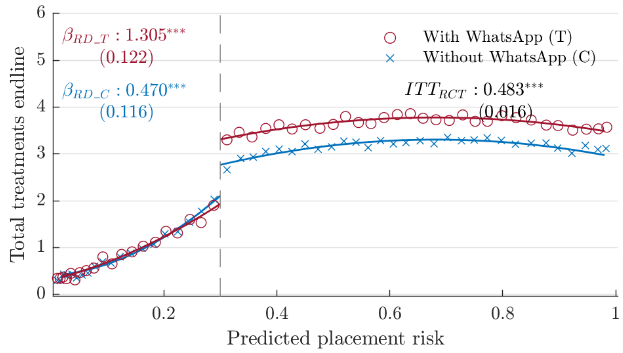
(c) Subjective and Predicted vs. True Placement Chances



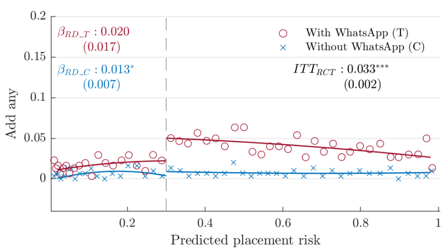
Notes. Panel A: distribution of true placement chances and survey-reported subjective placement chances. Vertical lines display means of each distribution. Panel B: distribution of optimism, defined as difference between subjective and true placement chances. Panel C: mean subjective placement belief within bins defined by true placement probability. The bottom bin includes applicants with placement probability less than 1%, and the top bin includes applicants with placement probability of 99% or more. The middle eight groups split the remaining observations into equally-sized bins. Dashed line is linear fit. Shaded areas are IQRs for subjective beliefs and risk predictions (within survey sample). 45-degree line displayed for reference. Sample: survey completers.

Figure 2.6: WhatsApp RCT Outcomes

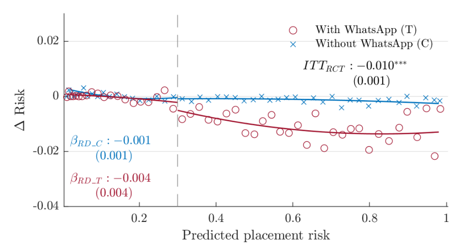
(a) Count of Feedback Messages Received



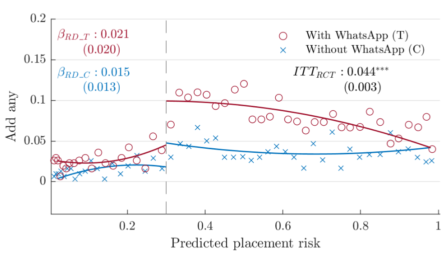
(b) Add at Least One School – 44 Hours



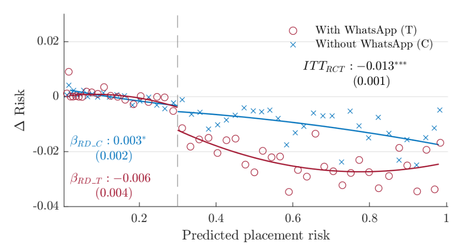
(c) Change in Risk – 44 Hours



(d) Add at Least One School – Endline

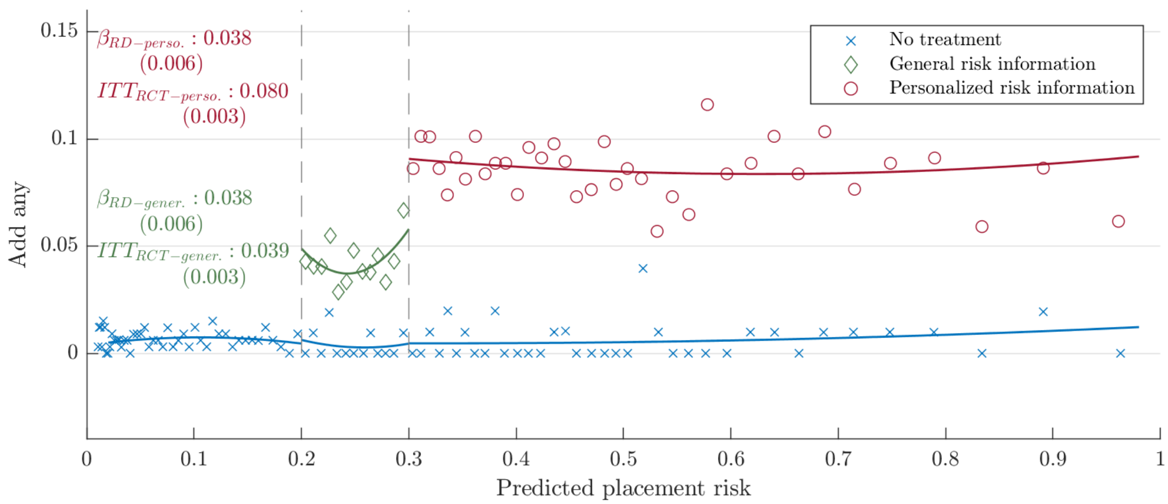


(e) Change in Risk – Endline



Notes. Binned means and global fits of message receipt, application behavior, and risk outcomes by predicted placement risk in RCT sample. Points are centered means of 50 quantile-spaced bins of the support of the predicted placement risk $\in [0.02; 0.98]$. Solid lines show the quadratic fit. Figures split by RCT treatment and control group, above and below treatment threshold. "With WhatsApp" group receives WhatsApp warning when above cutoff. "Without WhatsApp" group receives no warning. Below 0.30 predicted risk cutoff, the treatment group receives WhatsApp message with no risk-related information. Reported β_{RD} coefficients are RD estimates within treatment and control group, computed from local linear specifications using $+-0.1$ bandwidth. See section 2.5 for details. Because coefficients are local while polynomial fits are global, there may be minor differences between displayed fits and reported coefficients.. Reported ITT_{RCT} estimate is the experimental RCT effect for all above-cutoff students on the listed outcome. Outcomes, listed in panel titles, are as follows. Panel A: count of warnings messages received over full application period. Panel B: add any school in 44-hour window between WhatsApp message and SMS followup. Panel C: change in risk within 44-hour window between WhatsApp message and application followup. Panel D: add any school between WhatsApp message and application close. Panel E: change in risk by application close. See section 2.5 for details.

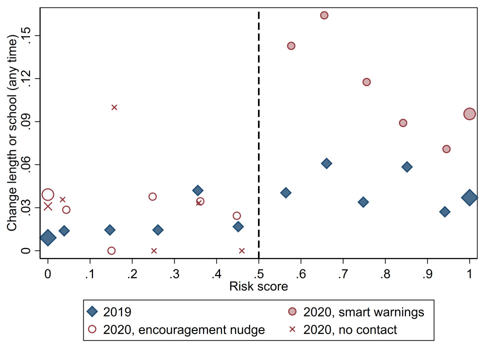
Figure 2.7: 2021 WhatsApp RCT with Personalized and Aggregate Information Treatments



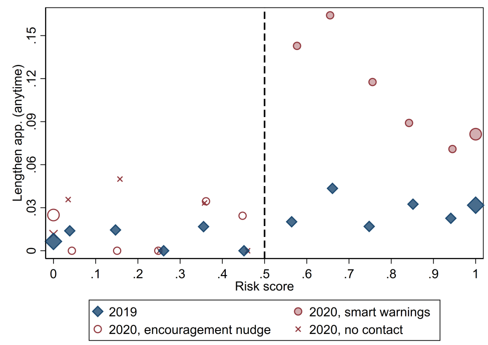
Notes. Results from 2021 WhatsApp RCT. Outcome is adding any school to the choice application. Treatments are as follows. "No treatment": control group that receives no WhatsApp message. "General risk information": treatment group that receives information about nonplacement risk in aggregate, not personalized to own application. "Personalized risk information": treatment group that receives information about own application risk, as in 2020 WhatsApp RCT. $\beta_{RD-general}$ is the RD estimate of general risk treament group against the control group at the 0.2 cutoff. $\beta_{RD-personal}$ is the RD estimate of the personalized risk information treatment group relative to the general risk treatment group. $ITT_{RCT-personal}$ and $ITT_{RCT-general}$ are RCT estimates of treatment effects for the personal and general info treatments (respectively) relative to the control group in the same risk range. See section 2.6 and Appendix B.3 for design details and additional results. Reported RD coefficients and standard errors are from local linear specifications using $+-0.1$ bandwidth. See section 2.5 for details.

Figure 2.8: Smart Warnings and Encouragement Nudges in New Haven

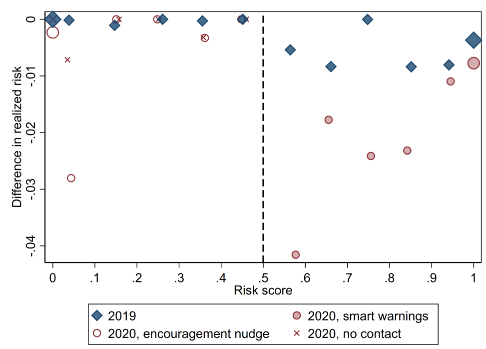
(a) Modify Application



(b) Add School to Application

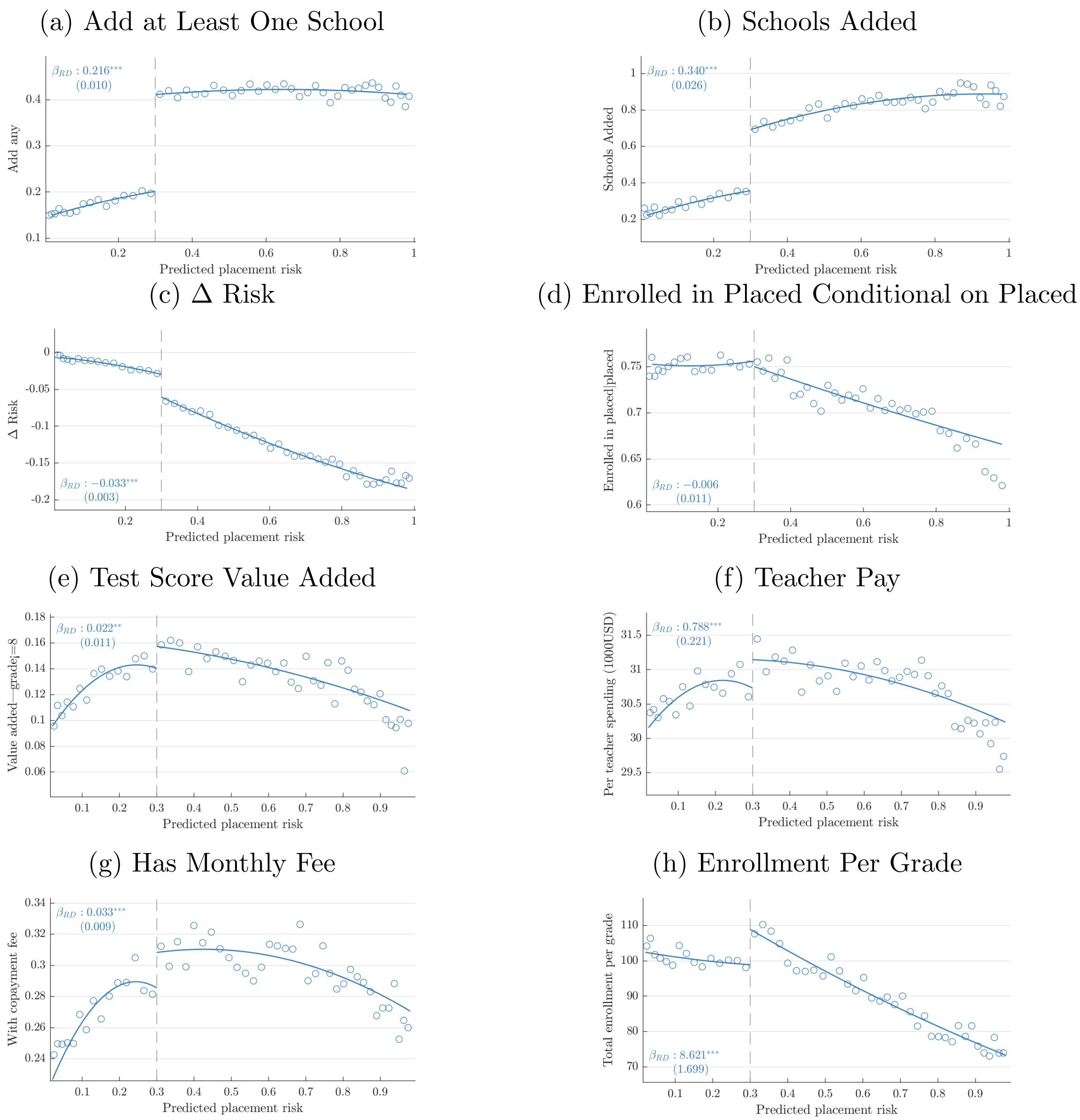


(c) Change Nonplacement Risk



Notes. Outcomes of warnings intervention in New Haven centralized choice system. Figures show changes in application behavior by risk score as of 7 days prior to application deadline in 2019 and 2020. Points are centered binned means within intervals of width 0.1, except for top- and bottom-most points, which are for students with risk scores of 1 and 0, respectively. In 2020, all applicants with risk scores above 0.5 received the warnings intervention. Randomly chosen applicants with risk scores below 0.5 received an encouragement nudge (a non-personalized message encouraging them to learn more about their assignment chances); the remaining non-risky applicants received no intervention. In 2019, no applicant received any intervention. Panel A: any change in application. Panel B: lengthen application. Panel C: change in risk from initial to final portfolio. See section 2.7 for details.

Figure 2.5: Choice Behavior and Enrollment Outcomes in the Platform Pop-Up RD



Notes. Binned means and global fits of choice outcomes by predicted risk for initial application. Points are centered means of 50 quantile-spaced bins of the support of the predicted placement risk $\in [0.02; 0.98]$. Solid line shows the quadratic fit. Reported coefficients and standard errors are from local linear specifications using $+-0.1$ bandwidth. See section 2.5 for details. Because coefficients are local while polynomial fits are global, there may be minor differences between displayed fits and reported coefficients. Outcomes by panel are as follows. Panel A: add at least one school to application. Panel B: count of schools added. Panel C: change in risk from initial to final application. Panel D: Enroll in placed school conditional on placement. Panel E: value added at enrolled school. Panel F: teacher compensation at enrolled school. Panel G: indicator for monthly fee at enrolled school. Panel H: students per grade at enrolled school.

Table 2.1: Descriptive Statistics for Chilean Choice Applicants

	(1) All	(2) Economically Vulnerable	(3) Not Economically Vulnerable	(4) Pop-up eligible	(5) Risky (predicted risk>.3)
N	1,168,706	575,521	593,185	848,795	233,678
%	1.00	0.49	0.51	0.73	0.20
A. Demographics					
Economically Vulnerable	0.49	1.00	0.00	0.51	0.37
Rural	0.05	0.07	0.03	0.06	0.02
B. Application behavior					
Length initial attempt	2.77	2.61	2.93	2.70	2.36
Length final attempt	3.14	2.92	3.36	3.06	3.20
Total attempts	1.41	1.35	1.46	1.38	1.74
Any modification	0.25	0.22	0.27	0.24	0.43
Add any	0.21	0.19	0.23	0.21	0.41
C. Placement					
Placed in pref.	0.79	0.84	0.74	0.80	0.45
Placed 1st	0.54	0.61	0.47	0.56	0.18
Particip. in 2nd round	0.09	0.08	0.10	0.08	0.15
Placed in 2nd round	0.07	0.06	0.07	0.06	0.11
D. School capacity available after placement (at local market level defined for each student)					
Share of total seats	0.42	0.41	0.42	0.42	0.50
Share of seats in free schools	0.46	0.45	0.47	0.47	0.55
E. Attributes of enrolled school					
Enrolled at some school	0.97	0.98	0.96	0.97	0.95
Enrolled at placed	0.62	0.66	0.57	0.63	0.31
Have value added measure$\mid$grade$\leq$8	0.77	0.76	0.78	0.75	0.77
Value added$\mid$enrolled at placed	0.11	0.06	0.14	0.10	0.20
Value added$\mid$not enrolled at placed	0.04	-0.03	0.10	0.04	0.08
School monthly fee (USD)	17.02	10.20	24.05	15.14	24.25
Share of vulnerable students	0.61	0.66	0.56	0.62	0.56
F. Classification by true risk of initial attempt					
Mean risk	0.24	0.18	0.30	0.23	0.69
Zero risk	0.59	0.67	0.51	0.62	0.05
Risky (risk>.3)	0.30	0.23	0.37	0.29	0.86

Notes. N: 1,168,706 (20% from 2018, 41% from 2019 and 39% from 2020). All statistics are means in the population defined by the column header. "Pop-up eligible" (col. 4) are students who submitted applications that received a risk prediction. "Risky" (col. 5) is applicants whose first attempt had a predicted risks > 0.3. Selected row variable definitions are as follows. "Economically vulnerable" is an SES measure computed by Mineduc. "Rural" is an indicator if students live in rural areas. "Length of initial/final attempt" is the number of schools on an applicants first and final choice application. "Total attempts" is the number of times an applicant submitted an application to the centralized system. Application change and addition variables describe the share of applicants making different kinds of changes applicants make between their first and final submission. "Placed in pref/1st" are indicators for any placement or for placement in the school ranked 1st. "2nd round" variables describe participation and placement outcomes in the second centralized placement round. "Share of total seats/seats in free schools" is the share of seats in all schools/in schools without fees unfilled after the first application round in a student's local market. Value added and school characteristic variables described in Online Appendix B.4. VA is calculated only for grades 8 and below. True risk of initial attempt variables describe the nonplacement risk of an applicant's initial application, evaluated using ex post observed applications.

Table 2.2: RD Estimates of Platform Pop-Up Effects

	(1)	(2)	(3)	(4)	(5)
	All		2018	2019	2020
		IV			
A. Balance					
Economically Vulnerable	-0.004		-0.014	0.016	-0.012
	(0.010)		(0.029)	(0.018)	(0.013)
Rural	-0.007		-0.002	-0.009	-0.008
	(0.003)		(0.007)	(0.005)	(0.003)
B. Choice Behavior					
Any modification	0.214		0.164	0.217	0.224
	(0.010)		(0.025)	(0.018)	(0.013)
Add any	0.216		0.176	0.224	0.223
	(0.010)		(0.024)	(0.018)	(0.013)
Schools Added	0.340	1.573	0.379	0.317	0.344
	(0.026)	(0.090)	(0.068)	(0.050)	(0.033)
Δ Risk	-0.033	-0.155	-0.039	-0.040	-0.029
	(0.003)	(0.013)	(0.009)	(0.007)	(0.004)
Add as first	-0.003	-0.012	-0.007	-0.005	-0.000
	(0.003)	(0.013)	(0.008)	(0.005)	(0.003)
Add to middle	0.017	0.078	0.017	0.023	0.014
	(0.004)	(0.018)	(0.012)	(0.007)	(0.005)
Add as last	0.205	0.949	0.172	0.207	0.213
	(0.009)	(0.018)	(0.023)	(0.017)	(0.012)
Drop any	-0.001	-0.003	-0.009	0.018	-0.008
	(0.004)	(0.019)	(0.010)	(0.008)	(0.005)
Re-order	0.014	0.065	0.026	0.005	0.015
	(0.005)	(0.022)	(0.013)	(0.009)	(0.006)
C. Choice outcome					
Placed to preference	0.038	0.178	0.033	0.086	0.020
	(0.009)	(0.041)	(0.026)	(0.018)	(0.011)
Enrolled in placed	0.024	0.113	0.008	0.055	0.018
	(0.010)	(0.049)	(0.029)	(0.020)	(0.013)
Enrolled in placed|placed	-0.006	-0.025	-0.021	-0.009	0.003
	(0.011)	(0.045)	(0.031)	(0.022)	(0.013)
D. Congestion-related outcomes					
Add any undersubscribed	0.073	0.339	0.052	0.081	0.075
	(0.007)	(0.026)	(0.016)	(0.012)	(0.009)
Δ prob. placed to undersubscribed	0.019	0.088	0.015	0.032	0.014
	(0.003)	(0.014)	(0.008)	(0.007)	(0.004)
NL	20,359	20,359	2,834	6,076	11,449
NR	21,145	21,145	2,776	6,015	12,354

Notes. Local linear RD estimates of pop-up effects from warning pop-up on application platform. Computed using triangular kernel with bandwidth 0.1. Heteroskedasticity-robust nearest neighbor variance estimator with minimum of 3 neighbors reported in parentheses; computed as in Calonico et al. (2014). We report estimates in the pooled sample and for each year. IV (column 2) shows the instrumental variable specifications, where the endogenous regressor is the add any school indicator. Panel A: predetermined covariates. Panel B: measures of choice behavior from initial to final application. Δ risk is change in application risk from first to final attempt. "Add to X" are additions of schools in given place on list, relative to initial application submission. Panel C: outcomes of choice process. "Enrolled in placed" is equal to one for students who receive a placement and enroll in the placed school. "Enrolled in placed | placed" is the enrollment rate in the placed school, conditional on receiving a placement. Panel D: congestion attributes of behavior and placement outcomes. "Undersubscribed" schools are those with excess capacity.

Table 2.3: RD Estimates of Platform Pop-Up Effects on Enrolled School Outcomes

	(1) Pooled	(2) Pooled IV	(3) $E[Y\|X=0.3^{-}]$
A. First stage and enrollment			
Add any	0.216		0.199
	(0.010)		
Enrolled	-0.004		0.966
	(0.004)		
Have value added measure|grade≤8	0.014		0.753
	(0.010)		
B. Attributes of enrolled school			
Distance (km)	0.050	0.239	3.022
	(0.243)	(1.158)	
Value added|grade≤8	0.022	0.103	0.138
	(0.011)	(0.051)	
Per teacher spending (1000USD)	0.788	3.714	30.646
	(0.221)	(1.065)	
Per student spending (1000USD)	0.002	0.007	2.245
	(0.015)	(0.071)	
With copayment fee	0.033	0.152	0.279
	(0.009)	(0.044)	
School monthly fee (USD)	2.016	9.237	21.839
	(0.815)	(3.778)	
Share of vulnerable students	-0.006	-0.029	0.567
	(0.003)	(0.013)	
Total enrollment per grade	8.621	39.467	98.981
	(1.699)	(7.964)	
NL	19,550	19,550	
NR	20,222	20,222	

Notes. Local linear RD estimates of platform pop-up effects. Computed using triangular kernel with bandwidth 0.1. Heteroskedasticity-robust nearest neighbor variance estimator with minimum of 3 neighbors reported in parentheses; computed as in Calonico et al. (2014). IV estimates in the second column report instrumental variable specifications where the endogenous regressor is the “add any school” indicator. The third column reports below-cutoff means of the variable listed in the row in the analysis sample. Sample for value added outcomes is restricted to grades eight and below. Reported sample sizes are counts of enrolling students. See section 2.5 for discussion and Online Appendix B.4 for detailed variable definitions

Table 2.4: WhatsApp RD and RCT Results – 2020

	(1)	(2)	(3)	(4)
	RCT		RD	
	ITT	IV	ITT	IV
A. Balance				
Economically Vulnerable	-0.019		-0.012	
	(0.006)		(0.039)	
B. Message receipt				
WhatsApp read	0.466		0.528	
	(0.005)		(0.030)	
SMS reminder received	-0.028		0.459	
	(0.004)		(0.034)	
Total treatments before final SMS	0.506		0.845	
	(0.016)		(0.116)	
Total treatments endline	0.483		1.305	
	(0.016)		(0.122)	
C. Outcomes in clean 44 hours before SMS followup				
Any modification	0.035		0.015	
	(0.002)		(0.017)	
Add any	0.033		0.020	
	(0.002)		(0.017)	
Schools Added	0.075	2.281	0.103	5.260
	(0.007)	(0.136)	(0.042)	(3.194)
Δ Risk	-0.010	-0.297	-0.004	-0.209
	(0.001)	(0.018)	(0.004)	(0.131)
D. Endline outcomes				
Any modification	0.046		0.012	
	(0.004)		(0.021)	
Add any	0.044		0.021	
	(0.003)		(0.020)	
Schools Added	0.112	2.550	0.138	6.681
	(0.011)	(0.175)	(0.065)	(4.764)
Δ Risk	-0.013	-0.301	-0.006	-0.307
	(0.001)	(0.022)	(0.004)	(0.206)

Notes. ITT and IV effects of 2020 WhatsApp warnings intervention. RCT columns: effects of random assignment to treatment group vs. control group for students with predicted risk > 0.30. Robust SEs in parentheses. N=17,970. RD columns: regression discontinuity evaluation in treatment group around 0.30 cutoff. RD specifications computed using local linear fit with a bandwidth of 0.1. Standard errors are heteroskedasticity-robust nearest neighbor variance estimator with minimum of 3 neighbors, as in Calonico et al. (2014). ITT column shows effects of group assignment. IV columns show the instrumental variable specification, where the endogenous regressor is the add any school indicator, instrumented with group assignment for the RCT, and with a dummy of crossing the risky threshold for the RD. Panel A: balance tests on predetermined characteristics. Panel B: message receipt. "WhatsApp read" is an indicator equal to one if applicant views the WhatsApp treatment message. "SMS remainder received" is indicator for receiving SMS reminder 44 hours later. Panel C: outcomes within 44 hour window between WhatsApp intervention and followup SMS. Panel D: endline choice behavior and placement outcomes. See section 2.5 for details.

Table 2.5: RD Estimates of Platform Pop-Up Effects on Subjective Beliefs

	(1) 2020	(2) $E[Y\|X = 0.3^{-}]$
A. Survey takeup and completion		
Survey take-up	-0.020	0.173
	(0.010)	
Answered expectation questions	-0.013	0.150
	(0.010)	
B. Application behavior in survey sample		
Add any	0.196	0.265
	(0.033)	
Δ Risk	-0.016	-0.027
	(0.008)	
C. Subjective beliefs		
Subjective P(not assigned to any)	0.036	0.165
	(0.017)	
Subjective P(assigned to 1st)	-0.049	0.754
	(0.021)	
D. Stated preferences		
Satisfaction if assigned to 1st choice (1–7)	-0.017	6.855
	(0.047)	
NL	1,381	
NR	1,500	

Notes. Local linear RD estimates of platform pop-up effects on survey reported subjective beliefs. Computed using triangular kernel with bandwidth 0.1. Heteroskedasticity-robust nearest neighbor variance estimator with minimum of 3 neighbors reported in parentheses; computed as in Calonico et al. (2014). The second column reports the below-cutoff means of the row variables. Panels B and C restrict the sample to applicants who completed the beliefs module of the 2020 survey. See section 2.6 for details.

Table 2.6: RCT Estimates of Behavioral Nudge Effects

	(1)	(2)	(3)	(4)	(5)
	Pooled	Risk>0.01	By message type		
			More schools, higher chances	Range suggestion	Role model
Add any	0.015	0.027	0.006	0.008	0.031
	(0.009)	(0.018)	(0.012)	(0.012)	(0.013)
N Treatment	1,402	479	463	455	484
N Control	648	215	648	648	648

Notes. RCT effect estimates for behavioral nudge interventions conducted as part of the 2016 choice process. These interventions encouraged people to add schools to their lists but did not include information on nonplacement risk. The sample is limited to the Puntarenas region, which was the only region with centralized choice in 2016. Estimates are differences in the share of students adding any school to their baseline application between the treatment group and a control group that did not get any message. Columns 1 and 2 are pooled estimates of the treatments from columns 3-5. Column 2 limits the sample to applicants facing non-zero application risk. See section 2.6 and Online Appendix B.3 for details.

Chapter 3

Can Information on School Attributes and Placement Probabilities Direct Search and Choice?

3.1 Introduction[1]

Platforms that allow users to search through and select from a variety of options are common tools for commercial goods, but are also used for services like health and education. Just as Google and Amazon carefully manage the information provided to users and generate value for consumers and sellers, school choice platforms provide these services for parents, students, and educational communities. In fact, centralized school choice and assignment platforms facilitate access to educational opportunities for students and families in over 50 countries worldwide (Neilson, 2019). Despite their popularity, little research has investigated how their design and the information they provide on options or the application itself affects the ensuing matches made by the centralized system.

Recent studies show that many users interact with these search and choice platforms with limited information and biased beliefs, which leads to inefficiencies and, in some cases, inequities in the resulting placements.[2] Arteaga et al. (2022), for example, provide evidence that the search for schools is costly, and that applicants exhibit biased beliefs on placement probabilities and list too few options as a result. They also show that correcting these beliefs by providing additional information leads applicants to expand their list of schools,

[1]This chapter was published as a working paper in 2022 (Arteaga et al., 2022), and was coauthored with Gregory Elacqua, Thomas Krussig, Carolina Méndez, and Christopher A. Neilson.

[2]Kapor et al. (2020) document that users have heterogeneous assumptions about assignment probabilities that can drive decision-making behavior, noting that many families engage in strategic behavior based on biased beliefs. Allende et al. (2019) provides evidence of limited awareness of the set of potential options and finds that giving information on the existence and characteristics of nearby choices can result in changes to users' applications and assignments. Their experiment shifted students' decisions toward schools farther from their homes with higher average test scores, higher added value, and higher prices.

suggesting that users made additional efforts to search for more schools. Still, there is no evidence on how giving information about available options might affect the school choice process in this centralized choice context.

This paper studies the role of information about available options on school applications in a centralized school choice platform. We partnered with the Ministries of Education (MoE) in Ecuador and Peru to send "report cards" to applicants during the pilot phase of centralized school choice and assignment platforms.[3] These report cards were given to a randomly selected group and included a list of suggested schools. We study the effect of this intervention on the users' subsequent application and enrollment decisions. We also collected survey data to evaluate the effect of our intervention on subjective measures.

The intervention proposed is based on previous work revealing that giving information can improve the application process. Based on a model of the costly search for schools, Arteaga et al. (2022) show that over-optimism about placement chances can lead to insufficient searching for options, potentially reducing the probability of finding a placement. Furthermore, the authors provide survey evidence that applicants have biased beliefs about their chances of finding a placement, incomplete information about their options, and a belief that searching for schools is costly. Their model motivates our intervention through the theoretical channel of lowering application costs.

The information we gathered on parent beliefs shows that the biases studied in Arteaga et al. (2022) are present in our two contexts. The online survey distributed to the chosen participants echoed their results in three respects. First, applicants have limited information about their options, and learning about new schools is costly. Second, their beliefs on their chances of admission inform their application strategy, and those beliefs are overly optimistic when we compare them with actual placement probability. Third, non-placement is much less desirable than being placed in their lowest-ranked (or any) option, which suggests large welfare stakes at play.

We tested the effects of information about the available options and replicated the feedback mechanism developed in Arteaga et al. (2022), who document how providing applicants in Chile and New Haven with feedback on their chances of admission helped them to search more effectively and ultimately increased their placement chances. We also provided the treatment group with additional information on schooling options.

We designed three report cards to test the effect of different levels of detail. The first report card only included the current application, a warning about the placement risk, and a general recommendation to add more schools. On the second report card, we added a personalized list of 10 schools that the student did not consider in their initial portfolio. The third report card differed between the 2021 and 2022 implementations: in 2021, we included a personalized list of 10 schools plus information about the popularity and congestion of each school, the most comprehensive report card we delivered. In 2022, this report card only

[3] The Inter-American Development Bank (IADB) and the NGO ConsiliumBots provided technical support for the information interventions, as well as for the design and implementation of the centralized school choice and assignment pilots.

included three schools with no extra information on popularity or congestion. In trying more than one version, we aim to examine the trade-off between more information and too much information, an issue that has received little attention in the literature (Gabaix, 2019).

We randomly assigned applicants to one of the three treatment arms. Between 4 and 7 days before the last day of the application process, we identified the students with a positive non-placement probability —i.e., risky applicants— and sent them a link to the report card by email and through WhatsApp. In Peru, we implemented the intervention during the 2021 and 2022 intake years, and applicants also received an additional WhatsApp message with a non-placement warning. In Ecuador, the intervention was conducted only during the 2021 admission process, and we did not send the non-placement warning via WhatsApp.

Our results show that applicants that received the treatment with school suggestions were more likely to add those schools compared to applicants who received the report card with no suggestion list. They were also more likely to add *more* schools to the list. All report cards included a non-placement warning but varied in the suggestions and the extra information on popularity and congestion. The RCT design thus allowed us to estimate the causal effect of the additional information, since we generated a list of suggested schools for every applicant, but did not show it to students who received the basic report card. In Peru, the proportion of students that add a school from the suggestion list increases between 52% and 120%, depending on the year and treatment. Meanwhile, adding additional information on popularity and congestion does not affect the probability of adding a school. For the Ecuadorian context, we cannot rule out a zero effect on the shifting of preferences.

Our study contributes to the literature on information provision policies in educational markets (Allende et al., 2019; Hastings and Weinstein, 2008; Andrabi et al., 2017) by integrating information within the centralized school choice process, and testing new channels that can potentially help to distribute the information at scale. We also build on an emerging strand of empirical market design work focused on educational markets[4] (Arteaga et al., 2022; Kapor et al., 2020; Ajayi and Sidibe, 2020) by assessing how new information can affect search in a context with incomplete information about all the options.

The remainder of the paper is organized as follows. Section 3.2 describes the Ecuadorian and Peruvian schooling context and provides details on the intervention, sample, and survey design. Section 3.3 discusses the results of the post-application survey. In Section 3.5, we present the findings from the information intervention on choice behavior and beliefs. Finally, Section 3.6 concludes.

3.2 Setting

We study the effect of information provision in the regions of Manta, Ecuador and Tacna, Peru, both of which were implementing a centralized school choice system for the first time. These pilots offered a unique opportunity to test the same policy design in different contexts.

[4]See Agarwal and Budish (2021) for a recent review.

In both pilots, parents applied to schools using an online platform, and the educational authorities then assigned students to schools using a deferred acceptance (DA) assignment mechanism. We applied similar information treatments in both pilots.

We were granted access to the applications and enrollment outcomes, and complemented our dataset with a parent-participant survey. We can therefore observe the universe of applicants, the history of their applications, and information related to the options available on the platform as well as off-platform alternatives. We also have information on the family's final enrollment decision for the 2021 academic year.

The IADB supported the pilots in both countries, but their origins were quite different. In Ecuador, the government wanted to introduce parental choice to improve efficiency and equity in school access. The country had previously used a centralized system that assigned students to schools based on the applicant's location, which parents reported through their electricity account code (CUEN). This process was costly and time-consuming as it required considerable effort to ensure that the assignment results were consistent with existing transportation options, and that routes to school were not blocked by hills, rivers, or other geographic barriers. The reporting system also created incentives to obtain electricity bills from areas near the most selective schools. These challenges reduced the overall transparency and predictability of the assignment system.[5]

In Peru, the government's objective in introducing centralized assignments was instead to improve the transparency of the school system. There were many reported cases of parents paying fees or bribes to ensure their children received admission to certain oversubscribed schools. Parents also often waited in long lines for days to apply for a vacancy in a selective school (Elacqua et al., 2022).

Ecuador

As mentioned above, families in Ecuador have historically been assigned to the closest public school based on household location as reported through the family's electricity bill code, a process that was costly, inefficient, and inequitable. In an effort to improve the system, the government partnered with IADB and ConsiliumBots to introduce parental choice through a centralized process. The region of Manta was chosen for the pilot, where local authorities supported the policy change.[6]

For the 2021 admission, the Ministry of Education (MoE) collected applications to national public preschools (ages 3-5) in Manta through the new centralized online platform. Based on families' submitted rank-order lists (ROL), students were assigned to one of their options using the deferred acceptance algorithm (also employed in Peru). The system covered three districts (namely, Manta, Jaramijó, and Montecristo) or area representing 2.5% of the national school enrollment.

[5]See Elacqua et al. (2022) for details on the distance-centric algorithm used in Ecuador and qualitative evidence for distortions in the electricity bill registration process.

[6]See Elacqua et al. (2022) for further information on the Ministry of Education's rationale for choosing the region of Manta.

The application process consisted of a single round. The online platform opened at the beginning of February 2021 and families had three weeks to complete their applications, with no limits on the number of schools they could include on their lists. They could furthermore modify their rank-order list multiple times.

Applicants could also apply to three types of institutions providing preschool education that were not listed on the online platform. These were municipal public schools, subsidized private schools, and private schools, with different ownership and management characteristics. Schools outside the national public network represented 45% of the possible options, and defined their own application processes.

The universe of applicants assessed here consists of around 4,000 children aged 3-5, with a balanced number of boys and girls. Column 1 of Table 3.1 shows that 3,984 applicants submitted a rank-order list for the 2021 admission process. The average length of the final portfolios was 1.9 schools, and around 66% of the families applied to schools in Manta, the largest district. Forty-three percent of the applicants requested a place in "pre-pre-kindergarten" (three-year-olds), and 43% in "pre-kindergarten" (four-year-olds). If we count the schooling options within 2 kilometers of each household, we observe that, on average, there are slightly more off-platform options (municipal public schools, subsidized private schools, and private schools) than on-platform options (national public ones).

Peru

In Peru, families have historically applied directly to each school. In 2021 and 2022, in an effort to improve transparency and efficiency in student assignment, the government worked with IADB and ConsiliumBots to introduce a centralized student assignment pilot in the region of Tacna—one of the objectives being to eventually scale this reform up to more regions.[7] The government chose Tacna because it was a small region with a significant concentration of schooling options. Additionally, the local government was a strong proponent of the reform.[8] For the 2021 and 2022 admission processes, all applications to public schools in Tacna were submitted on the new centralized online platform using a rank-order list (ROL), following which families potentially received a placement offer from the Ministry of Education (MoE).

The system covered 10 districts,[9] representing close to 1% of the national school enrollment, and was specific to placement in preschool through grade 1 (ages 3-6). As in Ecuador, students were assigned to one of their options using the deferred acceptance algorithm. The application process consisted of three rounds. We focus on the first round, as that is when our information intervention was implemented. Specifically, the online platform opened at the beginning of December in both study years. Families then had seven weeks to com-

[7]This process has been delayed due to the Covid-19 pandemic. In 2023, the government will begin introducing the reform in two additional regions: Arequipa and Madre de Dios.

[8]See Elacqua et al. (2022) for more details on the government's rationale for choosing the region of Tacna.

[9]Namely, Alto de la Alianza, Calana, Ciudad Nueva, Gregorio Albarracín, Inclán, La Yarada los Palos, Pachía, Pocollay, Sama, and Tacna.

plete their application, with no limits on the number of schools they could include on their lists. For the 2021 intake, applicants had only one chance to submit their rank-order list, and were not permitted to modify the latter unless the system authorized additional access. This restriction was relaxed in 2022, allowing applicants to adjust their application multiple times.

Applicants could also apply to private schools that were not listed on the online platform. In Peru, private schools compete in the provision of PK-11 education,[10] and interested families can apply directly to each school, following a decentralized process that is not coordinated with the public school choice process. In Tacna, the largest district participating in the pilot, private schools represent 40% of the available options. These schools charge tuition fees, do not receive funding from the government, and since a reform in 2012, are not allowed to engage in active selection or discrimination of students.[11]

The universe of assessed applicants in Peru for both years includes around 11,700 families with children aged 3-6, with a balanced gender ratio. Columns 4 and 7 of Table 3.1 show that 6,876 applicants submitted an ROL in the 2021 intake, while 4,856 applicants did so in 2022. The average length of the ROLs submitted was 3.3 schools, and around 40% of the families applied to schools in the largest district. Forty-nine percent of applications were for a grade 1 seat and 35% for "pre-pre-kindergarten" (three-year olds). If we count the schooling options within 2 kilometers of each household, we observe that, on average, there are more private options (off-platform), than public ones (on-platform).

Differences between Manta and Tacna

The regions in Ecuador and Peru where the pilots were implemented differently in several observed aspects. Consideration of these differences is important for understanding potential disparities in the behavioral response to our intervention. Specifically, local cultural attitudes toward school admissions, the availability of on-platform options, and rurality could all affect the application strategies and placement results.

In Ecuador, families had not been able to choose schools prior to the pilot, while a decentralized choice system already existed in Peru. As described above, applicants in Ecuador were assigned to the closest schools using the address on their electricity bill as a proxy of home location. Peru meanwhile had a decentralized system in which families applied directly to individual schools, with no coordination between institutions. This difference in school choice culture could help explain, for example, the number of schools included on each list. While the average length of the ROLs on the initial applications (pre-intervention) was 3.15 schools in Peru, it was only 1.83 in Ecuador (see Table 3.1). The market's underlying characteristics may partly explain why the ROLs were 65% longer in Peru.

[10] Ages 3-17; see Allende (2019) for an in-depth discussion of the Peruvian school system.

[11] For details, refer to *Directiva N 014-2012-MINEDU/VMGP "Normas y Orientaciones para el Desarrollo de Año Escolar 2013 en la Educación Básica."* However, recent evidence shows that private schools continue to use different instruments to screen students and families, including cognitive tests and interviews (Balarin et al., 2019).

Second, the density of schools is lower in the Ecuadorian context. Table 3.1 shows that an average applicant from Peru had 19 off-platform options and 16 on-platform ones. In Ecuador, in contrast, families had an average of 12 off-platform and 11 on-platform options. Furthermore, Figure C.1 in Appendix C.1 provides an example of the market concentration for both Manta and Tacna, the largest districts in each region. We see that a student who lives in the city center of Tacna has a larger potential choice set than her peers in Manta.

Finally, the interventions were not exactly identical in both countries—arguably the likely cause of most of the differences in the outcomes. In Ecuador, the warning related to placement risk was not sent as a separate message via WhatsApp; it was only included on the report card. Applicants in Peru received a specific WhatsApp warning message, which was also included in the report card. We provide further details on the intervention in Section 3.4.

3.3 Survey

We conducted an online survey to elicit participants' beliefs about placement probabilities, their level of knowledge about schooling options, and the difficulty of the application process. The MoEs distributed the surveys once the application processes had been closed, but before the placement results were made public. The evidence shows limited awareness and biased beliefs, suggesting that information interventions such as ours may prove beneficial.

Survey design and implementation

Our questionnaire aims to gain a better understanding of participants' knowledge and beliefs relating to the application process.[12] We included novel questions on parents' understanding of the mechanism, interpretation of school popularity, and awareness of private (off-platform) options. We distributed links to the survey (on the Qualtrics platform) through WhatsApp messages. Thirty-two percent of families completed the survey. They are more likely to come from the largest urban areas and are representative in terms of grades and gender.

The survey covered five aspects of the application process and was personalized for each applicant, taking into account the ROL submitted and the family's home location. Specifically, questions were asked about the (1) perception of the application platform, (2) application strategy, (3) level of awareness of ranked and non-ranked schooling options, including on- and off-platform alternatives in the applicant's neighborhood, (4) beliefs on assignment probabilities, and (5) satisfaction relative to hypothetical placement alternatives.

As mentioned, the online survey was implemented just after the application process and before the results were made public, to avoid potential changes in beliefs based on placements. Parents with two or more applicants were surveyed only once, choosing randomly between the associated students. The survey was not incentivized in any way, though we did send a reminder to parents who did not answer on the first day.

[12] The content of the survey is based on the questionnaire of Arteaga et al. (2022).

Survey completion rates were higher in Ecuador than in Peru, and the populations in the largest districts were more likely to answer. Columns 3, 6, and 9 of Table 3.1 show that the completion rate in Peru was respectively 25% and 39% in the two years of the study, and 47% in Ecuador. Families that responded to the survey tended to reside in zones with a slightly higher density of schools and, on average, applied to more schools.

Survey results

Our survey results show that applicants are overly optimistic in terms of placement probabilities and that their awareness of the available options is very limited. We also observe that families have a strong desire to be placed, and that finding out more about a given school is hard.[13] Generally, respondents have an imperfect notion of the optimal strategy and the signal of popularity does not correlate with school quality for everybody.

Applicants with a positive probability of non-placement hold a belief about their admission chances that is around 30pp higher than the actual probability, i.e., a considerable optimism bias. In Ecuador, an average applicant thinks that her child's chances of being placed in at least one of the options in her ROL are 37pp higher than the true likelihood (Panel 3.1a). In Peru, this optimism is 29pp and 20pp in 2021 and 2022, respectively (Panels 3.1b and 3.1c). In Panel 3.1a we observe that a significant number of applicants in Ecuador have virtually no chance of being assigned to one of their options while simultaneously indicating their complete confidence that they are going to be placed: around 20% of Ecuadorians at high risk of non-placement have an optimism bias of over 80%. In contrast, in Peru, this group represents less than 2% of the risky applicants.[14]

Biased beliefs on admission chances affect application strategies. Panels 3.1d, 3.1e, and 3.1f show that the two most common reasons why applicants did not add more schools are optimism and a lack of options. Respondents from the two countries differ in terms of the modal reason. In Ecuador, the most common reason is the availability of schools, which makes sense given the lower density of education establishments. Meanwhile in Peru, optimism bias appears to be the most common reason for not adding more schools to the application.

The responses to a satisfaction question concerning different placement scenarios suggest that a non-assignment outcome has relevant welfare implications for participants. Panels 3.1g and 3.1h indicate that "not being placed" is a scenario that most families dislike. More than 90% of families give a failing grade to the scenario of non-placement, while 90% give an excellent grade to placement in first preference. There is also a considerable decline in satisfaction in the hypothetical scenario where a participant is moved from their first to their last option.

[13] Our survey results are consistent with the main survey findings of Arteaga et al. (2022).

[14] This is likely related to the context. In Peru, families were used to a competitive school admission process, albeit a decentralized one. In Ecuador, the previous admission system assigned students to the closest school, a less useful experience for forming beliefs on centralized admissions processes.

We added three survey components geared toward understanding the results of the information intervention. The first provides a sense of applicants' sophistication in a setting with a strategy-proof mechanism. Panels 3.2a to 3.2c show the proportion of respondents who answered correctly from the perspective of a user who knows how deferred acceptance works and reports his ranking truthfully. The first question asks "Imagine that you find a school that you like very much, even more than your first preference, but it has 100 applicants and 30 seats. What would you do?" The correct answer is to rank the school in first place, but most families (73% in Ecuador and 78% in Peru) answered they would add it below the current first preference or would not include it on their list. The second and third questions relate to the effect of adding more schools to the list. Seventy-nine percent and 75% of the applicants in Peru and Ecuador mistakenly said that this will decrease the chance of being placed in their first preference, while 55% and 54% answered correctly that it reduces overall placement risk.

A second novel insight comes from the (declared) inference that families make from schools that are popular. It seems that there is no consensus on the signal that generates high demand. We asked "If you find out that there is a school that many other families are applying to, but that you have not added to your list, you would say that:" Panels 3.2d to 3.2f show that Ecuadorian parents are more likely to answer "I don't know," and that in both cases, the proportion of parents who chose another option increases with the mother's education. Less than a third of respondents said that a popular school is probably a good school, while a similar proportion answered that its popularity provides no insights into the quality of the school.

Finally, the survey reveals that families are not well informed about the private options in their neighborhood. Panels 3.2h and 3.2g show that close to 40% of applicants have never heard of the largest private school within a radius of 3km of their home address. This proportion is around 60% when we asked about the closest private school. A random school in the area is less known than either of the latter two, as expected. Our benchmark of a high level of awareness is provided by the same respondents. Figure C.2 shows that only 4% and 9% of the applicants have no knowledge of the first option on their ROL in Peru and Ecuador, while 9% and 30% have no knowledge of the third school. We also asked about a fake school, to check the quality of the responses. Around 90% of applicants stated that they did not know about the school, and only 1% declared themselves to be familiar with it.

Families have an imperfect understanding of the deferred acceptance assignment mechanism. This is reflected in their declared strategies, which neither benefit their application nor the stated effects on their beliefs from hypothetical strategies. This is unsurprising given that 2021 was the first year in which the centralized mechanism was implemented. Applicants do not necessarily infer that a popular school is a good school, and have very limited knowledge about private options.

3.4 Intervention design

The survey evidence suggests that there is scope for helping parents to form more accurate beliefs about their children's chances of admission, and to become informed about neighborhood schooling options.[15] We designed an information intervention that included feedback on admission chances following Arteaga et al. (2022), to which we also added a suggested list of schools that was tailored to each applicant based on their current application, grade, and geographic location.

Our intervention included a warning to applicants with a positive chance of non-assignment along with a list of suggested schools that parents could potentially add to the application. The implementation team drew best practices from past experiences in order to maximize the probability of success of the process. One relevant aspect was the need to tackle the optimism bias over placement chances. The process of warning families about the risk of non-placement created a communication channel where we could innovate. Based on the same costly search framework in Arteaga et al. (2022), we complemented the warning with information about alternatives that were not considered in the families' initial ranking. This new information was intended to lower the search cost, potentially affecting the conformation of the final portfolio.

In practical terms, our research team worked with the MoE in both countries to identify applicants with a predicted probability of non-placement higher than 1% in 2021 for Ecuador and Peru, and 30% or higher for Peru in 2022. Before the end of the application process, we sent a communication—or what we call a warning—to these parents about the chance that their child might not being assigned to any of their choices.[16] In addition to the warning, we randomly assigned one of three different report cards that contained the following information:

T_1: Only warning

T_2: Warning + list with 10 suggestions

T_{3-2021}: Warning + list with 10 suggestions + information on popularity and congestion

T_{3-2022}: Warning + list with 3 suggestions

In theory, providing information about the available options ($T2$ and both $T3$) would reduce the application cost, inducing marginal applicants to add schools to their lists. Survey evidence shows that gathering information about a school is costly. Panels (a) to (c) of Appendix Figure C.3 document that at least 84% of the respondents value information about a school's academic performance, extracurricular activities, and infrastructure. Close

[15] Survey evidence also shows the need to educate applicants on the consequences of a strategy-proof mechanism. Though beyond the scope of our intervention, future research might explore this topic.

[16] In Tacna, the warning was given four days before the end of the process, while in Manta, it was sent six days beforehand.

to two-thirds also value references from other people about the schools, interviews with staff, and information on the school's website or Facebook page. We also asked participants about how important they feel it is to have information on the families that attend the school. Around 45% of Peruvian and 60% of Ecuadorian respondents agreed that it is important.

Our intervention does not eliminate search costs entirely, but rather aims to facilitate the search process for families that marginally stopped looking for alternatives. The additional information in T_{3-2021} works in at least two potential ways. The popularity was designed to signal what other families like, which could potentially focus the search, or simply be used as an additional school attribute to consider. Congestion information could be employed as a tool to evaluate which schools would be safer to apply to, but also as a proxy for popularity. Since we did not randomize the allocation of popularity or congestion information, we are not able to differentiate their particular effects.

Details on inputs and construction

For the warning, we used the same message as Arteaga et al. (2022), adding a "fire rating" symbol to show the level of risk. Figure 3.3a shows the warning included in the report card. It displayed the following message (all treatment arms):

> We have detected that many families are applying to the same schools as you, so there is the possibility that you will not be granted a spot in any of them. Remember that to increase the chances of obtaining a spot, we recommend adding all the schools that you would be willing to attend to your application.

The school suggestions for T_2 and T_{3-2021} consisted of a list of 10 schools that the student did not include in her initial ROL, while T_{3-2022} included only three schools. We built each personalized list by adding alternatives located as far as 3km from the declared home address. The 10 schools sent in T_{3-2021} included at least one popular *under*subscribed school, one popular *over*subscribed school, two non-popular *under*subscribed schools with at least 5 applicants, and two non-popular *over*subscribed schools with at least 5 applicants. To round out the 10 schools, we added random schools from the student's neighborhood.

To create proxies of popularity, we used the applications collected at the time of the intervention. We classified schools according to the number of applications. The minimum number to be considered "popular" was the number of applications received by the most demanded school with some available seats. This definition allowed us to classify at least one school as *under*subscribed within the set of popular schools and, potentially, many oversubscribed schools.[17] The process was conducted at a district level, meaning that only applicants from the district were considered for the definition of popularity within each specific geographic zone.

[17]We define a school as "*over*subscribed" if the probability of a regular applicant being placed there is less than 100%, which is equivalent to having more demand than seats. A school is "*under*subscribed" if every potential applicant to the school can be placed there. We follow the same procedure as in Arteaga et al. (2022) to obtain the placement probabilities for each school.

The information provided to families who received lists of suggested schools included the school's name, the distance from the applicant's address on the application form, and the levels of education offered.[18] The Peruvian report card also included whether the school was single-sex or co-ed.

The information on popularity and congestion provided in T_{3-2021} incorporates two additional pieces of information for each of the 10 schools on the personalized suggestions list. The first was a discrete category called "popular," which was based on the number of applications from the same district, as explained above. We displayed this on the report card as "High" or "Low" demand. The second additional component was the number of applicants and open seats available.[19] Extracts of the report cards are shown in Figures 3.3a and 3.3b. The full report card is presented in Figure C.4 in Appendix C.1.

Sample

Four days before the end of the process in Peru and six days in Ecuador, we used the total sample of filed applications accumulated up to that time to estimate the probability of non-placement for each participant. We randomly assigned applicants with a predicted non-placement probability of higher than 1% (in 2021) or 30% (in 2022) to one of the three treatments. We then sent a message through the WhatsApp mobile application that included a link to the report card containing the warning and, for T_2 and T_{3-2021}, the list of suggested schools.[20] In Peru 2021 and 2022, we also sent a separate WhatsApp message related only to the warning right before the link to the report card.

In Peru in 2021, the online platform allowed only one submission attempt per applicant. The authorities provided families assigned to the treatment group with additional access to log in and modify their applications.

Columns 2, 5, and 8 of Table 3.1 describe the RCT sample population for Ecuador and Peru, with all choice participants exhibiting some level of placement risk. We intended to treat 51% of the applicants in Ecuador, 25% in 2021 Peru and 39% in 2022 Peru, reflecting a more congested pre-intervention scenario in the first setting. The lower proportion of intended recipients of the treatment in Peru in 2021 vs. 2022 is partially explained by the fact that in 2021 we treated only students applying to PPK and grade 1.[21] Compared to the

[18]For Ecuador, the report card explicitly showed which educational levels were offered at the school (Inicial, EGB, and Bachillerato, which correspond to preschool, elementary, and high school, respectively). For Peru, this information was limited to whether or not the school was classified as integrated (integrado), meaning that it offered both preschool and some higher levels of education (e.g., preschool + elementary or preschool + elementary + high school).

[19]The number of applicants corresponds to the mean of the number of admitted plus waitlisted students from 500 simulations of the assignment based on the current demand. In this case, we did not differentiate from the applicant's geographic origin, we included all applicants.

[20]In Ecuador 2021 and Peru 2021, we also sent the link to the report card by email. A full description of outreach strategies is presented in Table C.1 in Appendix C.2.

[21]PK and K both had low congestion levels in Peru across both years. Since the number of potentially treated applicants was small, we decided not to implement the intervention in those grades.

average student, applicants assigned to the treatment group filed shorter pre-intervention portfolios and were likelier to belong to the largest districts, namely Tacna in Peru and Manta in Ecuador.

Delivery of information and treatment take-up

We used the WhatsApp messaging app to distribute the links to the report cards with the information for each treatment arm.[22] In our first message, we told parents that we had information about the application to share with them, and asked if they were interested. For those that answered positively, in Peru, we sent a warning about the chances of non-placement followed by a link to the personalized report card. In Ecuador, we only sent the link.[23]

Table 3.2 presents the main statistics on the intention to treat and messaging reception. Panel B shows that WhatsApp messaging was more effective in Peru. We sent an introductory message to 100% of the applicants assigned to the treatment, and 89% of them read it in the 2021 version and 92% in the 2022 version.[24] In Ecuador, we sent WhatsApp messages to only 22% of the targeted population, and 90% read them.[25]

All applicants who replied to the initial message were sent a link to the report card (panel D), which was preceded by an initial warning message in the case of Peru (panel C). Panels B and C of Table 3.2 reveals that 69% and 86% of the 2021 and 2022 Peruvian applicants assigned to the treatment received the warning message and a link to the report card, while only 19% of Ecuadorian applicants received the message with the link. The proportion of parents who read the message related to the report card is very similar to the sent rate since this group had already replied to our introductory message.

In the 2021 Peru and Ecuador admission processes, we also sent the link to the report card by email (panel A), a strategy that we did not use in Peru in 2022. The last row of Panel D in Table 3.2 shows that the proportion of the population that viewed the report card was 43, 63, and 53% for 2021 Ecuador, 2021 Peru, and 2022 Peru, respectively. In the case of Ecuador, this outreach would not have been possible without the outreach by email, as clearly seen in the second row of Panel D in Table 3.2, which shows the mean proportion of applicants who did not receive the WhatsApp message but still opened the report card: 36% and 46% in 2021 Ecuador and Peru and just 5% in 2022 Peru.

[22] In both countries in 2021, we also sent the information by email. Table C.1 contains a summary of the interventions and channels.

[23] The warning message was included in the report card in both countries. The difference was that in Peru, we also sent it as a separate WhatsApp message. For more details on the messages, the original text, and translations to English, see Appendix C.3

[24] A particular feature of the WhatsApp messaging app is that it provides insights into message status since it distinguishes between messages that have been sent, delivered, and read.

[25] The low rate of messaging in the Ecuadorian context was not by design but rather the result of implementation difficulties.

3.5 Choice behavior and choice outcomes

Survey evidence shows that applicants have imperfect knowledge about the options around them, and are overly optimistic about their admission chances. An information intervention could therefore play a potentially relevant role in this setting. In theory, a non-placement warning reduces the under-search behavior by correcting the biased beliefs on admission probabilities. Meanwhile, providing alternative options reduces the search cost. Both interventions should affect the construction of the rank-ordered list. In this section, we present the results from warning messages and the randomly assigned information intervention (T_2 and T_{3-2021} and T_{3-2022}), compared to the basic report card (T_1), which does not contain the suggestion list.

Our survey evidence also suggests a channel that can potentially reduce the response to our intervention. A meaningful proportion of families have incorrect beliefs about the impact of adding a new school on the placement probabilities of alternatives they have already considered, which could lead to them not adding more schools to the list. Furthermore, many families do not make inferences regarding a school's quality based on its popularity, potentially making the information provided in T_{3-2021} less useful.

First, we document that the warning affects parent behavior. Figure 3.4 shows the regression discontinuity plots for 2022 Peru and for Arteaga et al. (2022) Chilean pooled sample from 2018 to 2020 (Figure 5b in their paper). The horizontal axis represents the predicted placement risk (probability of non-placement), the metric used to assign the warning message in both contexts. Only applicants with a risk level higher than 30% received a warning. We observe a discontinuous behavior, reflected in applicants to the right of the threshold adding more schools.[26]

Second, we test the causal effect of providing school suggestions on a sample that is restricted in two ways with respect to the universe of applicants. First, the sample includes only applicants with elevated placement risk,[27] and therefore all treated applicants received a report card with a non-placement warning. Second, we limit our sample analysis to applicants who opened the report card. We define this group as the compliers to the information campaign. This was possible as all applicants received a link to a report card, regardless of whether or not they were assigned to the additional treatment that included a suggestion list. We do not find different selection into the analysis sample between T_1 –our "control"– and the other treatments.

First, we focus our analysis on the differential behavioral response between applicants who received the suggestion list of ten schools (T_2) and those who did not see such a list

[26] There are differences between the two studies in terms of the channel used to deliver the non-placement warning. The plot from Arteaga et al. (2022) represents a message shown in a pop-up on the application platform, displayed as families prepared to submit their applications, while in the case of 2022 Peru, the warning took the form of a WhatsApp message. The levels of precision in Figure 4 also obviously differ. Figure 5b from Arteaga et al. (2022) was built using considerably more observations.

[27] That is, applicants with predicted placement risk > 0 for the 2021 school choice processes and applicants with predicted risk > 0.3 for the 2022 Peru admission process.

on the report card (T_1). Since we implemented T_2 on all three contexts, we can pool the individual samples to calculate aggregate results. In column 1 of Table 3.3 we compare the pre-treatment and post-treatment ROLs for the sub-sample to which we assigned the basic version of the report card with no suggested schools (T_1).[28] Column 2 reports the differential effect of the information added in treatment T_2 (suggestion of 10 schools) on the changes in the pre- and post-intervention ROLs compared to T_1.

We observe a marginally significant effect on the number of schools added. Applicants that receive the lists add, on average, 23% more schools to their list. Students assigned to T_2 are more likely to include schools from the list. When we don't show the list, 12% of the families add a school from what could have been their list. When we show it, 19% of the applicants add at least one suggestion. We can't rule out a null effect on the type of school added. Despite our point estimates for "Add popular" school or "Add congested" school are positive, they are imprecise.

Now we look at the results by implementation country/year. In this analysis, we consider the effect of treatments T_3 that are specific to each implementation year. In columns 1, 4, and 7 of Table 3.4 we compare the pre-treatment and post-treatment ROLs for the sub-sample to which we assigned the basic version of the report card with no suggested schools (T_1). Columns 2, 3, 5, 6, 8, and 9 report the differential effect of the information added in treatments T_2 and T_{3-2021} and T_{3-2022} on the changes in the pre- and post-intervention ROLs compared to T_1.

Columns 1 to 3 show the effect on Ecuador. The first column indicates that 10% of the applicants who opened a report card with a non-placement warning added at least one school to their list. On average, these families extended their portfolios by 2.4 schools. Columns 2 and 3 show that there is no statistically significant effect of providing school suggestions in the context of Ecuador.[29]

Columns 4 to 6 report the response in the Peruvian 2021 context. One-third of the families that opened our link added a school to their application. In this case, we observe a statistically significant differential effect between applicants assigned to treatments T_1 and treatments with suggestion lists. Applicants who received the school suggestion list (T_2 and T_{3-2021}) were more likely to add a school from the list. Columns 5 and 6 suggest that the list of schools also shifted preferences toward the suggested options. The proportion of applicants who added schools from the list was 68% (13 pp) higher in T_2, and 51% (or 10pp) higher in T_3.

If we examine the type of school families add, we observe no significant differential effect

[28] This is not an estimated causal effect of the warning. We maintain, however, that our contribution comes from the marginal effect of the list of suggested schools. We are interested in comparing subgroups that were *all* exposed to the warning, but to which we randomly assigned differing levels of information (T_2 and T_3). See Arteaga et al. (2022) for experimental and quasi-experimental evidence on the effect of the warning.

[29] There are two marginally statistically significant results. First, applicants assigned to T_{3-2021} added fewer schools, which could imply a negative effect of providing too much information. Second, the same population is more likely to not add schools that are not on the list.

between T_2 and T_{3-2021} for 2021 Peru.[30] This suggests that either the extra information on congestion and popularity is not a relevant input for families, or that they are already aware of these characteristics and thus the information is not new to them.

If we look at the effect one year later (2022 Peru, columns 8 and 9), we observe that families who did not receive suggestion lists added fewer schools than families with the list, but our estimates are not precise to reject zero effect. This result can be partially explained by the fact that they did not receive the warning message on the report card (see Table C.1). We observe that suggestion lists of varying lengths (3 or 10 schools) have a similar effect on the proportion of applicants that add a school from the suggestion list: T_2 increased the proportion of families that added schools from the list by 120% (or 5pp), the same absolute magnitude as the effect of T_{3-2022}. We also observe a marginally significant effect of the short suggestion list on the proportion of families that add at least one school. The proportion of applicants that added a new school to their portfolio increased by 49% (or 7pp) when they were assigned to the list of three suggestions compared to when they did not receive a list of suggestions.[31]

Panel C of Table 3.4 shows that applicants who received suggestion lists are no more likely to be placed or enrolled in one of the schools on the list.

Although the treatments affect search, we find no significant differences in the observable characteristics of the assigned schools across different treatments. One hypothesis is that applicants who did not receive the suggestions could find schools as easily as their peers who did receive them. If this is the case, the intervention may still help treated applicants to decrease the search costs faced by families at almost zero marginal cost.

Discussion

Despite being relatively similar interventions, the results from the three contexts (2021 Ecuador and Peru and 2022 Peru) somewhat differ. Four factors may help to understand these differences. First, the implementation of the information campaigns and the application systems were not identical. Second, the availability of options may have also played a role. Third, the underlying cultural differences between Peru and Ecuador could shape the behavioral response, as noted in Section 3.2. Finally, a minor change was made to the school choice process in Peru between 2021 and 2022.

Our intervention in Ecuador differed in two key respects. First, the WhatsApp conversation did not include a separate warning on the placement risk (see Table 3.2). While the warning message was included in the report card in every context, it was arguably more salient to families who received it as a separate WhatsApp message, as was the case for both years in Peru. The report card contained information about the current application, the warning, and the suggested list, which may have been an overload of information for many

[30] Table C.2 in Appendix C.2 show the estimates for the differential effect of T_3 versus $T2$..

[31] Treatments T_{3-2022} and T_1 also differ in that the former included the warning message in the report card, while the latter did not (see Table C.1 for details).

applicants.[32] Second, implementation issues meant that we were only able to reach around 22% of families on WhatsApp, which may have affected the precision of our estimates.

Another difference between the two countries that may have shaped the results is the density of schooling options. As discussed in Section 3.2, Ecuadorian applicants had a lower density of local schooling options to choose from (see Table 3.1 or Figure 3.3b). Thus, information about all available local options may have been easier to collect. Table 3.4 shows that, of the participants who added a school and were *not assigned* the suggestion list (T_1), 86% of the Ecuadorian applicants added a school from our list, a figure that was 55% and 33% for 2021 and 2022 Peru. Since Ecuadorian applicants were already choosing schools from the list without us revealing this information to them, the potential effect of showing the list was constrained to a much smaller population than in Peru.

There are also significant differences in the choice culture in Ecuador and Peru. Families in Ecuador have historically had no choice as to where their children go to school. Rather, the latter are centrally assigned to the nearest establishment. We do observe baseline differences in application behavior. The first portfolio that families submitted (i.e., before our intervention) was 42% shorter in Ecuador. In contrast, Peru has historically had a decentralized choice system in which families need to apply directly to schools, such that they are already accustomed to searching for schools. In our model, this could be interpreted as the population has a lower search cost, which would make them more likely to react to changes in their beliefs.

Lastly, one detail may help to understand the differences between years in Peru. The application process changed subtly between 2021 and 2022: in the first year, applicants could only apply once, with no opportunity to modify their application. When we sent our report cards, the platform granted special access to the families we reached with our intervention. In the 2022 version, all applicants could return to the platform and modify their respective lists of schools.

Survey results

We evaluated whether the additional suggestion lists (in T_2, T_{3-2021} and T_{3-2022}) and the information on popularity and congestion (T_{3-2021}) impacted subjective measures captured by our survey of applicants. We find evidence related to the perception of the application process: applicants that received suggestion lists in 2021 Peru were less likely to say that it was hard to search for schools. In 2022 Peru, students assigned to T_2 or T_{3-2022} were more likely to declare that they received the warning message. This is consistent with the implementation, since families in T_1 only received the warning through WhatsApp, and not in the report card.[33]

[32]There is an emerging literature on people's limited capacity to pay attention to all the potential attributes in the choice process, and efforts have been made to incorporate this into economic models. See Gabaix (2019) for a review.

[33]See Table C.1 for details on the contents of each treatment.

Another statistically significant result (at the 1% level) indicates that applicants in Ecuador assigned to T_2 rated the quality of the "Information about schools available on the application platform" lower than other groups. We interpret this result with caution. First, the treatment is not directly related to the information available on the platform. Second, the result does not hold for T_{3-2021}. Third, we are testing 16 hypotheses in Table C.3, meaning there are high chances of a type I error.

The treatments had no effect on the declared satisfaction with hypothetical placement results (Panel B). Our intervention did not aim to promote changes in the first preference, but we did expect to have an effect on the preference vis-à-vis the lowest-ranked option, since the invitation was to "add more schools to the list." Applicants that received the suggestion lists did not declare a lower level of satisfaction with the schools chosen at the bottom of the rank-order list.

Our treatment affected participants' level of knowledge of the schools. We asked them to rate their knowledge of five schools out of the ten listed in the report card. Peruvian applicants in 2021 who did not receive the list (T_1) declared that they were aware of 36% of the schools. For students assigned to T_2, this proportion increases by 14pp, equivalent to being aware of 0.7 more schools listed on the report card. There is an opposite effect in Ecuador, but with half the magnitude: students assigned to T_2 are less likely to declare that they are aware of the schools.[34]

3.6 Conclusions

This paper builds on Arteaga et al. (2022) and shows that inaccurate beliefs about admission probabilities extend to the contexts of Manta, Ecuador and Tacna, Peru. Motivated by their framework, in which searching for schooling options is costly and choice participants hold biased beliefs, we implemented a warning strategy and designed a new information intervention that included school suggestions tailored to applicants with positive non-placement probabilities. We find that the new information has a causal effect on search, increasing the number of schools added to the list, and shifting preferences towards the suggested schools.

The goal of the information intervention was to test the effect of providing personalized school suggestions together with a non-placement warning message. The theoretical framework suggests that lowering the cost of searching for new alternatives should affect choice behavior. The main implications of providing a list of suggested schools are two. First, they can reduce search costs, increasing the number of schools added. Second, it can shift preferences, increasing the likelihood of adding the suggested schools, and potentially inducing changes in the search process. An additional implication is that this result is both context- and implementation-dependent. We test the effect of the suggestions in different regions of Ecuador and Peru, observing a shift in preferences only in the Peruvian context.

[34]For further details, see Panel C of Table C.3.

These findings shed light on the need to help applicants understand centralized choice before and during the application process with better information. Our survey evidence showed a clear misunderstanding of the rules of the assignment mechanism, reflected in sub-optimal declared strategies and incomplete knowledge about the available options.

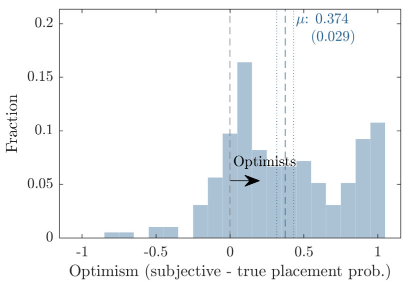
The interventions also provide additional insights: communication channels and content matters. Where it was possible to communicate through WhatsApp, the messages were more effective than by email. Furthermore, the 2022 Peru implementation reveals that it is better to provide information using more than one channel. Applicants that received an additional warning message on the report card were more likely to respond than the families that received it only via WhatsApp.

We conclude that understanding the local context together with effectively designing platforms and carefully selecting the information they contain can shift search and choice behavior and should be seen as important aspects in the creation of centralized choice policy.

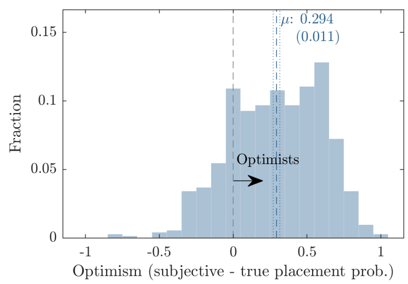
Figures and Tables

Figure 3.1: Main survey evidence

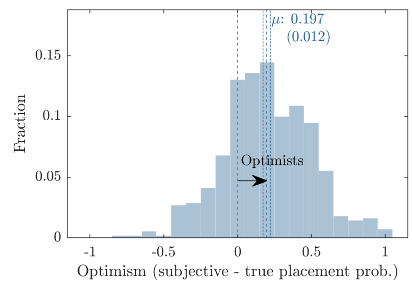
(a) Optimism on placement probability - Ecuador



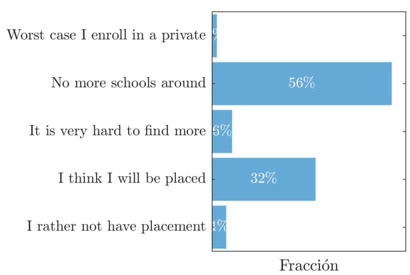
(b) Optimism on placement probability - Peru 2021



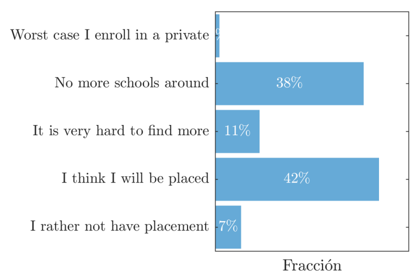
(c) Optimism on placement probability - Peru 2022



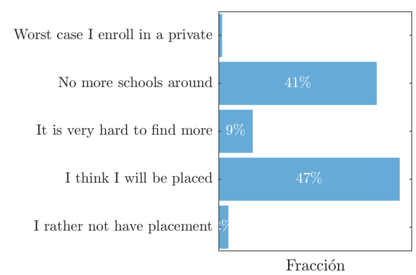
(d) Stated reason for not adding more schools - Ecuador



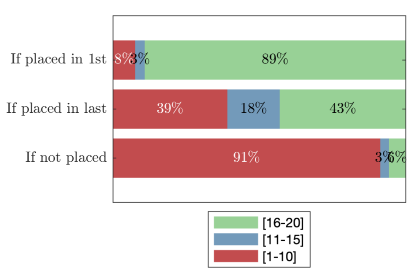
(e) Stated reason for not adding more schools - Peru 2021



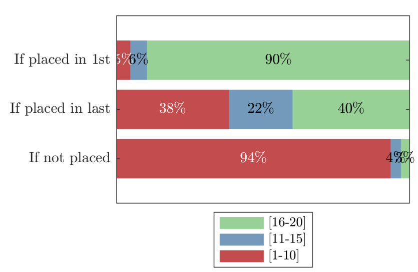
(f) Stated reason for not adding more schools - Peru 2022



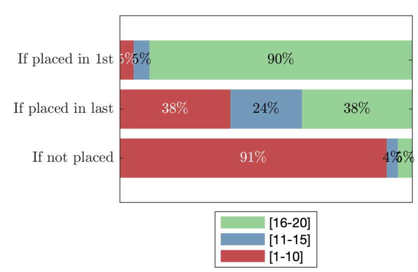
(g) Satisfaction with placement by rank - Ecuador



(h) Satisfaction with placement by rank - Peru 2021



(i) Satisfaction with placement by rank - Peru 2022



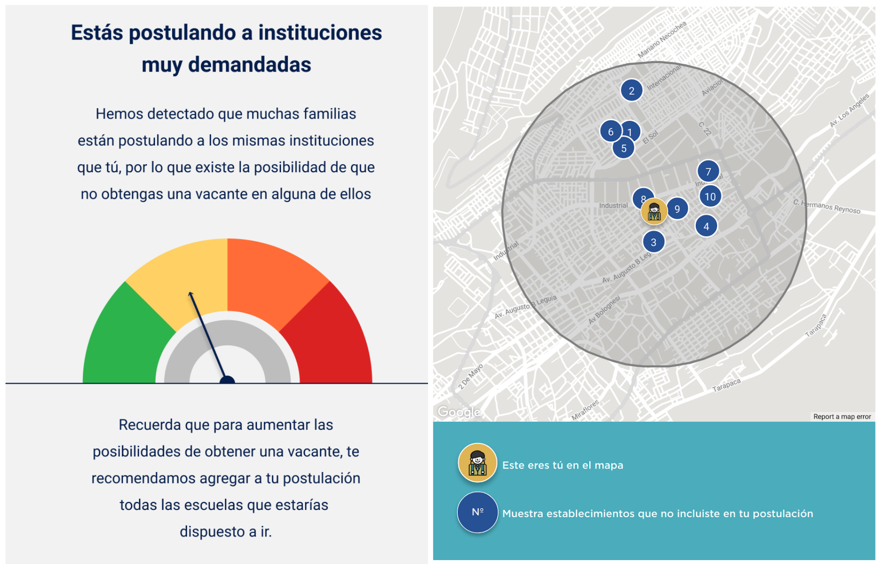
Notes: Panels (a), (b), and (c) show the differences between the subjective and true placement probabilities for the subset of applicants with placement risk>0.01. The subjective placement probability comes from the question "On a scale from 0 to 100, with what probability do you think that [applicant name] will obtain a spot in at least one of the [number of schools in ranking] schools in the ranking? Panels (d), (e) and (f) represent the answer to the question "Why didn't you add more schools to your application? (select the main reason)" for applicants with placement risk>0.01. Panels (g), (h), and (i) asked about the level of satisfaction for three scenarios: placed in first choice, last choice, and no placement ("If [applicant name] gets a spot in the following schools, from 1 to 20, how satisfied would you be?"). See Appendix Section C.4 for details on the survey questions.

Figure 3.2: New Survey Evidence

Notes: Panels (a), (b), and (c) show the answers to three questions related to the understanding of the mechanism. The first bar (from top to bottom) represents the answer "I add it to my list in 1st preference" to the question "Imagine that you find a school that you like very much, even more than your first preference, but it has 100 applicants and 30 seats. What would you do?". The second bar represents the answer "No" to "If you add more schools to your application, do you think the possibility of being assigned to your first preference decreases?" The third bar represents the answer "Yes" to "If you add more schools to your application, do you think the non-placement probability decreases?" Panels (d), (e), and (f) represent the answer to the question "If you found out that there is a school that many other families are applying to, but you haven't added it to your list, what would you say about its quality that...". Panels (g), (h), and (i) asked about the level of familiarity with four private schools (not available on the platform) located within 3km of their home address. See Appendix Section C.4 for details on the survey questions.

Figure 3.3: Report card extracts

(a) (b)

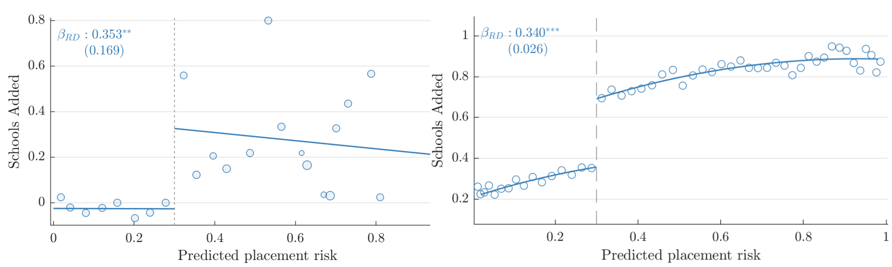


Notes: Both panels are extracts from a report card sent to applicants with positive placement risk that were assigned to T_2 (warning + suggestion list). Panel (a) represents the non-placement warning while panel (b) shows the map of the 10 suggested schools that the applicant did not include in her ranking. The full report card is presented in Figure C.4 in Appendix C.1.

Figure 3.4: Schools added RD

(a) Peru 2022

(b) Chile 2018-2020 (Figure 5b from Arteaga et al., 2022)



Notes: Binned means and global fits of schools added after the information campaign by predicted risk for the pre-campaign application. The non-placement warning was assigned only to applicants with a predicted risk higher than 0.3 (30%), as indicated by the vertical dashed line.

Table 3.1: Descriptive statistics for applicants

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Ecuador 2021			Perú 2021			Perú 2022		
	All	RCT Sample	Pre Placement Sample Survey	All	RCT Sample	Pre Placement Sample Survey	All	RCT Sample	Pre Placement Sample Survey
Female	0.49	0.50	0.50	0.49	0.48	0.50	0.50	0.54	0.50
From largest district	0.66	0.71	0.70	0.43	0.58	0.49	0.40	0.65	0.44
Length pre-treatment portfolio	1.83	1.79	1.96	3.14	2.97	3.24	3.16	2.85	3.30
Length final portfolio	1.90	1.91	2.05	3.34	3.81	3.50	3.22	3.09	3.39
In-platform opts in 2km radio	11.32	13.10	12.46	16.33	18.55	17.01	16.66	19.90	17.46
Off-platform opts in 2km radio	12.11	13.94	13.37	19.28	27.98	21.05	20.30	29.58	22.26
Grade									
PPK (3 yrs old)	0.43	0.43	0.43	0.35	0.37	0.38	0.35	0.28	0.37
PK (4 yrs old)	0.43	0.45	0.44	0.10	0.00	0.09	0.12	0.11	0.11
K (5 yrs old)	0.14	0.12	0.13	0.06	0.00	0.06	0.04	0.03	0.04
1st (6 yrs old)	0.00	0.00	0.00	0.49	0.63	0.48	0.49	0.58	0.48
N	3,984	2,021	1,872	6,876	1,708	1,721	4,856	1,140	1,501

Notes. All statistics are means in the population defined by the column header. Largest district is Manta for Ecuador and Tacna for Perú.

Table 3.2: Take-up of WhatsApp Messages and Report Card

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Ecuador 2021				Peru 2021				Peru 2022			
	In RCT	Treatment			In RCT	Treatment			In RCT	Treatment		
		T_1	T_2	T_3		T_1	T_2	T_3		T_1	T_2	T_3
A. Email with link to report card												
Sent	1	1	1	1	1	1	1	1	0	0	0	0
B. WhatsApp introduction												
Sent	0.22	0.21	0.21	0.25	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Deliverd	0.21	0.20	0.19	0.25	0.92	0.94	0.93	0.91	0.96	0.96	0.96	0.96
Read	0.20	0.19	0.18	0.23	0.89	0.90	0.88	0.88	0.92	0.92	0.92	0.92
C. WhatsApp warning												
Sent	0.00	0.00	0.00	0.00	0.69	0.68	0.72	0.67	0.86	0.85	0.88	0.86
Deliverd	0.00	0.00	0.00	0.00	0.69	0.68	0.72	0.67	0.86	0.85	0.88	0.86
Read	0.00	0.00	0.00	0.00	0.66	0.66	0.68	0.65	0.83	0.83	0.84	0.81
D. WhatsApp with link to report card												
Sent	0.19	0.18	0.16	0.22	0.69	0.68	0.72	0.67	0.86	0.85	0.88	0.86
Deliverd	0.19	0.18	0.16	0.22	0.69	0.68	0.72	0.67	0.86	0.85	0.88	0.86
Read	0.17	0.17	0.14	0.19	0.66	0.66	0.68	0.65	0.83	0.83	0.84	0.81
E. Opened link of report card (Google Analytics)												
Obs. with link sent by WhatsApp	0.76	0.73	0.79	0.77	0.71	0.66	0.72	0.75	0.53	0.56	0.48	0.53
Obs. without link sent by WhatsApp	0.36	0.32	0.35	0.41	0.46	0.48	0.44	0.46	0.05	0.04	0.04	0.07
All	0.43	0.39	0.42	0.49	0.63	0.60	0.64	0.65	0.46	0.48	0.43	0.47
N	2,021	676	673	672	1,708	568	572	568	1,140	377	380	383

Notes. All statistics are proportion in the population defined by the column header. Panels A to C show the mean of the status for the three WhatsApp messages. “Sent” means that we tried to reach the applicant, “delivered” that the applicant received the message on his app, while “read” that the applicant saw de message. Every message that is read is also delivered and sent, and every message that is delivered is also sent. “WhatsApp introduction” (Panel A) is the first message we sent to families, asking if they want to receive information about the application. “Whatsapp introduction” is the message in which we invited the families to recieve more information about their application. We sent additional messages only to applicants who answered positively to the initial message. “WhatsApp warning” (Panel B) is the message that contained the alert about the placement risk and a recommendation to add more schools. “WhatsApp with link to the report card” (Panel C) was sent after the previous one, and had the hyperlink to the personalized information treatment. Panel D shows the proportion of students that opened the link. The first row (“Link sent by WhatsApp”) is conditional on the report card link being delivered through WhatsApp, the second on not being delivered, while the third row is unconditional. The link was also sent by email in Ecuador 2021 and Peru 2022, but we do not have data on recepection status.

Table 3.3: RCT Results: Effect of Suggestions on Application Outcomes.

	(1)	(2)
Country	Combined	
	T_1	T_2
Intervention	Warning	+ list (10)
	(base)	(diff.)
A. Choice behavior		
Add any school	0.207	0.028
	(0.014)	(0.020)
Number of schools added	0.540	0.124*
	(0.045)	(0.075)
Add popular	0.153	0.026
	(0.012)	(0.018)
Add congested	0.192	0.017
	(0.014)	(0.019)
B. Add schools from list		
Add from list (10)	0.117	0.069***
	(0.011)	(0.017)
Add outside list (10)	0.180	-0.028
	(0.013)	(0.018)
Add popular from list (10)	0.094	0.028**
	(0.013)	(0.014)
Add congested from list (10)	0.169	0.045**
	(0.015)	(0.019)
Placed in list (10)	0.127	0.005
	(0.015)	(0.019)
Enrolled in list (10)	0.234	-0.003
	(0.019)	(0.023)

Notes. This table shows the aggregate effect of the information intervention on the applicants. Column 1 compares the portfolios before and after treatment for all applicants that were assigned to T_1: warning message but no suggestion's list. Column 2 shows estimates of the differential effect of showing a list of 10 suggested schools in addition to the warning (T_2) compared to only showing the warning (T_1). "list (10)" is the list of 10 suggestions. The sample considers only applicants that opened the link to the report card.

Table 3.4: RCT Results: Effect of Suggestions on Application Outcomes by implmentation country/year.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Country	Ecuador 2021			Peru 2021			Peru 2022		
Intervention	T_1 Warning	T_2 + list (10)	T_{3-2021} + list (10) + info pop/cong	T_1 Warning	T_2 + list (10)	T_{3-2021} + list (10) + info pop/cong	T_1 Warning	T_2 + list (10)	T_{3-2022} + list (3)
	(base)	(diff.)	(diff.)	(base)	(diff.)	(diff.)	(base)	(diff.)	(diff.)
A. Choice behavior									
Add any school	0.097	0.005	-0.011	0.337	0.048	0.034	0.134	0.011	0.066*
	(0.018)	(0.025)	(0.023)	(0.026)	(0.036)	(0.036)	(0.025)	(0.038)	(0.038)
Number of schools added	0.242	0.109	-0.112*	0.896	0.123	0.004	0.330	0.119	0.163
	(0.053)	(0.117)	(0.062)	(0.087)	(0.126)	(0.118)	(0.079)	(0.144)	(0.171)
Add popular	0.052	-0.001	-0.023	0.253	0.057*	-0.020	0.121	-0.008	0.037
	(0.013)	(0.019)	(0.016)	(0.024)	(0.034)	(0.033)	(0.024)	(0.035)	(0.036)
Add congested	0.086	-0.006	-0.021	0.311	0.034	0.014	0.134	0.004	0.049
	(0.017)	(0.023)	(0.022)	(0.025)	(0.035)	(0.035)	(0.025)	(0.037)	(0.038)
B. Add schools from list									
Add from list (10)	0.082	0.005	-0.009	0.186	0.126***	0.095***	0.039	0.047*	
	(0.017)	(0.024)	(0.022)	(0.021)	(0.032)	(0.032)	(0.015)	(0.026)	
Add from list (3)							0.011		0.050***
							(0.008)		(0.019)
Add outside list (10)	0.064	-0.012	-0.036**	0.303	-0.050	-0.069**	0.127	-0.022	
	(0.015)	(0.020)	(0.017)	(0.025)	(0.034)	(0.033)	(0.024)	(0.035)	
Add popular from list (10)	0.033	0.005	-0.011	0.071	0.031	0.005	0.018	0.038*	
	(0.011)	(0.016)	(0.014)	(0.014)	(0.021)	(0.020)	(0.010)	(0.020)	
Add congested from list (10)	0.071	-0.022	-0.021	0.145	0.085***	0.070**	0.034	0.039	
	(0.016)	(0.020)	(0.020)	(0.019)	(0.029)	(0.029)	(0.014)	(0.024)	
C. Assignment and Enrollment Outcomes									
Placed in list (10)	0.241	-0.018	-0.018	0.055	0.030	0.010			
	(0.026)	(0.035)	(0.034)	(0.013)	(0.019)	(0.018)			
Enrolled in list (10)	0.376	-0.044	-0.046	0.138	0.031	0.012			
	(0.028)	(0.040)	(0.037)	(0.019)	(0.027)	(0.027)			

Notes. This table shows the effect of the information intervention on the applicants of Ecuador 2021 (columns 1 to 3) and Peru (columns 4 to 9). Columns 1, 4 and 7 compare the portfolios before and after treatment for all applicants that were assigned to T_1: warning message but no suggestion's list. Columns 2, 5 and 8 show estimates of the differential effect of showing a list of 10 suggested schools in addition to the warning (T_2) compared to only showing the warning (T_1). Columns 3 and 6 show estimates of the differential effect of showing a list of suggested schools with information on pupularity and congestion in addition to the warning (T_{3-2021}) compared to only showing the warning (T_1). Column 9 show estimates of the differential effect of showing a list of 3 suggested schools in addition to the warning (T_{3-2022}) compared to only showing the warning (T_1). "list (10)" is the list of 10 suggestions, while "list (3)" is the list of 3 suggestions. "info pop/cong" refers to the additional information on popularity and congestion showed for each school in the list. The sample considers only applicants that opened the link to the report card.

Bibliography

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

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

Abdulkadiroğlu, A., N. Agarwal, and P. A. Pathak (2017). The welfare effects of coordinated assignment: Evidence from the new york city high school match. *American Economic Review 107*(12), 3635–89.

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

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

Abdulkadiroğlu, A., Y.-K. Che, and Y. Yasuda (2011). Resolving conflicting preferences in school choice: The "boston mechanism" reconsidered. *American Economic Review 101*(1), 399–410.

Abdulkadiroğlu, A., P. A. Pathak, and A. E. Roth (2005). The new york city high school match. *American Economic Review 95*(2), 364–367.

Abdulkadiroğlu, A., P. A. Pathak, A. E. Roth, and T. Sönmez (2005). The boston public school match. *American Economic Review 95*(2), 368–371.

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

Abdulkadiroğlu, A., P. A. Pathak, and C. R. Walters (2018). Free to choose: Can school choice reduce student achievement? *American Economic Journal: Applied Economics 10*(1), 175–206.

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

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

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

Agarwal, N. and P. Somaini (2018). Demand Analysis Using Strategic Reports: An Application to a School Choice Mechanism. *Econometrica 86*(2), 391–444.

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

Agrawal, A., J. Gans, and A. Goldfarb (2018). *Prediction Machines: The Simple Economics of Artificial Intelligence*. Harvard Business Press.

Ainsworth, R., R. Dehejia, C. Pop-Eleches, and M. Urquiola (2023, April). 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. and M. Sidibe (2020). School Choice Under Imperfect Information. *SSRN Electronic Journal*.

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

Akbarpour, M., C. Neilson, A. Kapor, W. van Dijk, and S. Zimmerman (2020). Centralized school choice with unequal outside options. *Industrial Relations Section Working Paper*.

Allende, C. (2019). Competition Under Social Interactions and the Design of Education Policies. Job Market Paper.

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

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

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association 91*(434), 444–455.

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

Arteaga, F., G. Elacqua, C. Méndez, T. Krussig, and C. Neilson (2022). Can Information on School Attributes and Placement Probabilities Direct Search and Choice? Evidence from Choice Platforms in Ecuador and Peru. Working Paper.

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

Ashlagi, I. and A. Nikzad (2020). What matters in school choice tie-breaking? how competition guides design. *Journal of Economic Theory 190*, 105–120.

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

Balarin, M., C. Fontdevila, P. Marius, and M. F. Rodríguez (2019, October). Educating on a budget: The subsistence model of low-fee private schooling in Peru. *Education Policy Analysis Archives 27*, 132–132.

Barahona, N., S. Otero, and C. Dobbin (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., F. Molinari, and M. Thirkettle (2021, June). Discrete Choice under Risk with Limited Consideration. *American Economic Review 111*(6), 1972–2006.

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

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

Bobba, M. and V. Frisancho (2020). Self-perceptions about academic achievement: Evidence from mexico city. *Journal of Econometrics*.

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

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

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica 82*(6), 2295–2326.

Calsamiglia, C., C. Fu, and M. Güell (2020, February). Structural Estimation of a Model of School Choices: The Boston Mechanism versus Its Alternatives. *Journal of Political Economy 128*(2), 642–680.

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

Chen, Y. and Y. He (2020). Information acquisition and provision in school choice: an experimental study. Technical report.

Chen, Y. and Y. He (2021). Information acquisition and provision in school choice: a theoretical investigation. *Economic Theory*, 1–35.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014, September). Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *American Economic Review 104*(9), 2593–2632.

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

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

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

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

DellaVigna, S. and E. Linos (Forthcoming). Rcts to scale: Comprehensive evidence from two nudge units. *Econometrica.*

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

Doraszelski, U., K. Seim, M. Sinkinson, and P. Wang (2017). Ownership concentration and strategic supply reduction. Technical report, National Bureau of Economic Research.

Dynarski, S., C. Libassi, K. Michelmore, and S. Owen (2021). Closing the gap: The effect of reducing complexity and uncertainty in college pricing on the choices of low-income students. *American Economic Review 111*(6), 1721–56.

Elacqua, G., I. Jacas, T. Krussig, C. Méndez, and C. Neilson (2022). What they want or where they live: A welfare comparison of coordinated assignments focused on location vs choice in the case of Ecuador. Technical report.

Elacqua, G., I. Jacas, T. Krussig, C. Méndez, C. Neilson, A. Roman, and S. Soares (2022). Centralized Student Assignment Systems: Implementation Guide. Technical report, Inter-American Development Bank.

Enke, B. and F. Zimmermann (2019). Correlation neglect in belief formation. *The Review of Economic Studies 86*(1), 313–332.

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

Fernandes, D., J. G. Lynch, Jr., and R. G. Netemeyer (2014). Financial literacy, financial education, and downstream financial behaviors. *Management Science 60*(8), 1861–1883.

Gabaix, X. (2019). Behavioral inattention. In *Handbook of Behavioral Economics: Applications and Foundations 1*, Volume 2, pp. 261–343. Elsevier.

Gobierno de Chile Ministerio de Educación (2017, September). *El primer gran debate de la Reforma Educacional: Ley de Inclusión Escolar*.

Grenet, J., Y. He, and D. Kübler (2021). Decentralizing centralized matching markets: Implications from early offers in university admissions. *arXiv preprint arXiv:2107.01532*.

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

Gruber, J., B. Handel, S. Kina, and J. Kolstad (2020, April). Managing Intelligence: Skilled Experts and AI in Markets for Complex Products. Technical Report w27038, National Bureau of Economic Research, Cambridge, MA.

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

Gurantz, O., J. Howell, M. Hurwitz, C. Larson, M. Pender, and B. White (2021). A national-level informational experiment to promote enrollment in selective colleges. *Journal of Policy Analysis and Management 40*(2), 453–479.

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

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

Hakimov, R., D. Kübler, S. Pan, et al. (2021). Costly information acquisition in centralized matching markets. Technical report, CRC TRR 190 Rationality and Competition.

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

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

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

Hsieh, C.-T. and M. Urquiola (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., J. Leshno, I. Lo, and B. Lucier (2020). Information Acquisition in Matching Markets: The Role of Price Discovery. SSRN Working Paper.

Jin, G. Z. and P. Leslie (2003). The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards. *The Quarterly Journal of Economics 118*(2), 409–451.

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

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

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

Lavecchia, A. M., H. Liu, and P. Oreopoulos (2016). Behavioral economics of education: Progress and possibilities. In *Handbook of the Economics of Education*, Volume 5, pp. 1–74. Elsevier.

Madrian, B. C. (2014). Applying insights from behavioral economics to policy design. *Annu. Rev. Econ. 6*(1), 663–688.

Mani, A., S. Mullainathan, E. Shafir, and J. Zhao (2013). Poverty impedes cognitive function. *Science 341*(6149), 976–980.

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

Manzini, P. and M. Mariotto (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 P. Zarembka (Ed.), *Frontiers in Econometrics*, New York, pp. 105–142. Academic Press.

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

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

Milgrom, P. and I. Segal (2020). Clock auctions and radio spectrum reallocation. *Journal of Political Economy 128*(1), 1–31.

Narita, Y. (2018, June). 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.

Neilson, C. (2021). Targeted vouchers, competition among schools, and the academic achievement of poor students.

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

Pathak, P. A. and T. Sönmez (2013, February). School Admissions Reform in Chicago and England: Comparing Mechanisms by Their Vulnerability to Manipulation. *American Economic Review 103*(1), 80–106.

Rees-Jones, A., R. Shorrer, and C. J. Tergiman (2020). Correlation neglect in student-to-school matching. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pp. 467–468.

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

Son, S. J. (2020). Distributional impacts of centralized school choice. *Yale University Working Paper*.

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

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

Thaler, R. H. and W. Tucker (2013). Smarter information, smarter consumers. *Harvard Business Review 91*(1), 44–54.

Tincani, M. M., F. Kosse, and E. Miglino (2021). Subjective beliefs and inclusion policies: Evidence from college admissions.

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

Tversky, A. and D. Kahneman (1974, September). Judgment under Uncertainty: Heuristics and Biases. *Science 185*(4157), 1124–1131.

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

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

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

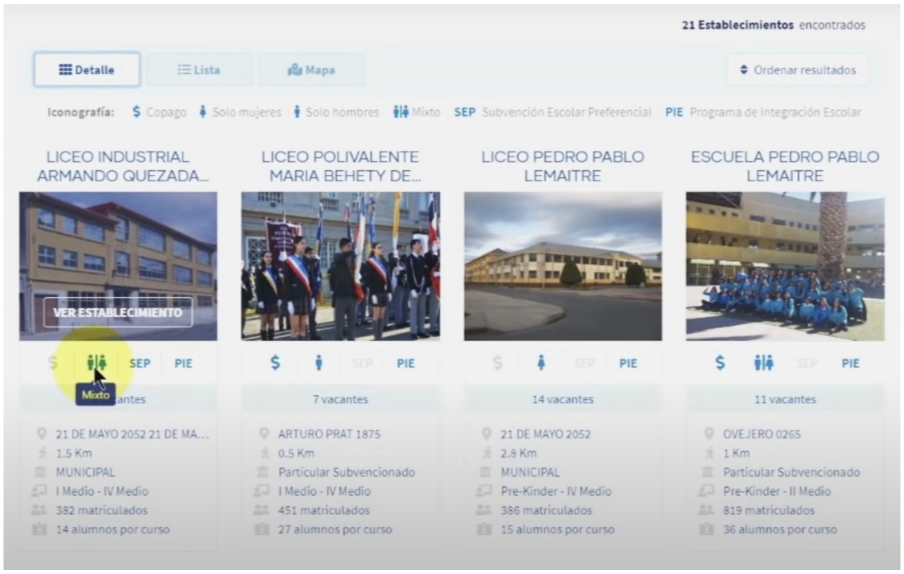
Appendix A

Appendix of Imperfect Information and Outside Options in Centralized School Choice

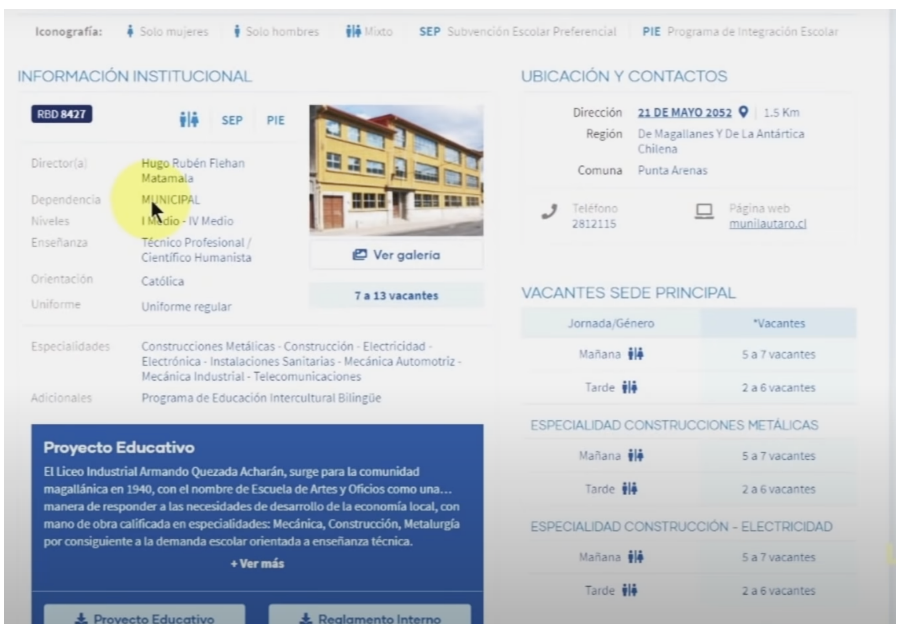
A.1 Additional Figures

Figure A.1: 2022 Online Application Platform Screenshots

(a) Gallery of schools

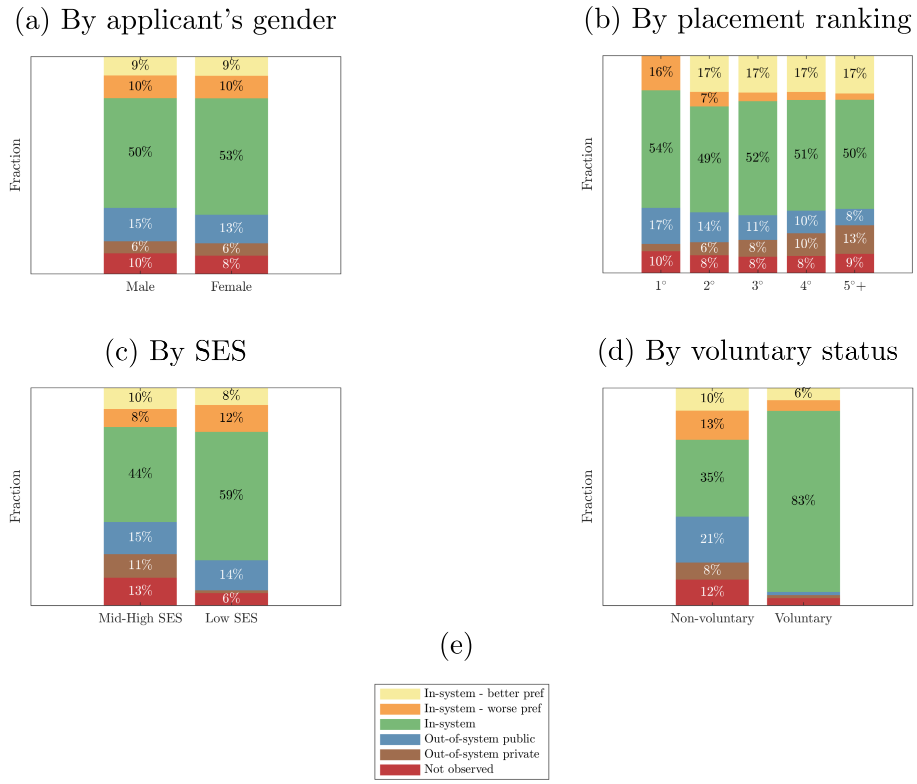


(b) Detailed Information of a School



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.

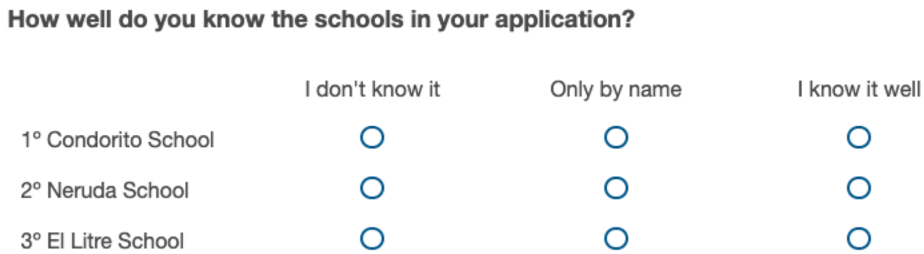
Figure A.2: Enrollment decision for non-compliers



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(does not) receive public funding.

Figure A.3: Screenshot of question about knowledge level of schools

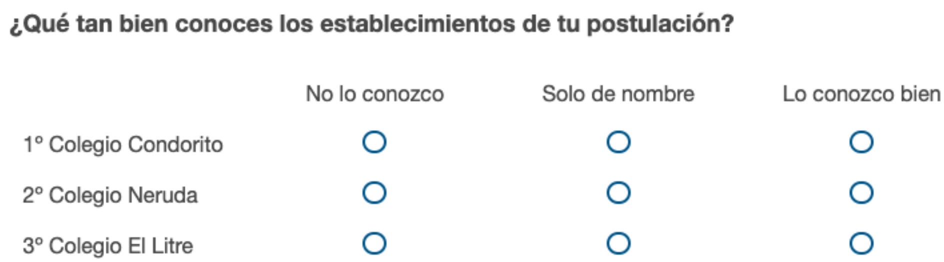
(a) Translated version to English



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	○	○	○
2° Neruda School	○	○	○
3° El Litre School	○	○	○

(b) Original version in Spanish



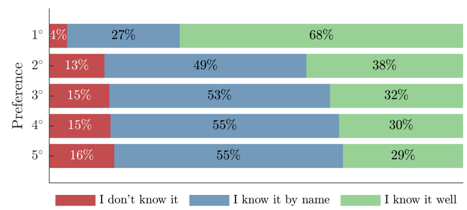
¿Qué tan bien conoces los establecimientos de tu postulación?

	No lo conozco	Solo de nombre	Lo conozco bien
1° Colegio Condorito	○	○	○
2° Colegio Neruda	○	○	○
3° Colegio El Litre	○	○	○

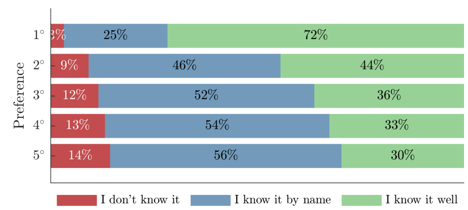
Notes: Figure shows how survey respondents saw the knowledge question. The survey was implemented in the Qualtrics platform.

Figure A.4: Knowledge level about schooling option by mother's education

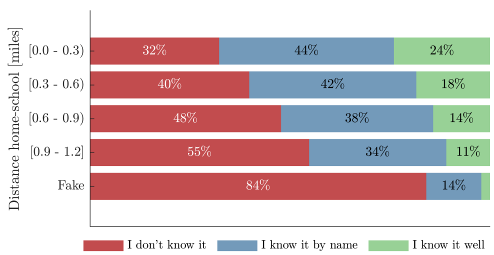
(a) Knowledge of ranked options - Low education



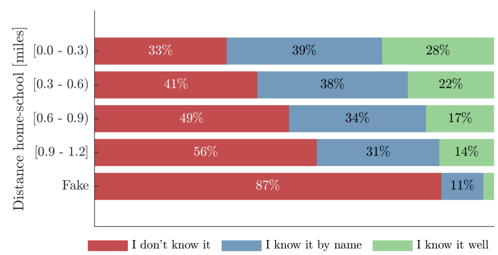
(b) Knowledge of ranked options - High education



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

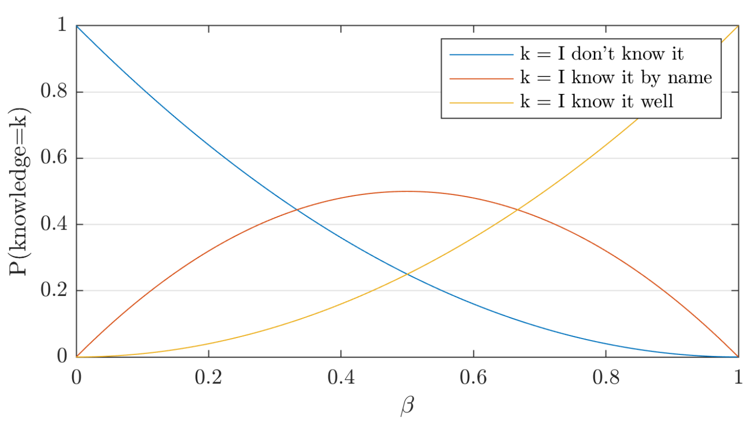


(d) Knowledge of non-ranked options - High education



Notes: "Low education" refers to families whose mother has at most secondary education (48% of the survey sample). "High education" refers to families whose mother has more than secondary education, ie complete or incomplete technical tertiary education or a college degree (52% of the survey sample). 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 A.5: Probability of knowledge by β



A.2 Additional Tables

Table A.1: 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 A.2: Descriptive Statistics for Schools

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

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

Table A.3: Counterfactual Results for Non-placed Applicants

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

Notes. 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 A.4: Counterfactual Results for Complier Placed in 2nd+ Preference

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

Notes. 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 A.5: Counterfactual Results for Complier Placed in 1st Preference

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

Notes. 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 A.6: Counterfactual Results for Non-complier

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

Notes. 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.

A.3 Simulated choice sets

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

The method uses simulation to approximate the integration over all potential choice sets. The procedure starts by calculating a consideration probability for each potential option for all applicants. Then, each simulated choice set is defined by a vector of iid uniform draws of length equal to the number of the potential options. If the draw is lower than the consideration probability, then the school is considered. Otherwise, it is not. Since the level of knowledge affects the utility in our framework, for the considered schools, we impute the knowledge using the prediction function described in Appendix A.7. In Sovinsky Goeree (2008) the consideration probabilities are calculated endogenously using advertisement measures as consideration shifters that don't affect choice probabilities. We use our survey data to estimate the consideration probability *offline*, approximating consideration with answers to our questions about knowledge of schools not in the ranking but in the neighborhood.

The detailed steps of the procedure are the following:

For each applicant i in the estimation sample:

1. Find the set of potential schools. Call J_i the cardinality of the set.
2. Predict the consideration probability $\hat{p}^c_{ij}$ with the function described on section A.3 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 $b_{ijs} = \mathbb{1}(\hat{p}^c_{ij} > u_{ijs})$.

Consideration probability of unranked school

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

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

ranking, we are gonna 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 1.5, and explained in detail in Appendix A.3.

To achieve this, we fit a binary logit model (Train, 2009) using the responses to the survey described in Section 1.3. We are going to assume that school is considered if the answer is 2: *I know it by name*, 3: *I know it well*, and not considered if the answer is 1: *I don't know it*. We assume that the consideration $c(i,j) \in \{0,1\}$ depends on an underlying continuous index C_{ij} defined as:

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

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

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

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

We observe the covariates that define C_{ij}, and we build c_ij from survey answers to the question "How well do you know the schools in your neighborhood?", with results summarized in Figure 1.3b in Section 1.3.

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

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

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

Given that $\epsilon_{ij} \sim Logistic(0,1)$, the probability $P(c(i,j) = a|\boldsymbol{\theta})$ has a simple analytical form once we condition on the vector θ, that contains the parameters that define C_{ij} and the threshold.[2] The log-likelihood function of observing c_{ij} for each school $j \in \mathcal{S}_i$[3] and survey respondent $i \in \{1 \ldots I\}$ is the following:

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

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

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

[2]See Train (2009) for details.

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

A.4 Counterfactual simulations

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

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 1.2 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 1.4 on the main article, and parameters related to the expected utility of the enrolled schools in stage 2 that are not present in stage 1 (λ and τ), displayed in Table 1.3.

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 A.7.

Simulation procedure

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

On each s simulation out of S:

- For each pair applicant-school $\{i, j\}$, $i \in \{1 \dots I\} \wedge j \in \mathcal{C}_i$

  1. Compute the observable part of the (indirect) expected utility of each school in the choice set ($\hat{V}_{ij}$), based on the estimated parameters, characteristics of student i and schools j
  2. Predict the knowledge level. We use the estimated ordered logit model to predict a probability for all three levels of knowledge: $\{\hat{p}^1_{ij}, \hat{p}^2_{ij}, \hat{p}^3_{ij}\}$. Then draw a knowledge level $\hat{k}(i, j)$ randomly from levels 1, 2 or 3 with probabilities $\{\hat{p}^1_{ij}, \hat{p}^2_{ij}, \hat{p}^3_{ij}\}$.

[4] This section is based on the explanation structure of Barahona et al. (2021).

[5] Home location for applicants with unreliable geocoding are imputed with the method detailed in Section A.4 of this appendix.

3. Simulate the noise component $\eta_{\hat{k}(i,j)} \sim N\left(0, \sigma_{\hat{k}(i,j)}\right)$
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 $\hat{EU}_{ij} = \hat{V}_{ij} + \hat{\eta}_{\hat{k}(i,j)} - \hat{\rho}^{s1}_{k(i,j)} + \hat{\epsilon}_{ij}$

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

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

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

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

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

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

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

C3: Information campaign

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

- Since we don't know the enrolled school for our simulated population of non-compliers, we impute $s(i)$ matching each simulated non-complier with a real non-complier based on geographic distance, and assume that the enrollment of the simulated non-complier will be the same as the matched real non-complier. We do this procedure on each stratum defined by gender, application grade, and geographic zone, following these steps:

  1. We count the amount of simulated and real non-compliers. If the set of simulated is larger, we bootstrap from the real until we get the same number.
  2. We first match all simulated applicants that share the same geolocation with a real applicant.[6]
  3. We generate a lottery for each remaining non-matched simulated non-complier. The simulated applicants are matched to the closest non-matched real applicant, following the order induced by the lottery.
  4. Then, we define $s(i)$ as the observed enrolled school of the matched real non-complier. It might be that $s(i) = 0$, that is the case when we observe the matched applicant enrolled none school.

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

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

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

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

- To analyze the impact of the "intensive margin", specifically, the extent to which families are informed about school $q(i)$, we examine varying levels of familial knowledge about the school the policy recommends. We introduce a parameter, β, to quantify this variation. At one extreme, where $\beta = 1$, families possess comprehensive knowledge about school $q(i)$, expressed as $k(i, q(i)) = 3$, eliminating any uncertainty penalization in the expected utility ($EU^{s1}_{iq(i)}$). Conversely, when $\beta = 0$, families possess only the predicted level of knowledge attributed to a school that does not feature in the rankings. The function to predict this knowledge level is estimated using survey data concerning non-ranked schools, as detailed in Appendix in A.7.

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.

Home location imputation procedure

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

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

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

A.5 Simulated maximum log-likelihood

We are interested in the parameters of our joint decision model, represented by the vector $\boldsymbol{\theta} = [\boldsymbol{\gamma}, \boldsymbol{\gamma}_X, \boldsymbol{\beta}_X, \boldsymbol{\beta}^\sigma, \boldsymbol{\delta}, \boldsymbol{\sigma}^\eta, \boldsymbol{\rho}, \boldsymbol{\psi}, \lambda, \tau]$. To estimate them we follow a log-likelihood maximization procedure. In Section 1.5 we defined the individual likelihood function, conditional on $\boldsymbol{\theta}$, as:

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

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

$$\begin{aligned}
\boldsymbol{\beta}^\sigma_i &= \boldsymbol{\phi}^\beta_i \cdot \boldsymbol{\sigma}^\beta & \boldsymbol{\phi}^\beta_i &\sim N(0, I_{|\boldsymbol{\beta}^\sigma|})\\
\eta_1 &= \phi^\eta_i \sigma^\eta_1 & \phi^\eta_i &\sim N(0,1)\\
\eta_2 &= \phi^\eta_i \sigma^\eta_2 & \phi^\eta_i &\sim N(0,1)
\end{aligned}$$

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

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

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

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

The vector $\hat{\boldsymbol{\theta}}$ represents our estimates. We calculate the covariance matrix as the inverse of the Fisher Information matrix, defined as the negative expectation of the Hessian matrix of the log-likelihood function. We use the outer product of the gradient (covariance matrix of the scores) to approximate the Hessian (Train, 2009).

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

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

- We start by generating all the draws necessary to approximate our step 1 integrals by simulation. Those correspond to the random parameters associated with preference heterogeneity ($\boldsymbol{\beta}^{\sigma}$) and the noise term ($\boldsymbol{\eta}$). We will use the same set of draws for step 1 and step 2.

- We estimate the parameters of the rank choice model (stage 1 in the model), i.e., without including the enrollment decision (stage 2 in the model), using simulated maximum likelihood. The simulated log-likelihood function that we maximize is:

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

- With the maximum likelihood estimates of the rank choice parameters in hand, for each applicant $i \in \{1 \ldots I\}$ and simulation $s \in \{1 \ldots S\}$ (i.e. conditional on $\boldsymbol{\beta}^{\sigma}_{is}$ and $\boldsymbol{\eta}_{ijs}$):

  1. We predict the observed part of the expected utility $V_{ijs} + \eta_{k(i,j)s} - \rho_{k(i,j)}$ using the estimated parameters and the generated draws ($\boldsymbol{\beta}^{\sigma}_{s}$ and $\boldsymbol{\eta}_{s}$) for every school in the ranking.
  2. We generate T approximate draws from $F(\epsilon_{ij} | \boldsymbol{\beta}^{\sigma}_{s}, \boldsymbol{\eta}_{s})$, performing the following procedure T times:

     a) We create a set of candidates $\{\hat{\epsilon}_{ijs}\}_{j \in \mathcal{C}_i}$ sampling $|\mathcal{C}_i|$ iid EVI draws.

     b) We use the candidates $\hat{\epsilon}_{ijs}$ to construct the expected utilities $\hat{EU}^{s1}_{ijs} = \hat{V}_{ijs} + \hat{\eta}_{k(i,j)s} - \hat{\rho}^{s1}_{k(i,j)} + \hat{\epsilon}_{ijs}$.

     c) We check if the constructed expected utilities are coherent with the ranking: $\hat{EU}^{s1}_{irs} > \hat{EU}^{s1}_{ijs} \quad \forall j > r, \forall r \in \mathcal{C}_i$. If the order of the constructed $\hat{EU}^{s1}_{ijs}$ matches the ranking, then we save our candidates $\{\hat{\epsilon}_{ijs}\}_{j \in \mathcal{C}_i}$ as a realization of the unobserved part of the expected utility of each school j in the ranking of i. If it does not match the ranking, we go back to step (a).[9]

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

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

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

A.6 Additional Proofs

Better outside option reduces the value of search

The benefit from search is represented by the expression:

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

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

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

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

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

A.7 Knowledge and consideration prediction functions

The counterfactuals detailed in Section 1.7 and the alternative choice set definition based on the work of Sovinsky Goeree (2008) explained in Section 1.5 rely on functions that predict a probability of knowledge level or consideration. This Section describes those functions.

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 1.3. The orderer discrete variable that we want to predict has three categories: 1: *I don't know it, 2: I know it by name, 3: I know it well* . We assume that the discrete level of knowledge $k(i,j) \in \{1 \dots 3\}$ depends on an underlying continuous index K_{ij} defined as:

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

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

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

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

We observe the covariates that define K_{ij}, and we collect $k(i,j)$ from survey answers to the question "How well do you know the schools in your application?", pictured in Figure A.3 on Appendix A.1, with results summarized in Figure 1.3a in Section 1.3.

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

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

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

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

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

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

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 1.7.

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

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

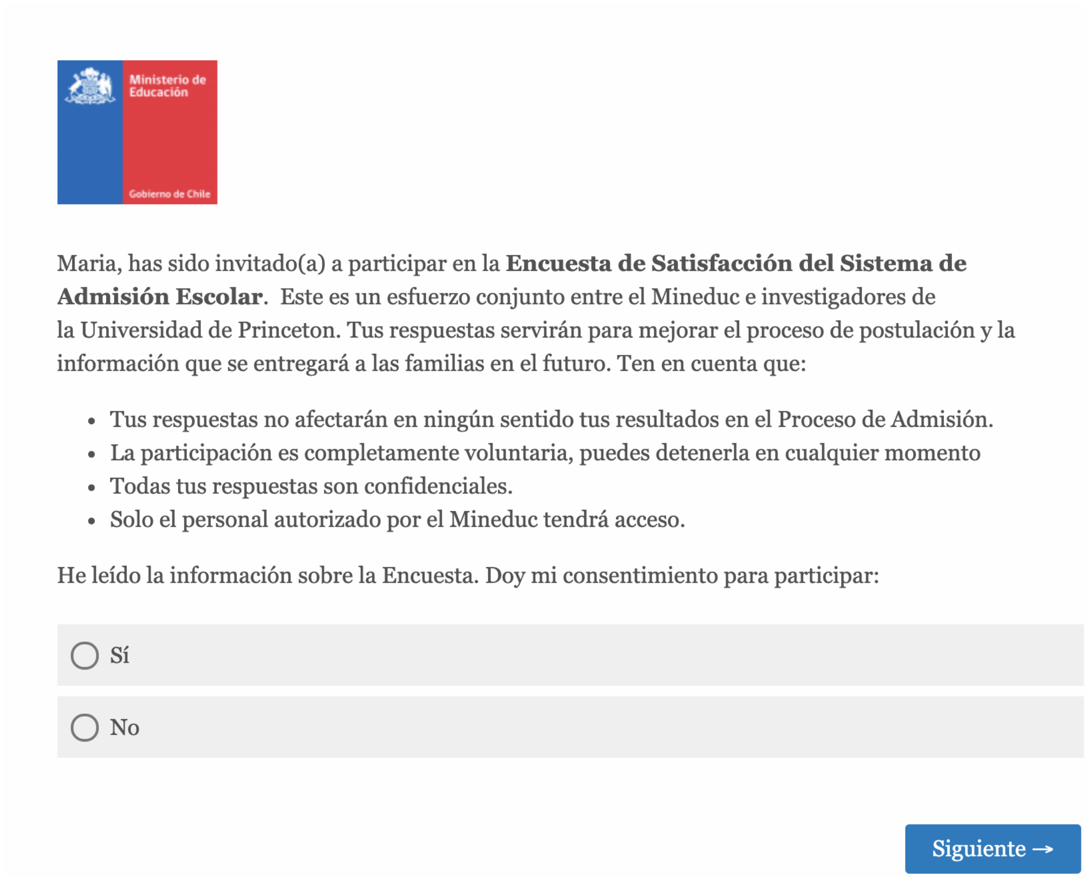
2. The data we used for the estimation comes from a different survey question. Besides asking for the knowledge of ranked options, we also asked about non-ranked options. We describe how we picked these options on Section 1.3.

[11] See Train (2009) for details.

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

A.8 Survey Translation

Figure A.6: 2020 Survey Landing Page



Ministerio de Educación

Gobierno de Chile

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

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

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

Sí

No

Siguiente →

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

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

a) Information on schools available (academic performance, collections, educational project, after school activities)
b) Availability of information on the application process (relevant dates, website, etc).
c) In general, what rating would you put to the application process?

3. How did you get information about of the application process? Select all that apply [*Select multiple*]
   a) Through the Municipality
   b) Through the current school/pre-school
   c) Through the newspaper or radio
   d) Through social networks (Facebook, Instagram, Twitter, Youtube)
   e) Through friends or relatives
   f) Through the website of the Ministry of Education
   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) Standarized test score
    c) Performance category
    d) Price
    e) Priority for economically-vulnerable students
    f) SAT scores
    g) Seats available
28. What is your preferred method of contact during the application process?
[*Select one*]
    a) E-mail
    b) Other
    c) SMS
    d) Telephone
    e) WhatsApp
29. What is your preferred method of contact during the application process? - Other
[*Open text*]
30. For registration purposes only, what is the highest educational level of the Mother (or Stepmother) of [applicant name]?
[*Select one*]
    a) Educación Básica Completa
    b) Educación Básica Incompleta
    c) Educación Media Completa
    d) Educación Media Incompleta
    e) Educación incompleta en una Universidad
    f) Grado de magíster universitario

g) No estudió

h) Titulada de un Centro de Formación Técnica o Instituto Profesional

i) Titulada de una Universidad

31. Do you know if [Field-nomPostulante] is a priority student (SEP)?
[*Select one*]

    a) He/she is not a beneficiary of the preferential subsidy

    b) I do not know

    c) He/she is a beneficiary of the preferential subsidy

32. Do you have any other comments, complaints or suggestions to make us?
[*Open text*]

Appendix B

Appendix of Smart Matching Platforms and Heterogeneous Beliefs in Centralized School Choice

B.1 Model Appendix

Additional Proofs

Note on Equation 2.3. Rewrite $V(\mathcal{C}) - V(\mathcal{C}_0)$ as a function of true beliefs and belief errors, so that

$$V(\mathcal{C}) - V(\mathcal{C}_0) = \Psi_r \times (1-a)^{r-1} \times (1 - R_s^*(1-a)) \times (u_s - \Gamma_r)$$

where $\Psi_r = \prod_{j<r} R_j^*$. The derivative of the log of $V - V_0$ follows.

Proof of proposition 1. By assumption, s is the last-ranked school in $\mathcal{C} = \mathcal{C}_0 \cup \{s\}$. This implies that $\Gamma_r = 0$ and that $\frac{d\Gamma_r}{da} = 0$, so we may ignore the third term of the sum in equation 2.3. Rewriting equation 2.3 without this term and setting it to be negative (because we are looking for conditions under which the value of adding s is decreasing in a), we have

$$\frac{d \log(V(\mathcal{C}) - V(\mathcal{C}_0))}{da} = \frac{1-r}{1-a} + \frac{R_s^*}{1 - R_s^*(1-a)} < 0.$$

Rearranging then yields

$$a > 1 - \frac{r-1}{rR_s^*}$$

Note also a) that $r = N_0 + 1 \geq 2$ (by assumption) and b) that $rR_s^* > 0$ (since both r and R_s^* are positive numbers). $\frac{r-1}{rR_s^*} > 0$ then follows immediately.

Proof of proposition 2. Let $\mathcal{C}$ denote a consideration set that an applicant obtains following an optimal search strategy under optimism level a. Additional search takes place when

$\kappa < U[\text{Search}|\mathcal{C}_0, a - \Delta_a]$. By optimality we have $\kappa > U[\text{Search}|\mathcal{C}_0, a]$. Hence the probability of additional search is equal to

$$\begin{aligned} Pr(\text{Search}; \mathcal{C}, a, \Delta_a) =& 1\left(U[\text{Search}|\mathcal{C}_0, a - \Delta_a] > U[\text{Search}|\mathcal{C}_0, a]\right) \\ & \times \left[\Phi(U[\text{Search}|\mathcal{C}_0, a - \Delta_a]) - \Phi(U[\text{Search}|\mathcal{C}_0, a])\right] \geq 0 \end{aligned}$$

Also note that search costs are immediately sunk once they are incurred. If an applicant searches and draws a school that has utility below the outside option value, he does not add it to the application, the value of future search is unchanged, and the applicant searches again. The implication is that the probability of adding at least one school is equal to the probability of search.

Finally, note that $U[\text{Search}|\mathcal{C}_0, a]$ is decreasing in $V(\mathcal{C}_0)$. The value of search declines as one adds more schools to the consideration set. Applicants for whom the expected value of search is negative given their current consideration set also have negative values of search for all larger consideration sets. $U[\text{Search}|\mathcal{C}_0, a] \leq \kappa$ is therefore a sufficient condition for terminating search.

Enrollment Choices and the Welfare Effects of Information.

We next consider how to use objects we observe in the data to assess the individual welfare effects of changes in optimism. We focus on changes in welfare accrued through the placement process; i.e. excluding search costs. The key insight here is that an applicant's decision to enroll in the school in which they are placed is a measure of how much they prefer that school to the outside option.

We model enrollment as a binary choice between the school where an individual is placed and the outside option. Timing is as follows. At the time of application, individuals observe school- (and person-) specific utilities μ_j, with the outside option normalized to zero. Following placement, they receive enrollment shocks ϵ_j, iid across schools. Students choose to enroll in the placed school j according to the rule

$$Enroll = 1[\mu_j + \epsilon_j > 0].$$

The utilities u_j defined above capture the expected value of placement at the time of application, so that $u_j = E[\max(\mu_j + \epsilon_j, 0)]$. Assume the ϵ_j are independent of μ_j and have distribution $G(\epsilon)$, which is differentiable with density function g and has an inverse that is also differentiable.

Let $q_j = Pr(\text{Enroll}|\text{place at } j)$ denote the probability of enrollment conditional on placement at j. An important observation is that, for each possible q_j, there is a unique utility level, $u^*(q_j)$, associated with enrollment probability q_j, and this utility level is increasing in q_j.[1]

[1] We have $q_j = 1 - G(-\mu_j)$ and $\mu_j = -G^{-1}(1 - q_j)$. Define $u^*(q_j) = E[\max(-G^{-1}(1 - q_j) + \epsilon_j, 0)]$. By construction, when utility is equal to $u^*(q_j)$, the associated enrollment probability is q_j. Note that

Our approach is to consider a surprise change in optimism from a to a' after an applicant has engaged in optimal search. Our framework lets us break down the effect of a change in information on payoffs into two channels: a placement channel, and a utility conditional on placement channel.

Before stating the result, we provide some definitions. Let $U^*(a', \mathcal{C}) = E(V(\mathcal{C}'); \mathcal{C}, a')$ denote the expected payoff an agent receives from his application after having followed an optimal search strategy, given an optimism level a', when endowed with consideration set $\mathcal{C}$, where the expectation is over the resulting consideration sets $\mathcal{C}'$ that may be obtained by further search.

Let $\mathcal{C}_0$ denote the initial consideration set that an applicant is endowed with. Let $\mathcal{C}_1$ be a consideration set that an applicant who is endowed with consideration set $\mathcal{C}_0$ reaches via an optimal search strategy under optimism a. By construction, an applicant with optimism a and consideration set $\mathcal{C}_1$ will not engage in further search. Hence we have $U^*(a, \mathcal{C}_1) = V(\mathcal{C}_1)$. However, when $a' \neq a$ the applicant may engage in additional search beyond the schools in $\mathcal{C}_1$.

Let $Pr(j|a', \mathcal{C})$ denote the probability that the person matches to school j when endowed with consideration set $\mathcal{C}$, optimism a', and the option to conduct further search if desired. Let $Pr(\text{place}|a', \mathcal{C}) = \sum_{j \in \mathcal{J}} Pr(j|a', \mathcal{C})$ denote the probability of any placement.

Proposition 3. Let $\mathcal{C}_1$ be a consideration set obtained by an optimal search strategy under optimism a. The individual utility gain from a change in optimism, $U^*(a', \mathcal{C}_1) - U^*(a, \mathcal{C}_1)$ where $a', a \in (0, 1)$, satisfies

$$U^*(a', \mathcal{C}_1) - U^*(a, \mathcal{C}_1) = [Pr(\text{place}|a', \mathcal{C}_1) - Pr(\text{place}|a, \mathcal{C}_1)]\, u^*(\overline{q}) + \sum_{j \in \mathcal{J}} [Pr(j|a', \mathcal{C}_1) - Pr(j|a, \mathcal{C}_1)]\, (u_j - u^*(\overline{q})),$$

where

$$\overline{q} = E(q_j|\text{place}, \mathcal{C}_1)$$

is the probability of enrolling in the inside option, conditional on receiving any placement, under consideration set $\mathcal{C}_1$. Moreover, for each $j \in \mathcal{J}$, the term $u_j - u^*(\overline{q})$ is nonnegative whenever $q_j \geq \overline{q}$.

Remark. This proposition shows that individual utility increases in proportion to the placement rate, except to the extent it is offset by declines in utility conditional on placement, where utility is an increasing function of enrollment probabilities.

This is particularly clear in our modal empirical case, which is when our intervention reduces optimism from a to $a' < a$ and the applicant adds a school $s \in \mathcal{J} \setminus \mathcal{C}_1$ to the bottom of his list. In this case, we would have $p_j(a, \mathcal{C}_1) = p_j(a', \mathcal{C}_1)$ for all $j \in \mathcal{C}_1$, and $p_s(a, \mathcal{C}_1) \leq p_s(a', \mathcal{C}_1)$ for $s \in \mathcal{J} \setminus \mathcal{C}_1$. When this happens, individual welfare increases at least

$\frac{dE[\max(\mu_j + \epsilon_j, 0)]}{d\mu_j} = Pr(\text{Enroll}|\text{place at j}) = q_j > 0$. We have

$$\frac{du^*(q_j)}{dq_j} = \frac{dE[\max(\mu_j + \epsilon_j, 0)]}{dPr(\text{Enroll}|\text{place at } j)} = \frac{Pr(\text{Enroll}|\text{place at } j)}{g(-\mu_j)} > 0.$$

proportionally to the placement rate as long as the probability of accepting a new placement satisfies $q_s \geq \overline{q}$.

We use this observation to guide our assessment of welfare effects.

Proof of proposition 3. We can write

$$\begin{aligned} U^*(a', \mathcal{C}_1) &= \sum_{j \in \mathcal{J}} Pr(j|a', \mathcal{C}_1) u^*(q_j) \\ &= \sum_{j \in \mathcal{J}} Pr(j|a', \mathcal{C}_1) u(\overline{q}) + \sum_{j \in \mathcal{J}} Pr(j|a', \mathcal{C}_1)(u_j - u^*(q)) \\ &= Pr(\text{place}|a', \mathcal{C}_1) u(\overline{q}) + \sum_{j \in \mathcal{J}} Pr(j|a', \mathcal{C}_1)(u_j - u^*(q)). \end{aligned}$$

The result follows. We have $u_j - u^*(\overline{q}) > 0$ if and only if $q_j > \overline{q}$ by monotonicity of $u^*(\cdot)$.

Additional Discussion

This subsection considers how violations of assumptions we impose in our theoretical model might mediate the effects of interventions that reduce optimism about application risk.

First, consider the possibility that applicants do not perfectly observe the utilities of schools they consider, but take schools' relative popularity as a signal of their utility. If so, informing an excessively optimistic applicant that the schools in his portfolio are more popular than he had thought may lead him to conclude that these schools are better, and schools outside his portfolio are worse, than he had previously believed, attenuating any increase in his incentives to search, and reducing the odds of placing a newly discovered school ahead of schools that the applicant already knows.

Second, our model makes the simplifying assumption that an applicant can surely discover a new school at a constant cost. Our assumption allows for uncertainty about the amount of effort needed to discover a school. For example, students may pay a flow cost k in order to discover an additional school with instantaneous probability λ. In this case, κ would denote the expected cost of searching until an additional acceptable school is found.

If search is uncertain, however, and in addition search costs are increasing or the chance of discovering a school is decreasing in the amount of effort that has already been exerted, then new information about placement chances may cause an applicant to engage in additional *unsuccessful* search before giving up. Thus this channel may also reduce the extent to which providing information to sufficiently optimistic applicants causes them to discover new schools.

B.2 Additional Figures and Tables

Table B.1: More Descriptive Statistics for Chilean Choice Applicants

	(1) All	(2) Economically Vulnerable	(3) Not Economically Vulnerable	(4) Pop-up eligible	(5) Risky (predicted risk>.3)	(6) Around Pop-up Cutoff	(7) RCT. sample (2020)	(8) Survey sample (2020)
N	1,168,706	575,521	593,185	848,795	233,678	84,517	19,213	48,929
%	1.00	0.49	0.51	0.73	0.20	0.07	0.02	0.04
A. Application behavior								
Add as last	0.18	0.16	0.20	0.18	0.38	0.26	0.27	0.22
Add to middle	0.03	0.02	0.03	0.02	0.03	0.03	0.03	0.03
Add as first	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.02
Change order	0.04	0.03	0.05	0.03	0.04	0.05	0.04	0.04
Change top 1	0.05	0.04	0.05	0.04	0.05	0.05	0.04	0.04
Delete any	0.05	0.04	0.05	0.04	0.03	0.04	0.04	0.04
Delete all	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
B. Placement								
Placed 2nd	0.13	0.13	0.14	0.13	0.12	0.19	0.12	0.14
Placed 3rd	0.06	0.06	0.07	0.06	0.08	0.10	0.07	0.06
C. School capacity available after placement								
Share of total seats	0.42	0.41	0.42	0.42	0.50	0.39	0.44	0.50
Share of seats in free schools	0.46	0.45	0.47	0.47	0.55	0.44	0.52	0.55
D. Attributes of enrolled school								
Value added	0.08	0.03	0.12	0.08	0.12	0.15	0.17	0.11
Total enrollment per grade	88.79	89.52	88.06	87.50	83.29	101.42	70.10	92.62
E. Classification by true risk of initial attempt								
.25 quantile |>0	0.28	0.23	0.32	0.29	0.52	0.15	0.41	0.30
.50 quantile |>0	0.62	0.56	0.66	0.64	0.79	0.28	0.63	0.65
.75 quantile |>0	0.92	0.91	0.93	0.94	0.99	0.42	0.87	0.94

Notes. N: 1,168,706 (20% from 2018, 41% from 2019 and 39% from 2020). Panels are as in Table 2.1. All statistics are means in the population defined by the column header. "Pop-up eligible" (col. 4) are students who submitted applications that received a risk prediction. "Risky" (col. 5) is applicants whose first attempt had a predicted risks > 0.3. "Around pop-up cutoff" (col. 6) are applicants whose first attempt had a predicted risk in [0.1,0.5]. "RCT sample" (column 7) is applicants in treatment or control group of the 2020 RCT design. "Survey sample" (column 8) is applicants who completed the 2020 school choice survey. Selected row variable definitions are as follows. "Economically vulnerable" is an SES measure computed by Mineduc. "Rural" is an indicator if students live in rural areas. "Length of initial/final attempt" is the number of schools on an applicants first and final choice application. "Total attempts" is the number of times an applicant submitted an application to the centralized system. Application change and addition variables describe the share of applicants making different kinds of changes applicants make between their first and final submission. "Placed in pref/1st/2nd/3rd" are indicators for any placement or for placement in the listed rank. "2nd round" variables describe participation and placement outcomes in the second centralized placement round. "Share of total seats/seats in free schools" is the share of seats in all schools/in schools without fees unfilled after the first application round in a student's local market. Value added and school characteristic variables described in Online Appendix B.4. VA is calculated only for grades 8 and below. True risk of initial attempt variables describe the nonplacement risk of an applicant's initial application, evaluated using ex post observed applications.

Table B.2: Descriptive Statistics for Chilean Choice Applicants- Alternate Samples

	(1) All	(2) Around Pop-up Cutoff	(3) RCT. sample (2020)	(4) Survey sample (2020)
N	1,168,706	84,517	19,213	48,929
%	1.00	0.07	0.02	0.04
A. Demographics				
Economically Vulnerable	0.49	0.42	0.25	0.42
Rural	0.05	0.02	0.00	0.04
B. Application behavior				
Length initial attempt	2.77	3.04	2.79	2.74
Length final attempt	3.14	3.57	3.32	3.22
Total attempts	1.41	1.51	1.53	1.45
Any modification	0.25	0.33	0.33	0.28
Add any	0.21	0.30	0.30	0.25
C. Placement				
Placed in pref.	0.79	0.77	0.42	0.79
Placed 1st	0.54	0.39	0.17	0.53
Particip. in 2nd round	0.09	0.12	0.20	0.09
Placed in 2nd round	0.07	0.09	0.16	0.07
D. School capacity available after placement				
Share of total seats	0.42	0.39	0.44	0.50
Share of seats in free schools	0.46	0.44	0.52	0.55
E. Attributes of enrolled school				
Enrolled at some school	0.97	0.96	0.91	0.97
Enrolled at placed	0.62	0.57	0.32	0.66
Have value added measure|grade$\leq$8	0.77	0.76	0.76	0.79
Value added|enrolled at placed	0.11	0.20	0.25	0.12
Value added|not enrolled at placed	0.04	0.06	0.11	0.09
School monthly fee (USD)	17.02	23.50	30.66	20.53
Share of vulnerable students	0.61	0.56	0.52	0.57
F. Classification by true risk of initial attempt				
Mean risk	0.24	0.24	0.61	0.25
Zero risk	0.59	0.19	0.02	0.59
Risky (risk>.3)	0.30	0.37	0.84	0.31

Notes. N: 1,168,706 (20% from 2018, 41% from 2019 and 39% from 2020). All statistics are means in the population defined by the column header. "Around pop-up cutoff" (col. 2) are applicants whose first attempt had a predicted risk in [0.1,0.5]. "RCT sample" (column 3) is applicants in treatment or control group of the 2020 RCT design. "Survey sample" (column 4) is applicants who completed the 2020 school choice survey. Selected row variable definitions are as follows. "Economically vulnerable" is an SES measure computed by Mineduc. "Rural" is an indicator if students live in rural areas. "Length of initial/final attempt" is the number of schools on an applicants first and final choice application. "Total attempts" is the number of times an applicant submitted an application to the centralized system. Application change and addition variables describe the share of applicants making different kinds of changes applicants make between their first and final submission. "Placed in pref/1st/2nd/3rd" are indicators for any placement or for placement in the listed rank. "2nd round" variables describe participation and placement outcomes in the second centralized placement round. "Share of total seats/seats in free schools" is the share of seats in all schools/in schools without fees unfilled after the first application round in a student's local market. Value added and school characteristic variables described in Online Appendix B.4. VA is calculated only for grades 8 and below. True risk of initial attempt variables describe the nonplacement risk of an applicant's initial application, evaluated using ex post observed applications. Panel F variables (School capacity available after placement) are calculated at a local market level defined for each student.

Table B.3: Platform Pop-Up RD Estimates of Main Outcomes (Table 2.2) with Alternate Bandwidths

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Bandwidth	Full	+-0.1	rdbwselect				
	Estimate	Estimate	Estimate	BW left	BW right	N left	N right
A. Balance							
Economically Vulnerable	-0.004	-0.004	-0.024	0.07	0.07	13,853	14,095
	(0.007)	(0.010)	(0.012)				
Rural	-0.000	-0.007	-0.013	0.05	0.05	10,794	11,078
	(0.002)	(0.003)	(0.003)				
B. Choice Behavior							
Any modification	0.206	0.214	0.213	0.10	0.10	20,863	21,697
	(0.007)	(0.010)	(0.010)				
Add any	0.210	0.216	0.217	0.09	0.09	19,280	19,947
	(0.007)	(0.010)	(0.010)				
Schools Added	0.335	0.340	0.343	0.13	0.13	26,024	26,546
	(0.018)	(0.026)	(0.023)				
Δ Risk	-0.031	-0.033	-0.034	0.10	0.10	20,746	21,614
	(0.002)	(0.003)	(0.003)				
Add as first	-0.002	-0.003	-0.002	0.10	0.10	20,725	21,565
	(0.002)	(0.003)	(0.003)				
Add to middle	0.014	0.017	0.017	0.08	0.08	16,250	16,695
	(0.003)	(0.004)	(0.004)				
Add as last	0.203	0.205	0.205	0.14	0.14	29,060	29,434
	(0.006)	(0.009)	(0.008)				
Drop any	-0.001	-0.001	-0.001	0.10	0.10	21,241	22,045
	(0.003)	(0.004)	(0.004)				
Re-order	0.011	0.014	0.022	0.07	0.07	13,645	13,947
	(0.003)	(0.005)	(0.006)				
C. Choice outcome							
Placed to preference	0.035	0.038	0.040	0.09	0.09	18,068	18,676
	(0.006)	(0.009)	(0.009)				
Enrolled in placed	0.022	0.024	0.024	0.08	0.08	15,879	16,276
	(0.008)	(0.010)	(0.012)				
Enrolled in placed|placed	-0.006	-0.006	-0.003	0.11	0.11	18,177	18,105
	(0.008)	(0.011)	(0.010)				
D. Congestion-related outcomes							
Add any undersubscribed	0.074	0.073	0.076	0.15	0.15	31,104	31,361
	(0.005)	(0.007)	(0.005)				
Δ prob. placed to undersubscribed	0.019	0.019	0.019	0.09	0.09	18,826	19,513
	(0.002)	(0.003)	(0.003)				
N left	71,075	20,359					
N right	166,699	21,145					

Notes. Local linear and full sample quadratic polynomial RD estimates of pop-up effects from warning pop-up on application platform. Computed using triangular kernel with different bandwidths. “Full” bandwidth uses 2nd order polynomial fit, while ”+-0.1” and rdbwselect uses 1st order (local) polynomial. Heteroskedasticity-robust nearest neighbor variance estimator with minimum of 3 neighbors reported in parentheses; computed as in Calonico et al. (2014).

Table B.4: RD Estimates of Platform Pop-Up Effects on Adding Any School, by City and Year

City	2020 applicants	2018	2019	2020
Santiago	158,057		0.24	0.25
			(0.04)	(0.02)
Viña - Valparaíso	26,215	0.01	0.28	0.22
		(0.08)	(0.07)	(0.05)
Concepción - Talcahuano	24,548	0.21	0.15	0.25
		(0.08)	(0.06)	(0.05)
Coquimbo - La Serena	13,994	0.18	0.38	0.11
		(0.10)	(0.10)	(0.07)
Rancagua	11,971	0.16	0.09	0.06
		(0.10)	(0.09)	(0.07)
Antofagasta	12,722	0.24	0.36	0.23
		(0.14)	(0.09)	(0.07)
Iquique - Alto Hospicio	10,251	0.25	0.23	0.25
		(0.09)	(0.09)	(0.07)
Temuco	10,176	0.22	0.31	0.29
		(0.10)	(0.08)	(0.06)
Puerto Montt - Puerto Varas	8,864	0.31	0.28	-0.02
		(0.15)	(0.08)	(0.09)
Talca - San Clemente	8,913	-0.03	0.11	0.17
		(0.13)	(0.09)	(0.07)
Arica	5,905	0.10	0.48	0.14
		(0.16)	(0.12)	(0.13)
Curicó	6,827	0.11	0.26	0.18
		(0.15)	(0.14)	(0.10)
Chillán	5,536	0.39	0.21	0.09
		(0.26)	(0.10)	(0.09)
Los Andes - San Felipe	5,006	0.11	0.03	0.42
		(0.32)	(0.24)	(0.13)
Los Ángeles	5,477	0.45	0.02	0.34
		(0.13)	(0.16)	(0.11)
Calama	5,565	0.00	0.32	0.08
		(0.21)	(0.17)	(0.10)
Copiapó	6,181	0.23	0.53	0.33
		(0.13)	(0.11)	(0.08)
Osorno	4,542	0.04	0.25	0.23
		(0.12)	(0.16)	(0.16)
Valdivia	4,599	0.10	0.37	0.13
		(0.23)	(0.12)	(0.18)
Algarrobo a San Antonio	4,705	0.43	-0.10	0.45
		(0.15)	(0.16)	(0.11)
Chile	454,226	0.18	0.22	0.22
		(0.02)	(0.02)	(0.01)

Notes. RD estimates of smart platform pop-up effects on adding at least one school to the choice application, split by city and year. Cities are sorted by count of 2020 applicants. Santiago is not displayed for 2018 because centralized admission had not yet been rolled out. Estimates from local linear specifications, computed using triangular kernel with bandwidth 0.1. Heteroskedasticity-robust nearest neighbor variance estimator with minimum of 3 neighbors reported in parentheses; computed as in Calonico et al. (2014). See section 2.5 for details.

Table B.5: RD Estimates of Platform Pop-Up Effects by Market-Level Choice Experience

	(1)	(2)	(3)	(4)	(5)	(6)
	1st year		2nd year		3rd+ year	
		IV		IV		IV
A. Balance						
Economically Vulnerable	-0.011		0.023		-0.029	
	(0.024)		(0.015)		(0.015)	
Rural	-0.001		-0.008		-0.009	
	(0.006)		(0.004)		(0.004)	
B. Choice Behavior						
Any modification	0.181		0.235		0.206	
	(0.023)		(0.016)		(0.015)	
Add any	0.194		0.232		0.210	
	(0.022)		(0.015)		(0.014)	
Schools Added	0.319	1.648	0.370	1.596	0.318	1.517
	(0.075)	(0.311)	(0.043)	(0.142)	(0.031)	(0.104)
Δ Risk	-0.037	-0.190	-0.032	-0.138	-0.033	-0.159
	(0.009)	(0.039)	(0.005)	(0.018)	(0.004)	(0.018)
Add as first	-0.011	-0.056	-0.004	-0.019	0.003	0.015
	(0.007)	(0.039)	(0.005)	(0.020)	(0.004)	(0.017)
Add to middle	0.028	0.147	0.017	0.072	0.012	0.057
	(0.010)	(0.049)	(0.006)	(0.026)	(0.006)	(0.027)
Add as last	0.184	0.951	0.222	0.956	0.197	0.940
	(0.021)	(0.050)	(0.015)	(0.026)	(0.014)	(0.028)
Drop any	0.004	0.021	-0.008	-0.033	0.004	0.021
	(0.010)	(0.052)	(0.007)	(0.029)	(0.006)	(0.030)
Re-order	0.007	0.038	0.021	0.091	0.010	0.046
	(0.011)	(0.059)	(0.008)	(0.033)	(0.007)	(0.033)
C. Choice outcome						
Placed to preference	0.059	0.306	0.049	0.211	0.017	0.083
	(0.023)	(0.119)	(0.014)	(0.060)	(0.013)	(0.063)
Enrolled in placed	0.009	0.046	0.041	0.177	0.014	0.067
	(0.025)	(0.130)	(0.016)	(0.071)	(0.016)	(0.077)
Enrolled in placed|placed	-0.045	-0.214	0.006	0.025	0.000	0.001
	(0.028)	(0.131)	(0.017)	(0.069)	(0.015)	(0.066)
D. Congestion-related outcomes						
Add any undersubscribed	0.059	0.303	0.075	0.322	0.079	0.376
	(0.015)	(0.068)	(0.011)	(0.040)	(0.010)	(0.041)
Δ prob. placed to undersubscribed	0.025	0.127	0.015	0.065	0.021	0.101
	(0.009)	(0.042)	(0.005)	(0.021)	(0.005)	(0.021)
NL	3,819	3,819	8,573	8,573	7,967	7,967
NR	3,571	3,571	8,880	8,880	8,694	8,694

Notes. Local linear RD estimates of pop-up effects from warning pop-up on application platform, split by years elapsed since city-grade combination first began using the centralized choice process. Computed using triangular kernel with bandwidth 0.1. Heteroskedasticity-robust nearest neighbor variance estimator with minimum of 3 neighbors reported in parentheses; computed as in Calonico et al. (2014). IV estimates reported in second column of each set show the instrumental variable specifications (fuzzy RD), where the endogenous regressor is the add any school indicator. Panel A: predetermined covariates. Panel B: measures of choice behavior from initial to final application. Δ risk is change in application risk from first to final attempt. "Add to X" are additions of schools in given place on list, relative to initial application submission. Panel C: outcomes of choice process. "Enrolled in placed" is equal to one for students who receive a placement and enroll in the placed school. "Enrolled in placed | placed" is the enrollment rate in the placed school, conditional on receiving a placement. Panel D: congestion attributes of behavior and placement outcomes. "Undersubscribed" schools are those with excess capacity.

Table B.6: RD Estimates of Platform Pop-Up Effects by Applicant's Socioeconomic Status

	(1)	(2)	(3)	(4)
	Economically Vulnerable		Not Economically vulnerable	
		IV		IV
A. Balance				
Economically Vulnerable	0		0	
	(0.000)		(0.000)	
Rural	-0.012		-0.004	
	(0.004)		(0.003)	
B. Choice Behavior				
Any modification	0.225		0.206	
	(0.015)		(0.013)	
Add any	0.227		0.209	
	(0.015)		(0.012)	
Schools Added	0.327	1.445	0.350	1.673
	(0.035)	(0.115)	(0.036)	(0.133)
Δ Risk	-0.041	-0.179	-0.028	-0.136
	(0.005)	(0.020)	(0.004)	(0.016)
Add as first	-0.001	-0.005	-0.004	-0.017
	(0.004)	(0.020)	(0.003)	(0.016)
Add to middle	0.013	0.057	0.020	0.095
	(0.005)	(0.023)	(0.006)	(0.025)
Add as last	0.217	0.957	0.197	0.943
	(0.014)	(0.025)	(0.012)	(0.025)
Drop any	-0.003	-0.014	0.001	0.006
	(0.006)	(0.028)	(0.006)	(0.027)
Re-order	0.020	0.090	0.009	0.045
	(0.007)	(0.031)	(0.006)	(0.031)
C. Choice outcome				
Placed to preference	0.019	0.085	0.052	0.251
	(0.014)	(0.062)	(0.012)	(0.056)
Enrolled in placed	0.008	0.036	0.036	0.173
	(0.016)	(0.071)	(0.014)	(0.067)
Enrolled in placed|placed	-0.009	-0.035	-0.004	-0.017
	(0.016)	(0.065)	(0.014)	(0.062)
D. Congestion-related outcomes				
Add any undersubscribed	0.071	0.315	0.075	0.359
	(0.011)	(0.040)	(0.009)	(0.035)
Δ prob. placed to undersubscribed	0.020	0.090	0.018	0.088
	(0.005)	(0.022)	(0.004)	(0.018)
NL	8,878	8,878	11,481	11,481
NR	8,721	8,721	12,424	12,424

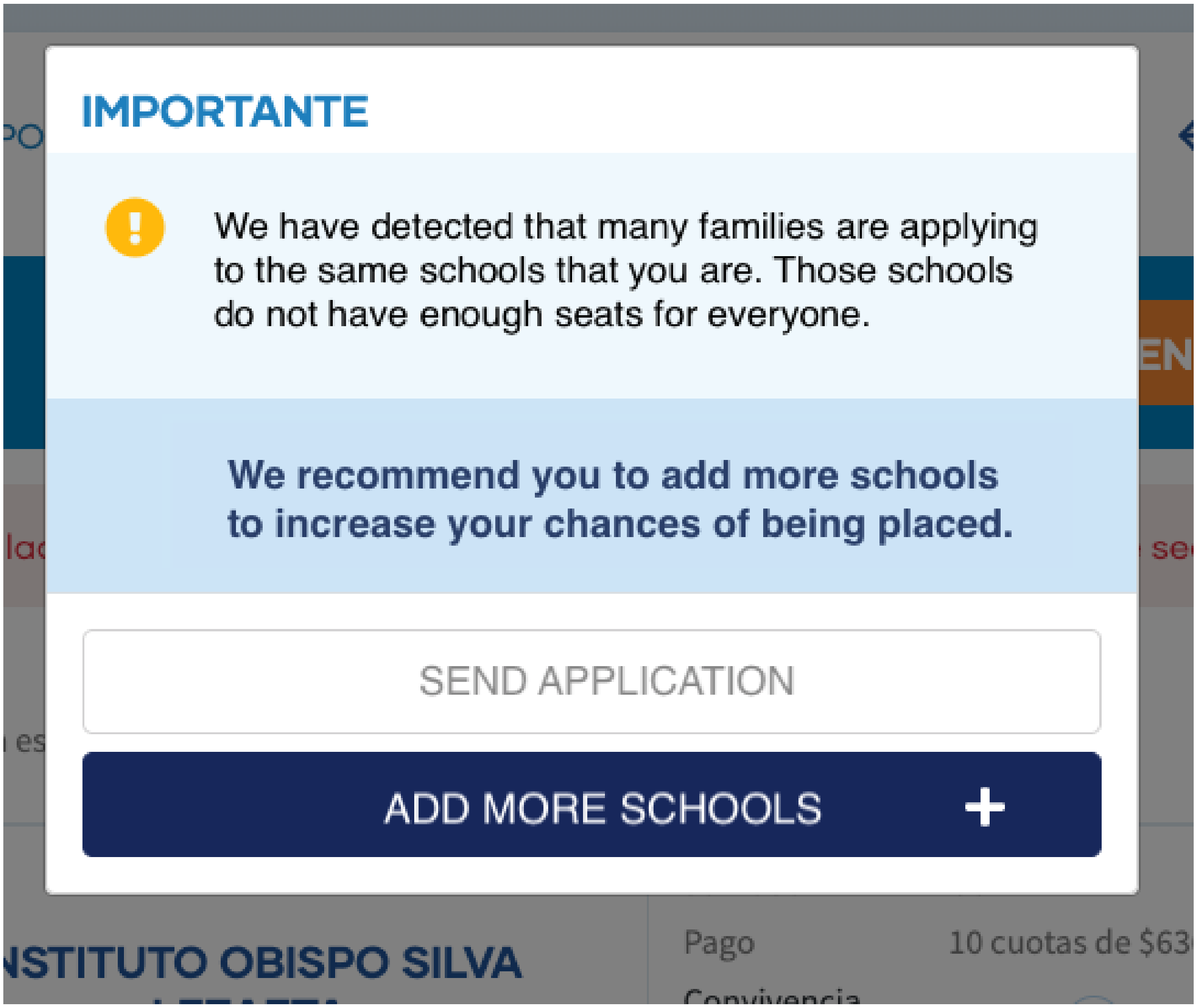
Notes. Local linear RD estimates of pop-up effects from warning pop-up on application platform, split by applicant's socioeconomic status. "Economically vulernable" individuals are the lower-SES group. See section 2.3 for details. Computed using triangular kernel with bandwidth 0.1. Heteroskedasticity-robust nearest neighbor variance estimator with minimum of 3 neighbors reported in parentheses; computed as in Calonico et al. (2014). IV estimates reported in second column of each set show the instrumental variable specifications (fuzzy RD), where the endogenous regressor is the add any school indicator. Panel A: predetermined covariates. Panel B: measures of choice behavior from initial to final application. Δ risk is change in application risk from first to final attempt. "Add to X" are additions of schools in given place on list, relative to initial application submission. Panel C: outcomes of choice process. "Enrolled in placed" is equal to one for students who receive a placement and enroll in the placed school. "Enrolled in placed \| placed" is the enrollment rate in the placed school, conditional on receiving a placement. Panel D: congestion attributes of behavior and placement outcomes. "Undersubscribed" schools are those with excess capacity.

Table B.7: RD Estimates of Platform Pop-Up Effects on Enrolled School Outcomes by Applicant's Socioeconomic Status

	(1)	(2)	(3)	(4)	(5)	(6)
	Economically Vulnerable			Not Economically Vulnerable		
		IV	$E[Y\|X=0.3^{-}]$		IV	$E[Y\|X=0.3^{-}]$
A. First stage and enrollment						
Add any	0.227		0.216	0.209		0.187
	(0.015)			(0.012)		
Enrolled	-0.009		0.979	-0.001		0.956
	(0.005)			(0.006)		
Have value added measure|grade≤8	0.061		0.730	-0.015		0.769
	(0.017)			(0.013)		
B. Attributes of enrolled school						
Distance (km)	-0.043	-0.214	3.294	0.112	0.516	2.841
	(0.466)	(2.345)		(0.281)	(1.292)	
Value added|grade≤8	0.001	0.006	0.103	0.034	0.151	0.160
	(0.018)	(0.091)		(0.014)	(0.061)	
Per teacher spending (1000USD)	1.202	5.992	30.224	0.519	2.361	30.926
	(0.358)	(1.855)		(0.289)	(1.327)	
Per student spending (1000USD)	0.001	0.005	2.252	0.002	0.008	2.240
	(0.022)	(0.112)		(0.020)	(0.091)	
With copayment fee	0.054	0.240	0.220	0.017	0.080	0.326
	(0.013)	(0.061)		(0.013)	(0.062)	
School monthly fee (USD)	3.925	17.390	15.534	0.507	2.378	26.851
	(1.043)	(4.794)		(1.202)	(5.639)	
Share of vulnerable students	-0.009	-0.040	0.604	-0.004	-0.018	0.539
	(0.004)	(0.018)		(0.004)	(0.018)	
Total enrollment per grade	14.382	63.825	104.184	4.702	21.964	94.904
	(2.758)	(12.771)		(2.193)	(10.347)	
NL	8,629	8,629		10,921	10,921	
NR	8,439	8,439		11,783	11,783	

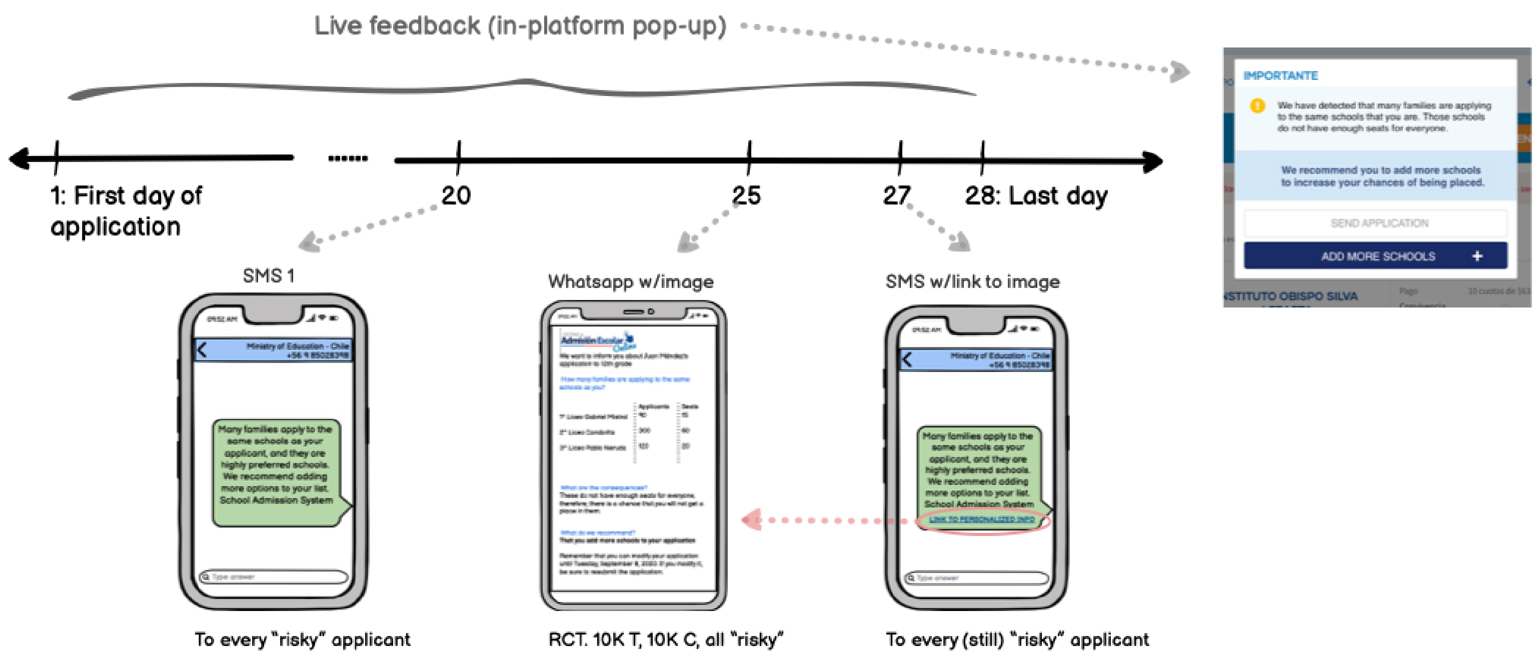
Notes. Local linear RD estimates of popup effects from warning popup on application platform. Computed using triangular kernel with bandwidth 0.1. Heteroskedasticity-robust nearest neighbor variance estimator with minimum of 3 neighbors reported in parentheses; computed as in Calonico et al. (2014). We report estimates split by whether students are economically vulnerable or not. IV estimates in columns 2 and 5 report instrumental variable specifications where the endogenous regressor is the "add any school" indicator. Columns 3 and 6 report below-cutoff means of the variable listed in the row in the analysis sample. Sample for value added outcomes is restricted to grades eight and below. Reported sample sizes are counts of enrolling students. See section 2.5 for discussion and Online Appendix B.4 for detailed variable definitions.

Figure B.1: Platform Pop-Up Intervention– 2018 and 2019



Notes. English translation of pop-up feedback shown to risky applicants on the application platform in 2018 and 2019. All applicants with predicted nonplacement risk of 30% or higher received this warning when they submitted their choice application. See section 2.3 for details.

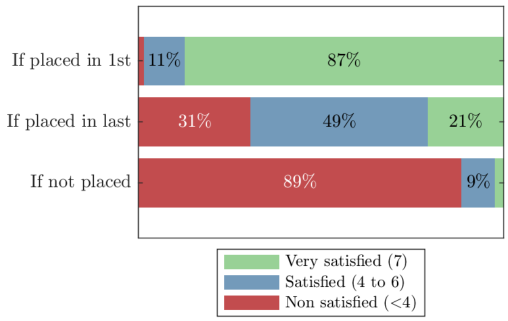
Figure B.2: Timeline of Feedback Interventions– 2020



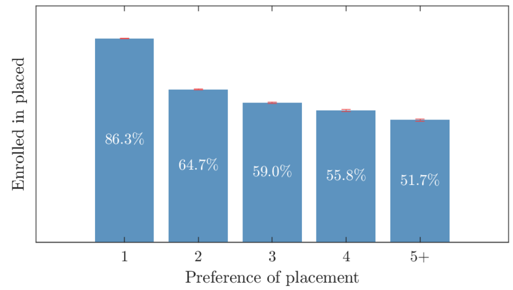
Notes. Sequence of 2020 application feedback for risky applicants. All text translated to English. The platform pop-up on the right was shown to all risky applicants at the time they submitted their application. The SMS and WhatsApp messages shown at center were sent to (subgroups of) still-risky applicants based on contemporaneous risk predictions on the day of the application cycle listed on horizontal axis, where day 28 is the final deadline for application submission. The SMS messages on day 20 and 27 were sent to all risky applicants, while the WhatsApp image at center was sent to randomly selected applicants on day 25. The schools displayed in the WhatsApp image are those the student listed on her choice application. See section 2.3 for details.

Figure B.3: Satisfaction with Placement Outcomes

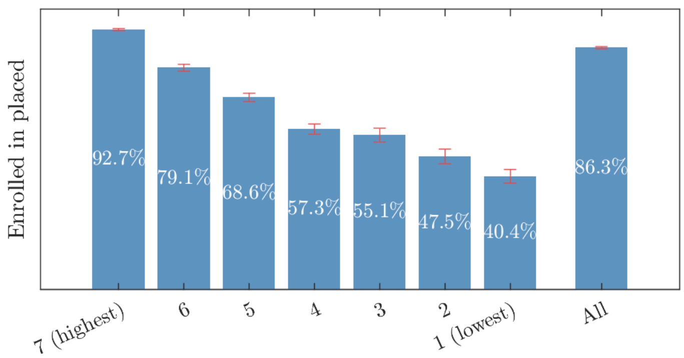
(a) Satisfaction with Placement by Rank



(b) Enrollment Rate by Preference



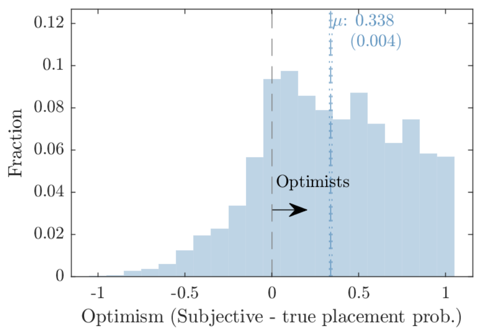
(c) Enrollment Rate by Declared Satisfaction



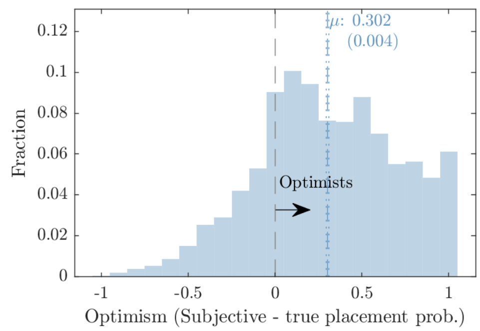
Notes. Panel A: stated satisfaction with hypothetical placement outcomes. Data are survey responses to questions about applicant satisfaction with being placed at their first-ranked school, last-ranked school, and nonplacement. Sample: survey completers. Results reported on a 1-7 scale, with 7 being very satisfied and 1 being not at all satisfied. Panel B: rates at which students enroll in the placed school, by rank of placed school. Unplaced students are not included. Sample: all placed students. Panel C: rate at which students enroll in the placed school, by survey reports of satisfaction with the placed school. Sample: survey completers who place in their first- or last-ranked school. See section 2.4 for details.

Figure B.4: Alternate Application Risk Framings

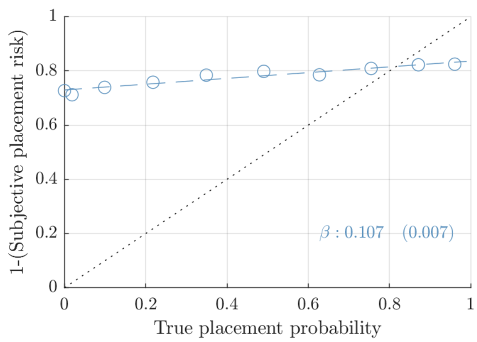
(a) Optimism– Negative Frame



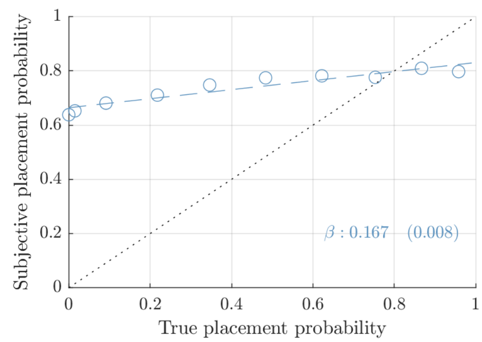
(b) Optimism– Positive Frame



(c) Subjective vs. True Beliefs– Negative Frame



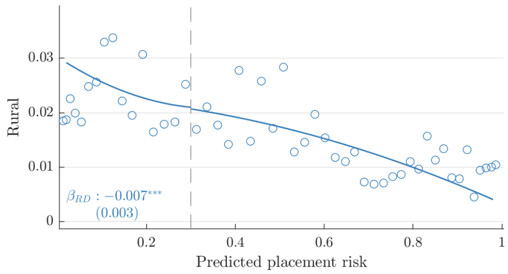
(d) Subjective vs. True Beliefs– Positive Frame



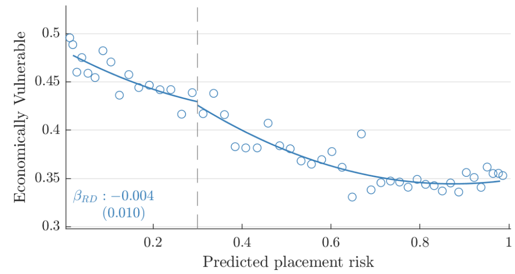
Notes. Panel A: distribution of optimism under "negative" framing for subjective risk question. Negative framing asks respondents about risk of non-placement under their submitted application. Panel B: distribution of optimism under "positive" framing for subjective risk question. Positive framing asks respondents about chance of placement under submitted application. Panel C: Binscatter of true placement probability vs. subjective placement probability under negative frame. Panel D: Binscatter of true placement probability vs. subjective placement probability under positive frame.

Figure B.5: Balance in Platform Pop-Up RD

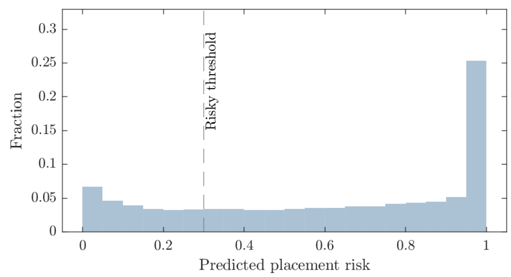
(a) Rural



(b) Economically Vulnerable

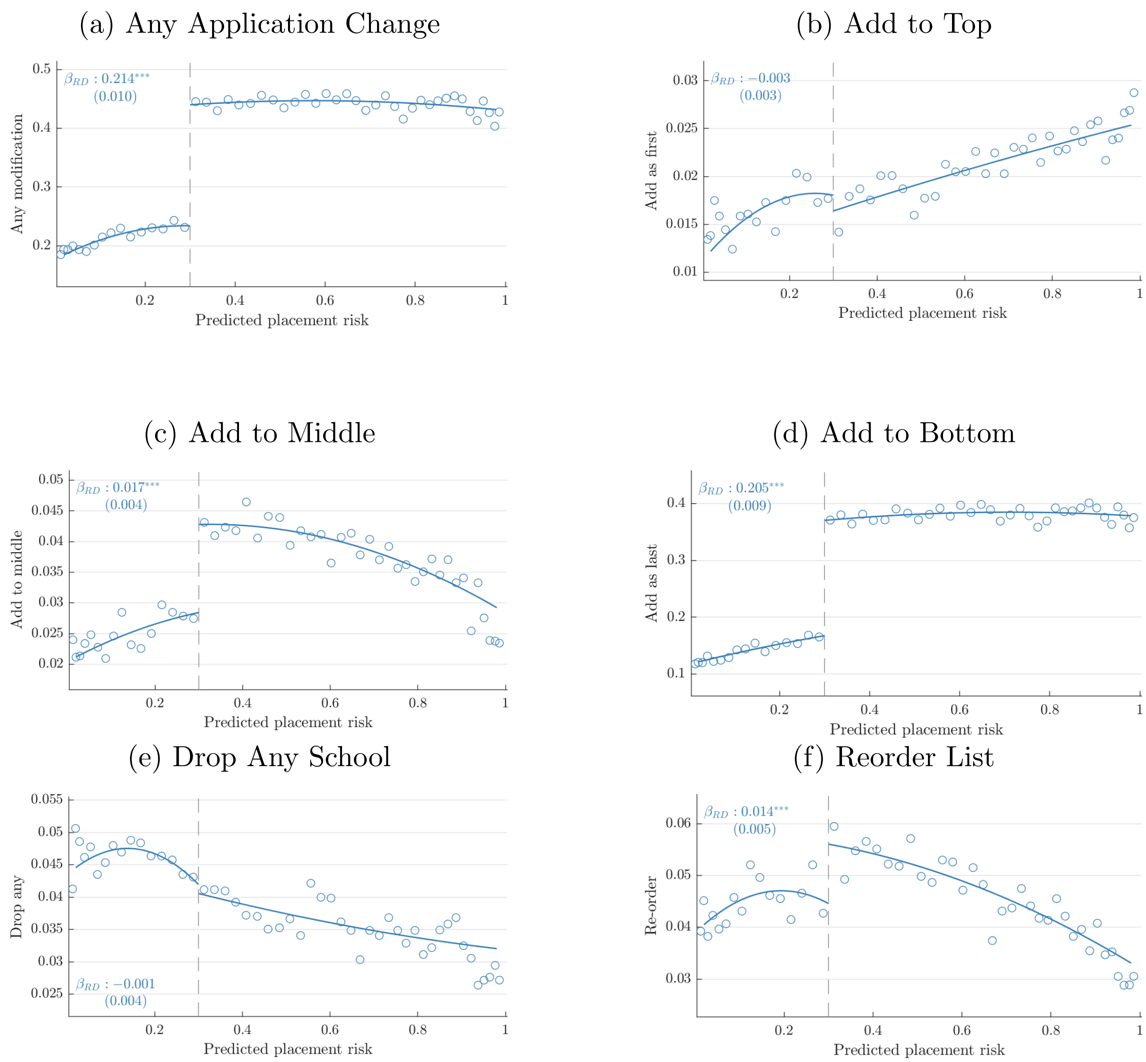


(c) Distribution of Predicted Risk (Only >0)



Notes. Binned means and global fits of predetermined characteristics by predicted risk for initial application. Points are centered means of 50 quantile-spaced bins of the support of the predicted placement risk $\in [0.02; 0.98]$. Solid line shows the quadratic fit. Reported coefficients and standard errors are from local linear specifications using $+-0.1$ bandwidth. See section 2.5 for details. Because coefficients are local while polynomial fits are global, there may be minor differences between displayed fits and reported coefficients. Panel A: vertical axis is indicator for rural location. Panel B: vertical axis is indicator for economic vulnerability (a measure of socioeconomic status). Panel C: histogram of predicted placement risk for initial application attempt, conditional on being greater than 0.01. Vertical lines display the 30% risk cutoff.

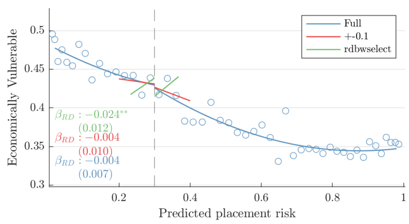
Figure B.6: Additional Platform Pop-Up RD Figures



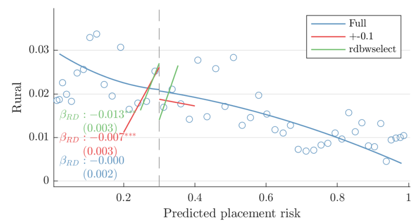
Notes. Binned means and global fits of choice outcomes by predicted risk for initial application. Points are centered means of 50 quantile-spaced bins of the support of the predicted placement risk $\in [0.02; 0.98]$. Solid line shows the quadratic fit. Reported coefficients and standard errors are from local linear specifications using $+-0.1$ bandwidth. See section 2.5 for details. Because coefficients are local while polynomial fits are global, there may be minor differences between displayed fits and reported coefficients. Outcomes by panel are as follows. Panel A: any application change. Panel B: add school to top of list. Panel C: add school to middle of list. Panel D: add school to bottom of list. Panel D: drop any school from list. Panel F: reorder existing schools.

Figure B.7: Multiple Bandwidths RD Plots of Platform-Based Pop-Up Warning Effects (Outcomes in Table 2.2)

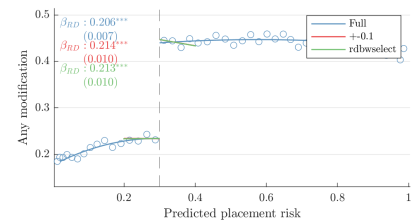
(a) Economically Vulnerable



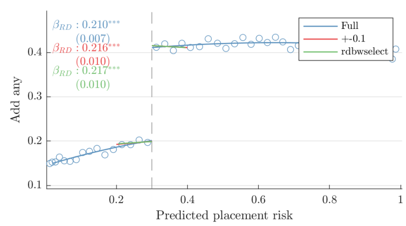
(b) Rural



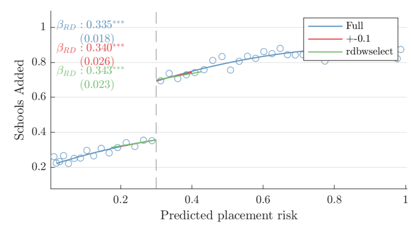
(c) Any Modification



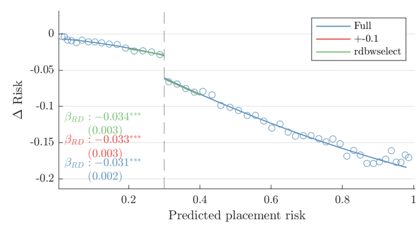
(d) Add Any



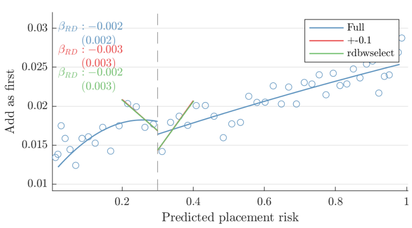
(e) Schools Added



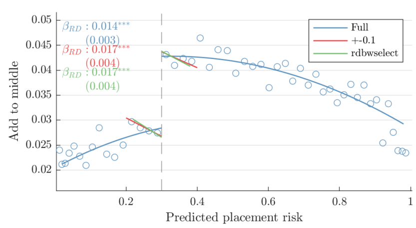
(f) Δ Risk



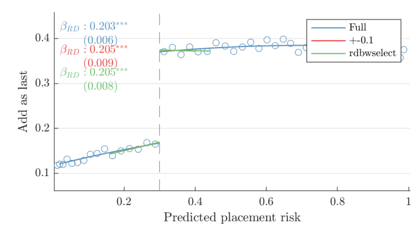
(g) Add as First



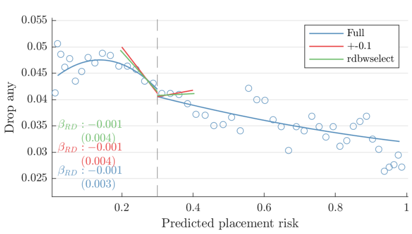
(h) Add to Middle



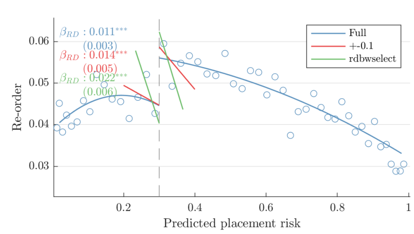
(i) Add as Last



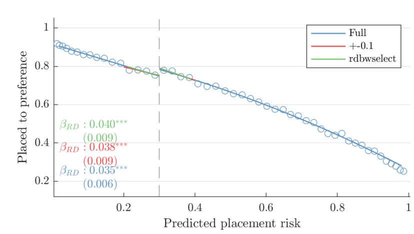
(j) Drop Any



(k) Re-Order



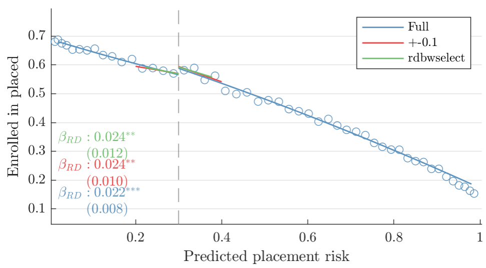
(l) Placed to Preference



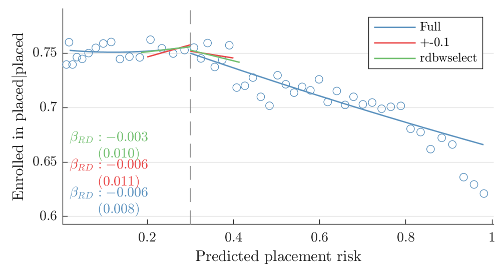
Notes. Pop-up warning RD effect fits and point estimates by bandwidth for outcomes listed in panel titles. "Full": global quadratic. "+/- 0.1": local linear within 0.1 bandwidth. "rdbwselect": optimal bandwidth selection using Calonico et al. (2014). See section 2.5 for details.

Figure B.8: Multiple Bandwidths RD plots of Platform-Based Pop-Up Warning Effects (Outcomes in Table 2.2)

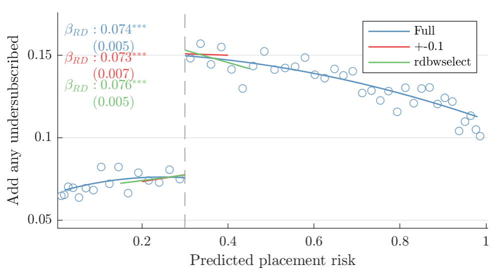
(a) Enrolled in Placed



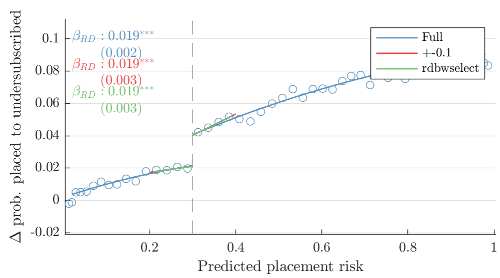
(b) Enrolled in Placed|Placed



(c) Add any Undersubscribed

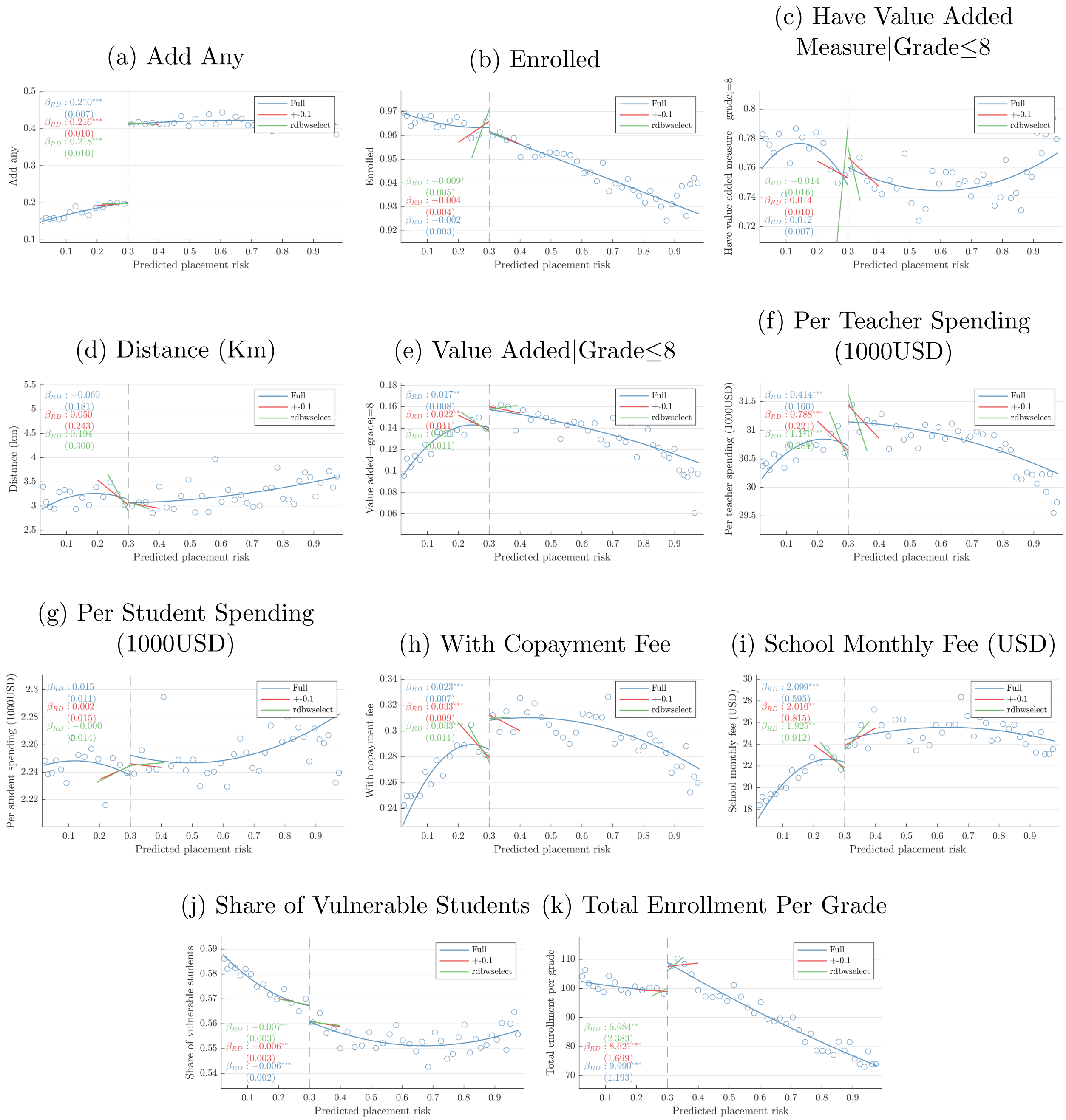


(d) Δ Prob. Placed to Undersubscribed



Notes. Pop-up warning RD effect fits and point estimates by bandwidth for outcomes listed in panel titles. "Full": global quadratic. "+/- 0.1": local linear within 0.1 bandwidth. "rdbwselect": optimal bandwidth selection using Calonico et al. (2014). See section 2.5 for details.

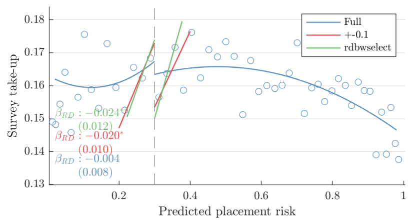
Figure B.9: Multiple Bandwidths RD Plots of Platform-Based Pop-Up Warning Effects (Outcomes in Table 2.3)



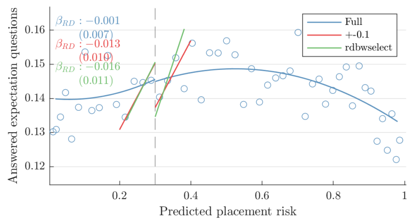
Notes. Pop-up warning RD effect fits and point estimates by bandwidth for outcomes listed in panel titles. "Full": global quadratic. "+/- 0.1": local linear within 0.1 bandwidth. "rdbwselect": optimal bandwidth selection using Calonico et al. (2014). See section 2.5 for details.

Figure B.10: Multiple Bandwidths RD Plots of Platform-Based Pop-Up Warning Effects (Outcomes in Table 2.5)

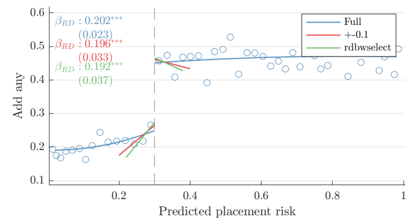
(a) Survey Take-Up



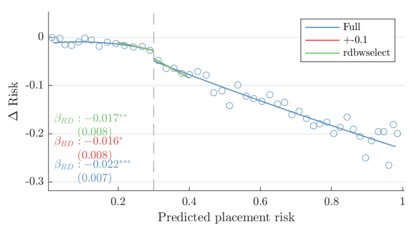
0.43
(b) Answered Expectation Questions



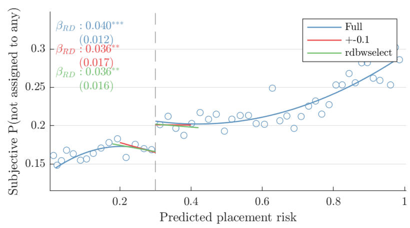
(c) Add Any



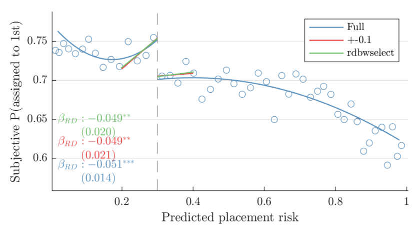
(d) Δ Risk



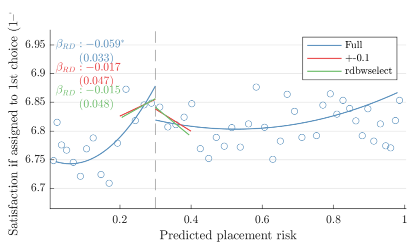
(e) Subjective P(Not Assigned to Any)



(f) Subjective P(Assigned to 1st)



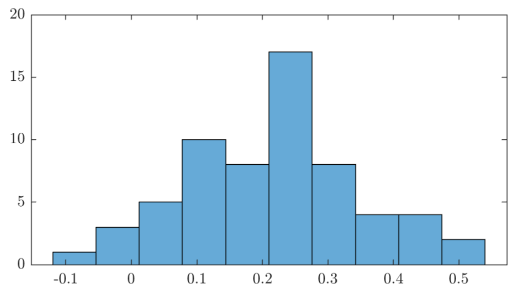
(g) Satisfaction if Assigned to 1st Choice (1-7)



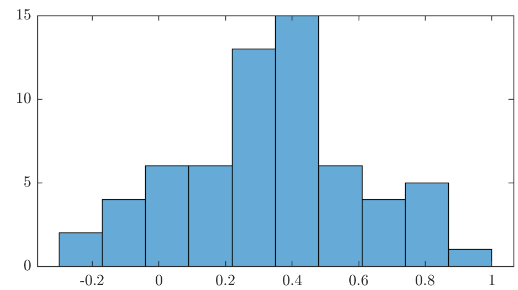
Notes. Pop-up warning RD effect fits and point estimates by bandwidth for outcomes listed in panel titles. "Full": global quadratic. "+/- 0.1": local linear within 0.1 bandwidth. "rdbwselect": optimal bandwidth selection using Calonico et al. (2014). See section 2.5 for details.

Figure B.11: Platform Pop-Up Effects over City-Years and by Market size

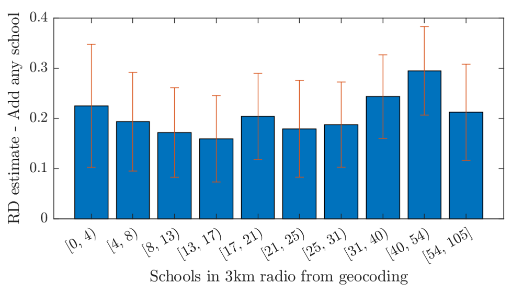
(a) Histogram: Estimates of Add Any School



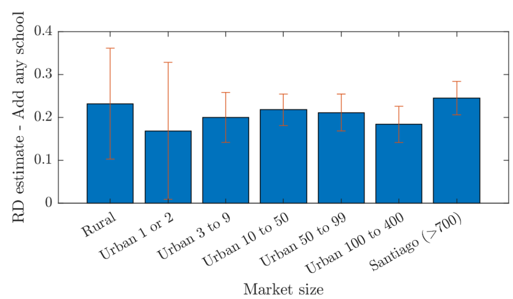
(b) Histogram: Estimates of Count of Schools Added



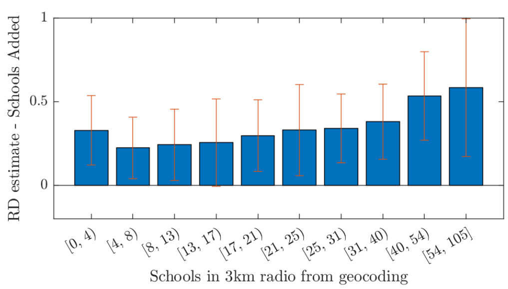
(c) Add Any School by Number of Nearby Schools



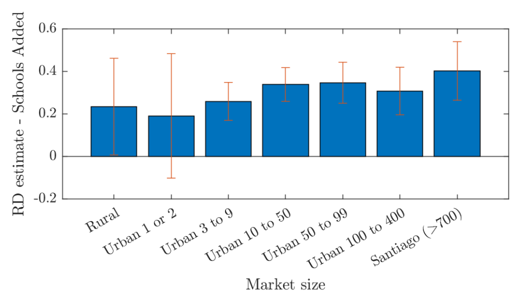
(d) Add Any School by Overall Market Size



(e) Schools Added by Number of Nearby Schools



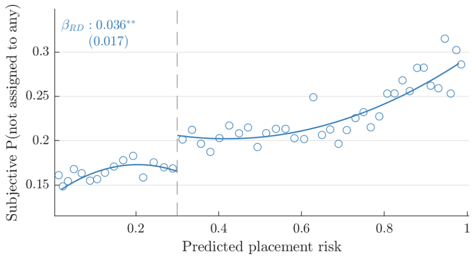
(f) Schools Added by Overall Market Size



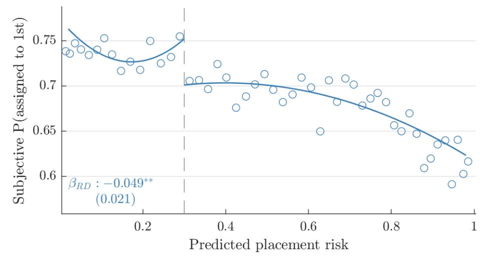
Notes. Panels A and B: distribution of estimated platform pop-up RD effects across city-year cells. Each city-year cell is one observation. Outcome in Panel A is add any school, outcome in panel B is count of schools added. Panel C: pop-up RD treatment effects on add any school split by count of nearby schools (within 3km of applicant address). Panel D: pop-up RD treatment effects on add any school split by overall market size, with size defined by the number of schools available to students in the city-year-grade cell and urban/rural status. Panel E: same as C, but with count of schools added as the outcome. Panel F: same as D, but with count of schools added as the outcome. See section 2.5 for details.

Figure B.12: Intervention Effects on Beliefs and Preferences

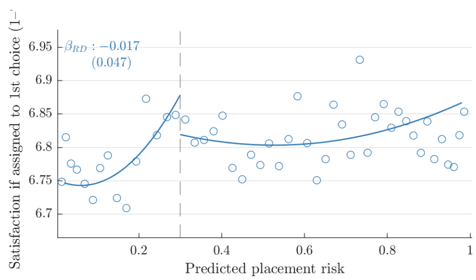
(a) Not Assigned to Any School



(b) Assigned to First Choice

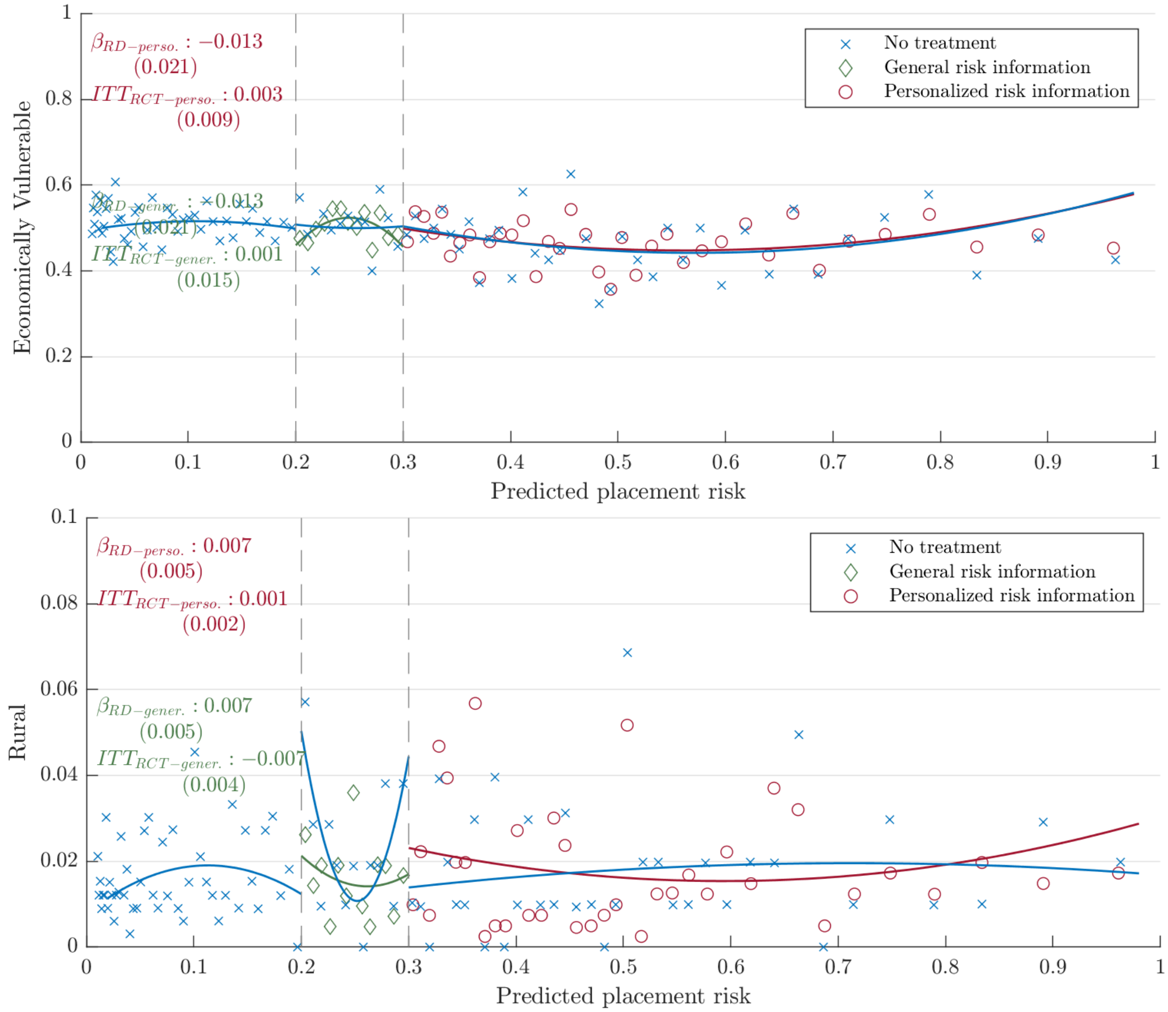


(c) Satisfaction if Assigned to First Choice



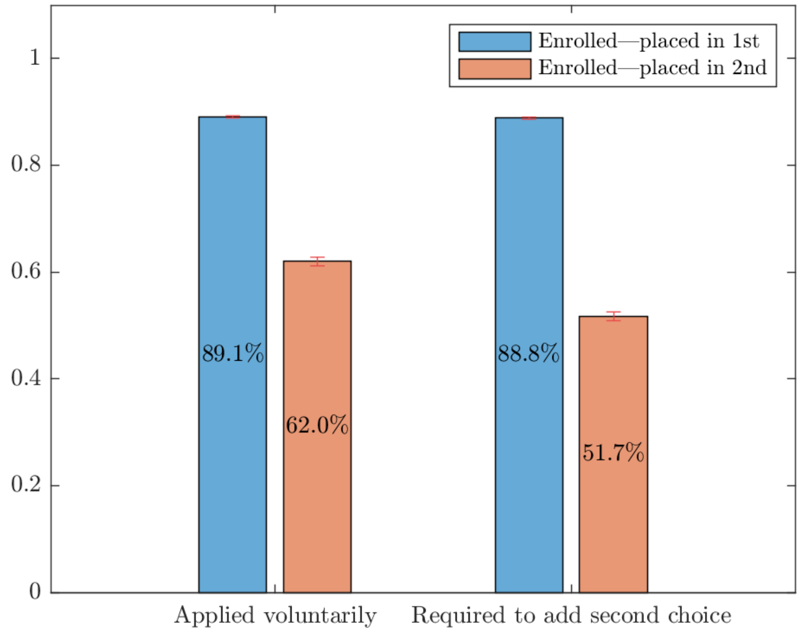
Notes. Binned means and global fits of choice outcomes by predicted risk for initial application. Points are centered means of 50 quantile-spaced bins of the support of the predicted placement risk $\in [0.02; 0.98]$. Solid line shows the quadratic fit. Reported coefficients and standard errors are from local linear specifications using $+-0.1$ bandwidth. See section 2.5 for details. Because coefficients are local while polynomial fits are global, there may be minor differences between displayed fits and reported coefficients. Sample is applicants who completed the belief modules of the endline survey. Panel A: outcome is survey-reported subjective belief about the chances of not being assigned to any school. Panel B: outcome is survey-reported subjective belief about placement chances at the first-listed school. Panel C: outcome is survey-reported satisfaction with placement at the first-choice school.

Figure B.13: Balance in 2021 WhatsApp RCT



Notes. Balance test results from 2021 WhatsApp RCT. Dependent variable is an indicator for (predetermined) economic vulnerability status (upper panel) and an indicator for rural location (lower panel). Treatments are as follows. "No treatment": control group that receives no WhatsApp message. "General risk information": treatment group that receives information about nonplacement risk in aggregate, not personalized to own application. "Personalized risk information": treatment group that receives information about own application risk, as in 2020 WhatsApp RCT. $\beta_{RD-general}$ is the RD estimate of general risk treament group against the control group at the 0.2 cutoff. $\beta_{RD-personal}$ is the RD estimate of the personalized risk information treatment group relative to the general risk treatment group. $ITT_{RCT-personal}$ and $ITT_{RCT-general}$ are RCT estimates of treatment effects for the personal and general info treatments (respectively) relative to the control group in the same risk range. See section 2.6 and Online Appendix B.3 for design details and additional results. Reported RD coefficients and standard errors are from local linear specifications using $+-0.1$ bandwidth. See section 2.5 for details.

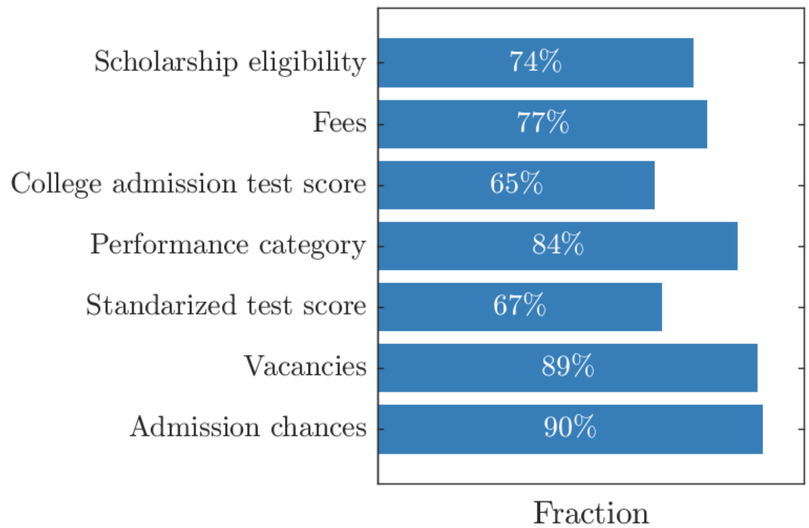
Figure B.14: Enrollment in Placed Rates Conditional on Placement Ranking



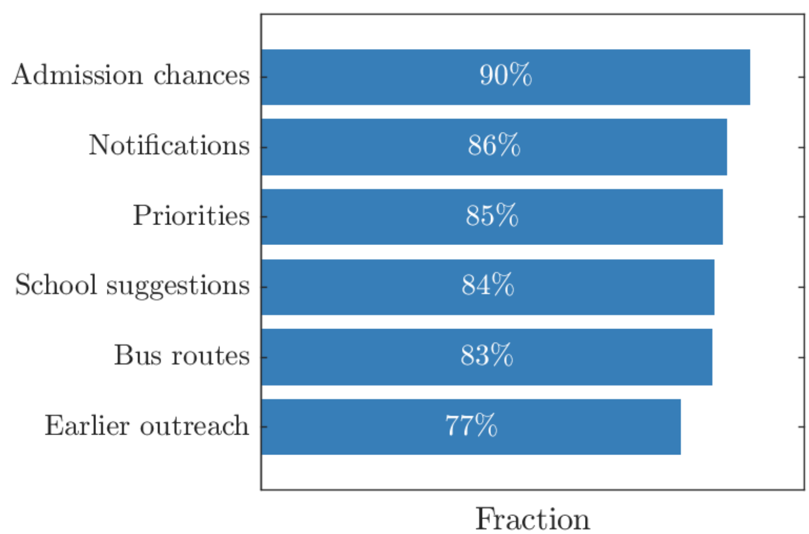
Notes. Share of students enrolling in the placed school, by rank of placed school and whether the applicant was forced to add the second school. The "applied voluntarily" group listed at least two schools on their initial application. The "Required to add second choice" group initially listed only one school and was required by the centralized system to add a second school.

Figure B.15: Demand for Information

(a) Information that You Would Have Liked to Have but You Did Not Have (Chile)



(b) Helpful Additional Information for Future Choice Participants (New Haven)



Notes. Share of survey respondents indicating desire for more information of the listed type, in response to the question listed in the panel title. Upper panel data source: Chilean school choice survey. See section 2.3 for details. Lower panel data source: email survey of 3,105 New Haven school choice participants in 2019 and 2020 (2,178 of those from applicants to simulator eligible grades). Bars refer to the following types of information (from top to bottom): personalized information about admission chances for school options, notifications about upcoming deadlines, detailed information about neighborhood and sibling priorities of school options, personalized suggestions for potential schools, information on bus routes to different school options, earlier outreach to allow more time to decide. See Online Appendix B.11 for more details.

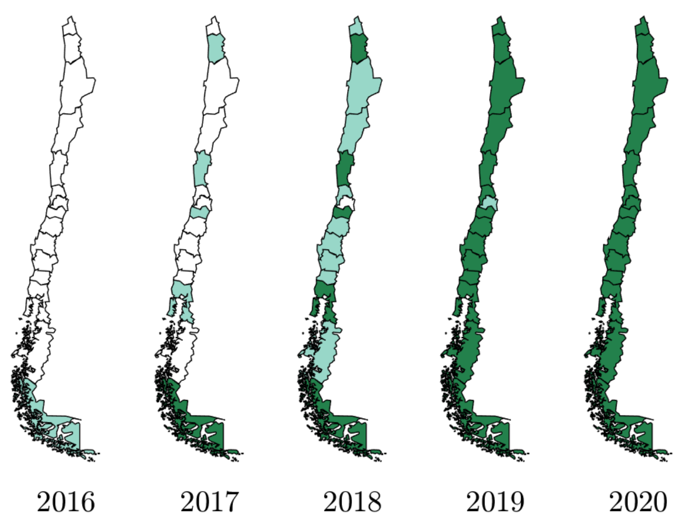
B.3 Chilean Admission Policy and Interventions

Policy

In May 2015, Chile's congress approved the "School Inclusion Bill." The goals of the bill included addressing discrimination in school assignment (Gobierno de Chile Ministerio de Educación, 2017). One major feature of the law was a change in the admissions process for public and private voucher schools in grades Pre-Kindergarten through 12. Schools of this type accounted for 92% of total primary and secondary enrollment. Between 2016 and 2020, the Ministry of Education implemented a nation-wide centralized school choice system. Rollout was staggered across regions and grades. Regions of Chile were divided into four sets. Each year a new set of regions was included in the system. For the first year following adoption in each region, only major "entry grades" were included in centralized choice. These grades were PK, K, 1st, 7th and 9th. In the second and following years of centralized choice in each region, all grades used the centralized system. The share of school-grade pairs included in the centralized system rose from 1% to 8% to 45% to 85% to 100% over the years 2016 through 2020. See Figure B.16 for an illustration of the policy rollout.

Schools that only enroll pre-Kindergarten and Kindergarten students were excluded from the centralized system. Lower-grade applicants could enroll through the centralized system in schools that also offer higher grades, but not in standalone early-grade institutions.

Figure B.16: Policy Rollout



Notes. This figure shows maps of Chile representing the implementation progress of the centralized application system by year. White regions represent where the system has not been implemented yet, light green reflects implementation only in entry grades (PK, K, 1st, 7th and 9th grades), and dark green means implementation in all grades of the region.

The centralized assignment process used a Deferred Acceptance algorithm with multiple tiebreakers (DA-MTB) to assign students to schools.[2] The law dictated that assignment include priorities for the following groups, in order: siblings, applicants with parents working at schools, and former students (i.e. students who previously were enrolled at a school but left). Ties were broken with lotteries. The law also imposed quotas for economically vulnerable students (15% of seats) and special needs applicants (count decided by each school, with a cap of two students per classroom). Finally, a small set of high schools was allowed to use quotas for high-performing students (30% of seats). In 2020, 39% of the schools offered seats for applicants with special needs in at least one level, while only 0.3% (23 schools) had a quota for high-performing students.

Following the initial assignment round, there is a second application round for applicants who do not receive or do not accept their initial offer. This round offers seats at schools with remaining excess capacity after the initial allocation.[3] Applicants without an offer after the second round are administratively assigned to the closest school to their registered address that has excess capacity.[4]

Students submit their applications to each choice round through an online platform. In addition to collecting applications, this platform provides information about schooling options, including test scores, fees, infrastructure, enrollment, religion, and extracurricular activities. See Figure B.17 for an illustration of the search engine and information. The Mineduc IT department developed the website, and a team from the Industrial Engineering Department of Universidad de Chile coded and ran the algorithm. The seed for the pseudo-random number generator is a mapping from the characteristics of the last six earthquakes recorded in Chile for a given date.

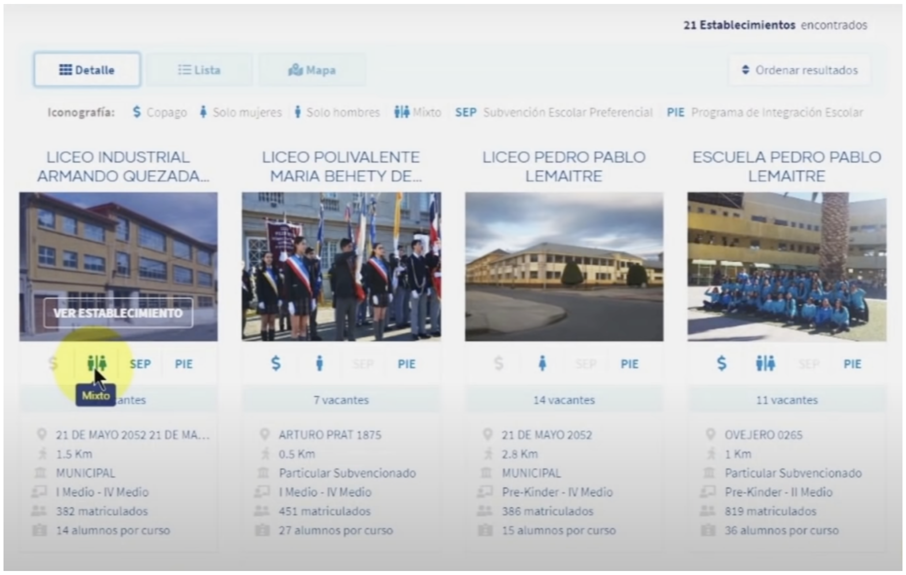
[2]For comparison over different lottery systems, see Ashlagi and Nikzad (2020).

[3]There is also a waitlist process between rounds that fills declined offers in over-demanded schools. This process works as follows. After assignments are made for the first application round, placed students are given the option accept their placements, to accept their placement, but consider waitlist options, and to reject the placement. The default choice for non-responders is to accept the placement. Mineduc then reruns the DA process using the already-submitted rank-order lists, with the set of applicants restricted to students who were unplaced, rejected their placements, or chose the "accept, but consider waitlist" option. The set of seats consists of those opened up by applicants rejecting placed schools. These additional "interim" placements account for a very small share of placements overall; in 2020, for example, they made up 1.2% of all placements. We do not include these interim placements in our main-text analysis of placement counts.

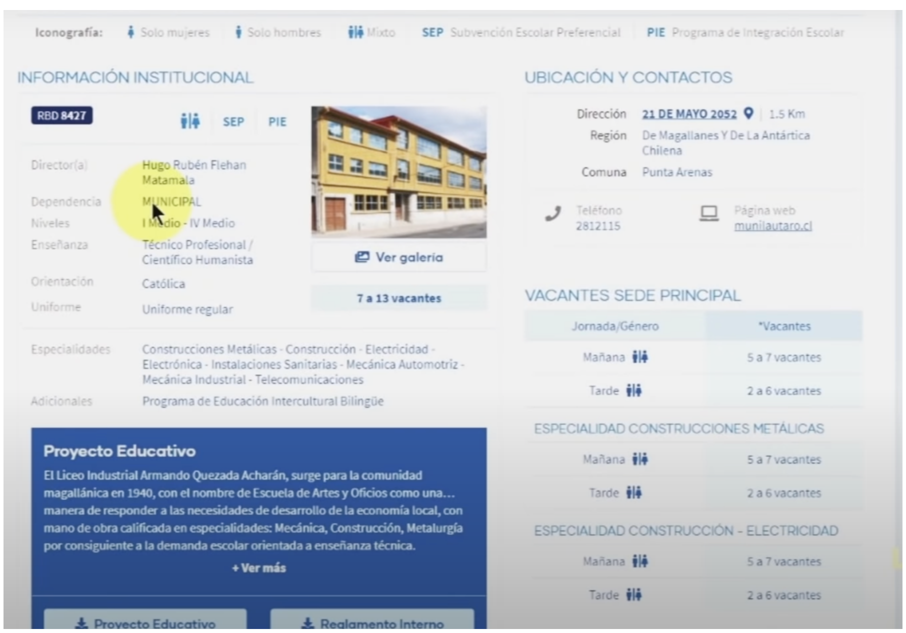
[4]In 2016, Mineduc also implemented default assignment to the closest school for the *first* application round. We do not use 2016 data in our analysis.

Figure B.17: Application Platform Screenshots

(a) Gallery of Schools



(b) Detailed Information of a School



Notes. Panel A shows an example of an applicant's view of the gallery of schools, what includes the main photo and a few characteristics as distance to home, enrollment, or price. Alternatively, users can also choose to see a list of schools or a map with all the options available close to home. Panel B is a screenshot of the information for a specific school, including educational project, an estimate of the seats available, religion, etc.

Policy Rollout, Placement Outcomes, and Outside Options

Policy Implementation and Policymaker Concerns.

Rollout of the centralized system proceeded more or less as planned. One concern for policymakers following the adoption of the centralized system was the proportion of applicants not assigned to any school they had listed. Reasons for this concern included the disutility of waiting for certainty about enrollment, the potential for unrealized matches between families and schools, the costs of aftermarket coordination, and the effect on the new system's reputation. To the latter point, policymakers were concerned that families might expect the centralized school choice system to assign all applicants.

Placement Outcomes over Time.

Table B.8 reports placement statistics for each year. It displays both aggregate placement statistics and statistics for the set of markets that first entered the centralized system in a given year, for each entry year between 2016 and 2020. Results in Panel A show that the share of applicants assigned to any preference has decreased. This is true both overall, and, to a lesser extent, within the set of markets entering the centralized system in each year between 2016 and 2020. Panels B and C show the share of applicants (in the full applicant sample) who are unplaced and are entering the schooling system or do not have the option to continue in their current school (Panel B, "Share not placed in pref. without continuation option") and the share who are unplaced but have the option to continue (Panel C, " Share not placed in pref. with continuation option"). Most of the rise in nonplacement comes through the latter channel.

Panel D of Table B.8 shows that mean application length declines over time, both overall and within each year-of-entry group.

Options Outside the Centralized System.

This subsection describes applicants' outside option behavior in detail. Figures B.18a and B.18b show the next-year enrollment of students that did not get a spot in the first round of the centralized process (i.e., those whom we classify as unplaced in our main text analysis). We split the sample by whether applicants have an option to continue in their current school or not. Figure B.18a shows enrollment outcomes for applicants without the continuation option. Roughly 60% of these applicants go on to enroll in some SAE school. Roughly 20% enroll in a voucher school outside of the centralized system. These students are mostly pre-kindergarten or kindergarten applicants enrolling in standalone preschools that do not participate in the main system. Roughly 12% enroll in a private school that does not accept vouchers; private schools that decline vouchers tend to be quite expensive. For a small fraction of unplaced students (4-11% depending on the year), we do not observe any enrollment outcome.

Table B.8: Aggregate Placement Results for Stable Populations by Year

	2016	2017	2018	2019	2020
A. Share placed in preference					
Aggregate	0.884	0.872	0.828	0.771	0.786
Since 2016	0.884	0.840	0.799	0.767	0.791
Since 2017		0.874	0.821	0.807	0.850
Since 2018			0.831	0.788	0.820
Since 2019				0.740	0.770
Since 2020					0.606
B. Share not placed in preference without continuation option					
Aggregate	0.088	0.087	0.089	0.104	0.094
Since 2016	0.088	0.122	0.121	0.152	0.117
Since 2017		0.085	0.110	0.116	0.096
Since 2018			0.081	0.102	0.094
Since 2019				0.100	0.096
Since 2020					0.079
C. Share not placed in preference with continuation option					
Aggregate	0.028	0.041	0.083	0.125	0.120
Since 2016	0.028	0.038	0.080	0.080	0.092
Since 2017		0.041	0.069	0.077	0.054
Since 2018			0.088	0.110	0.086
Since 2019				0.159	0.135
Since 2020					0.315
D. Mean of application list length					
Aggregate	3.4	3.5	3.2	3.2	3.0
Since 2016	3.4	3.2	3.2	3.2	2.9
Since 2017		3.5	3.3	3.1	2.9
Since 2018			3.1	3.1	2.9
Since 2019				3.4	3.3
Since 2020					2.9

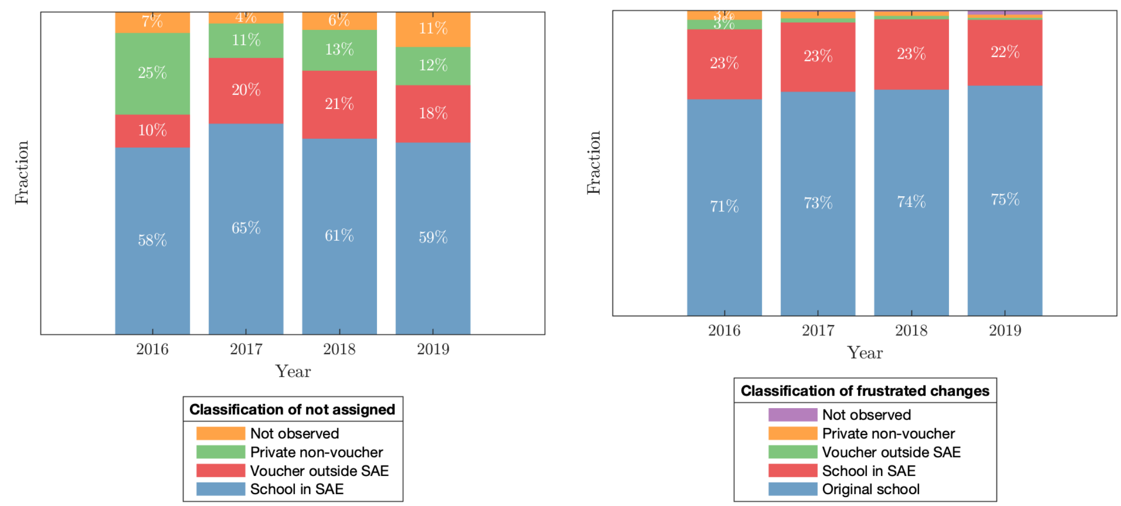
Notes. In Panel B “Share not placed in preference without continuation option” also includes applicants that do not have a current school because they are entering the schooling system. “Since 201X” represents the zones and grades where the new centralized system was implemented in 201X. Therefore, each row represents a stable population. The category “Change of school not realized” reflects students that applied from a school that offers the next grade and were not assigned to any submitted preference. They kept the seat at their school of origin.

Figure B.18b shows that roughly three quarters of unplaced applicants who have the option to continue in their current school choose to do so. Almost all of the remaining applicants enroll in another SAE school.

Figure B.18: Enrollment Outcomes for Unplaced Students

(a) No Continuation Option

(b) Continuation Option



Notes. Panel A shows where we observe the applicants enrolled for the set of students that were not placed to any school in the assignment and did not have the option to continue at their current school. Panel B is the same for the set of applicants who are unplaced in the main assignment round but have the option to continue at their current school.

Feedback Interventions

Mineduc worked with the Consiliumbots NGO to design and implement information interventions in the placement process each year between 2017 and 2020. Our analysis focuses on the 2018-2020 placement cycles, when the placement process was operating at or close to full scale. This section describes the interventions conducted in each year, including the 2017 intervention.

2016 SMSs were sent in three waves to applicants in the Puntarenas region (the only region with centralized choice at the time) in markets where there were overdemanded schools. The first wave was on day 25 of the application process. Three different general messages and no-message were randomly assigned and sent to all the applicants with fewer than five schools on their applications. For examples, see panels (a) to (c) of Figure B.19. These are the interventions for which we report results in Table 2.6.

We also conducted an additional intervention in 2016 that provided a leading indicator on the promise of personalized risk warnings. On day 36, applicants with a predicted risk higher than 0.01 were randomly assigned to be the recipients of an additional personalized SMS. This message represented a waypoint on the path from impersonal, no-information nudges to the personalized smart platforms deployed in 2017 and later. It included a personalized reference to the number of schools that the applicant had already placed on his list, which is risk-relevant for many applicants, but did not refer directly to application risk. See panel (d) of Figure B.19. On day 44, 5 days before the last application day, all applicants in the control group received the same SMS. The impact of the personalized SMS on the share of applicants who added schools prior to the message being sent to the control group is 0.051 (0.023), roughly double the size of the encouragement nudge effects reported in Table 2.6. The idea that personalized, risk-relevant information might be more effective than generic encouragement was an input to the choice to roll out smart platforms in subsequent years.

The personalized (but not "smart platform") SMS intervention was cross-randomized with the impersonal encouragement nudge interventions. When reporting the results of the encouragement nudge interventions in Table 2.6, we consider only changes to the application made in the "clean" 11 day window between the encouragement nudge and personalized non-smart intervention. This approach parallels our analysis of the 2020 WhatsApp RCT in section 2.5, and lets us capture the effects of the encouragement nudge in the absence of interactions with the subsequent personalized intervention. Estimates of encouragement intervention effects that take final applications as the outcome show null results similar to those reported in Table 2.6, both in the full sample and in the subsample of applicants assigned to the control group for the personalized intervention. These results are available upon request.

2017 There were two information interventions this year. The first was an on-platform pop-up warning to risky applicants (Figure B.21). The threshold that defined a "risky applicant" varied between 0.3, 0.5, and 0.7 values of predicted risk, depending on the region. The pop-up was active starting with the third day of the process. See Appendix B.9 for a discussion of results from this intervention.

The second intervention consisted of SMSs sent to risky applicants who applied during the first two days, when the pop-up was not active. Four different personalized messages were randomized. The basic content was composed by three concepts (1) risk, (2) consequences of risk, (3) a recommended action, the message were different on the order of the concepts and the wording. For examples, see panels (a) to (d) of Figure B.20.

2018 The implementation team at Mineduc kept on-platform pop-up from 2017, fixing the definition of "risky" as applicants with a predicted probability of non-assignment higher than 30%. It was active from the 3rd day of application. Only applicants applying after this date are included in our platform pop-up analysis. Figure B.21 displays the platform pop-up messages in 2018 and also in 2019-20.

Followup SMS messages were sent to risky applicants who applied during the first two days, when the pop-up was not active. A final SMS was sent four days before the end of the period to every risky applicant that did not receive the first SMS.

2019 As in 2018, but without the final SMS reminder to still-risky applicants. The platform pop-up was again active from the 3rd day of application. Only applicants applying after this date are included in our platform pop-up analysis.

2020 The platform pop-up was implemented again, this time from the first day of application. Mineduc sent an SMS to risky applicants eight days before the deadline.

Additionally, four days before the final application deadline, the NGO randomized the assignment of WhatsApp messages within risky applicants. The reason for the randomization was a cap on the number of messages that could be sent in a day. The message included one of three types of images, each with a slightly different type of information related to the risk and the application. See Figure B.22. Each image conveys the same idea as the pop-up warning. We pool over image types in our main analysis of the WhatsApp RCT. See Appendix B.10 for a breakout of effects by image type.

The set of students eligible for randomization into RCT treatment and control groups was restricted in several ways. First, the RCT sample included applicants to grades Pre-Kindergarten, Kindergarten, and 1st grade, from urban zones, without sibling priority at any school on their list. Second, the warnings RCT sample was layered on top of a parallel information intervention on school attributes also being conducted by the NGO and Mineduc. This second intervention involved sending emails about school attributes to choice applicants. The sample for the warnings RCT was drawn from the set of applicants who received the attributes email but had not yet opened it. The reason for this is the WhatsApp campaign was viewed at Mineduc as a reminder to check this report card. This sample selection approach does not affect the *internal* validity of the WhatsApp experiment, but does affect how one interprets the findings. Our view is that the approach to sample selection will tend to draw relatively low-interest applicants (those who had not opened other correspondence from the authority), but in an environment with relatively low search costs (they had access to the attribute report cards if they wanted them). Recall that that search costs may be relatively low for many applicants in this setting, given the attribute search tools embedded in the application platform (see section B.3).

Furthermore, two days after the WhatsApp message went out, Mineduc sent a final reminder using SMS messages to remaining still risky applicants with a link to the same image attached in the WhatsApp message, encouraging applicants to add more schools.

2021 The platform pop-up warning was implemented again, starting on the first day of the application window. As in the previous years, the risk cutoff for platform pop-up receipt was 0.3. Risk calculations were based on demand from 2020 and not updated with current demand until the 20th day. On day 20, the NGO re-calculated nonplacement risk for all applicants based on 2021 application data.

Starting on day 20, the NGO conducted a risk warning campaign with the goal of reaching out to applicants who had not received a platform pop-up warning based on the initial risk calculation, but who appeared to be at risk of nonplacement under the updated calculation. As in 2020, this campaign included a randomized WhatsApp component. Within the universe of applicants who had not received a platform pop-up warning, applicants with (updated) nonplacement risk above 0.30 were randomly assigned to either a control group that received no message or a treatment group that received a message with personalized application information similar to the pop-up warning. In addition, applicants with risk scores between 0.2 and 0.3 were randomly assigned to either a control group that received no message or a treatment group that received a *non-personalized* messsage about nonplacement risk in the aggregate. These two treatments form the basis for the analysis described in section 2.6. See Figure B.23 for screenshots of the WhatsApp messages in the two treatment arms. The number and timing of warnings was the same across the two treatment arms, except for the difference in warning text.

Two points are important to make here. First, note that sample selection into the 2021 WhatsApp RCT is somewhat different than for the 2020 WhatsApp RCT, because the 2021 RCT sample universe consisted of applicants who had not yet seen a pop-up warning on the application platform, whereas in 2020 most participants in the WhatsApp RCT had already received a platform pop-up, as described in section 2.5.

Second, risk calculations changed enough for enough applicants that the RCT sample universe of applicants who had not received platform pop-up had support across the distribution of updated risk scores.[5] Relatively low-risk applicants are over-represented among RCT-eligible applicants. However, as shown in Appendix Figure B.24, we observe RCT-eligible applicants over the full range of updated risk values, and the share of RCT-eligible applicants is smooth through the 0.3 cutoff of the updated risk distribution. This latter point speaks to the validity of the RD analysis described in section 2.6.

Five days after the WhatsApp messages went out to treatment groups, 28% of RCT participants (and 30% of applicants overall) received a report card that included the risk measures for every school in the application. For risky applicants the report card also had a warning with a similar message to the platform pop-up. As in our analysis of the 2020 WhatsApp intervention, we focus our analysis of the effects of the 2021 WhatsApp RCT on changes made to the application in the "clean" five-day window

[5] Risk updating was an important issue in 2021 because Covid-19 had depressed applications in 2020.

between the randomized WhatsApp intervention and the followup message sent to all risky applicants. Findings reported in Figure 2.7 and referenced in the main text are from this clean five-day window.

Table B.9 summarizes findings from the 2021 WhatsApp RCT beyond those reported in Figure 2.7. As expected, both the RCT and RD designs are balanced on applicant observables (specifically, economic vulnerability) and on eventual receipt of the report card intervention. Roughly 80% of applicants in both the general and personalized information treatments view the WhatsApp image. There is no difference in viewership rates across treatment type. We observe changes in choice behavior both in the "clean" five-day window between the WhatsApp treatment and the report card and at the endline. As in the 2020 WhatsApp RCT, treatment effects grow over time. Effects on choice behavior (i.e., the share of "compliers" who add schools, the number of schools added) are much larger than in 2020. IV estimates of risk reductions for compliers are somewhat smaller than in 2020, likely because the population receiving the WhatsApp treatment is less risky at baseline.

We have also conducted analyses that exclude the 28% of applicants who received the report card intervention from the sample. These findings are available upon request and yield the same conclusions as those we report here.

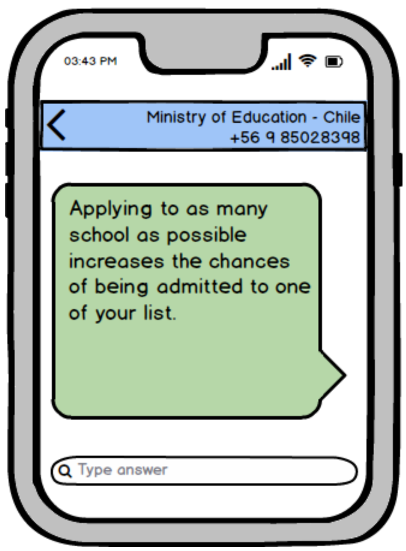
Table B.9: WhatsApp RD and RCT Results – 2021

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	General risk information				Personalized risk information			
	RCT		RD		RCT		RD relative to general risk treatm.	
	ITT	IV	ITT	IV	ITT	IV	ITT	IV
A. Balance								
Economically Vulnerable	0.001		-0.013		0.003		-0.008	
	(0.015)		(0.021)		(0.009)		(0.016)	
Rural	-0.007		0.007		0.001		0.006	
	(0.004)		(0.005)		(0.002)		(0.004)	
B. Message receipt								
WhatsApp read	0.793		0.798		0.792		0.005	
	(0.005)		(0.012)		(0.003)		(0.014)	
Report card intention to treat	0.004		0.013		0.008		0.026	
	(0.014)		(0.019)		(0.008)		(0.015)	
C. Outcomes in clean 5 days before report card								
Any modification	0.041		0.043		0.081		0.037	
	(0.003)		(0.007)		(0.003)		(0.008)	
Add any	0.039		0.038		0.080		0.038	
	(0.003)		(0.006)		(0.003)		(0.008)	
Schools Added	0.070	1.807	0.067	1.766	0.156	1.957	0.086	2.013
	(0.007)	(0.114)	(0.015)	(0.257)	(0.006)	(0.052)	(0.020)	(0.441)
Δ Risk	-0.002	-0.050	-0.001	-0.024	-0.013	-0.157	-0.000	-0.057
	(0.001)	(0.012)	(0.001)	(0.035)	(0.001)	(0.008)	(0.002)	(0.058)
D. Endline outcomes								
Any modification	0.054		0.053		0.099		0.055	
	(0.005)		(0.009)		(0.004)		(0.010)	
Add any	0.051		0.046		0.096		0.055	
	(0.005)		(0.009)		(0.004)		(0.009)	
Schools Added	0.098	1.921	0.085	1.856	0.203	2.108	0.148	2.978
	(0.015)	(0.236)	(0.022)	(0.313)	(0.010)	(0.067)	(0.025)	(0.611)
Δ Risk	-0.001	-0.029	-0.002	-0.034	-0.016	-0.163	-0.004	-0.149
	(0.001)	(0.017)	(0.002)	(0.049)	(0.001)	(0.009)	(0.003)	(0.061)

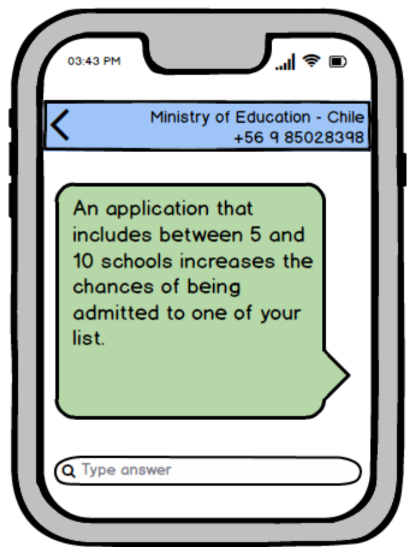
Notes. ITT and IV effects of 2021 WhatsApp warnings intervention. RCT columns (1) and (2): effects of random assignment to general risk information message vs. control group for students with predicted risk $\in (.2, .3]$. RCT columns (5) and (6): effects of random assignment to personalized risk information message vs. control group for students with predicted risk $> .3$. Robust SEs in parentheses. General information intervention N=6,819. Personalized information intervention N=18,763. RD columns (3) and (4): regression discontinuity evaluation of general risk information message vs no treatment around 0.20 cutoff. RD columns (7) and (8): regression discontinuity evaluation of personalized risk information message vs general risk information message around 0.30 cutoff. RD specifications computed using local linear fit with a bandwidth of 0.1. Standard errors are heteroskedasticity-robust nearest neighbor variance estimator with minimum of 3 neighbors, as in Calonico et al. (2014). ITT column shows effects of group assignment. IV columns show the instrumental variable specification, where the endogenous regressor is the add any school indicator, instrumented with group assignment for the RCT, and with a dummy of crossing the risky threshold for the RD. Panel A: balance tests on predetermined characteristics. Panel B: message receipt. "WhatsApp read" is an indicator equal to one if applicant views the WhatsApp treatment message. "Report card intention to treat" is indicator for receiving a report card 5 days later. Report card included additional information on nonplacement probability and school options. 28% of the RCT sample were assigned to received the report card. Panel C: outcomes within 5 days window between WhatsApp intervention and report card was sent. Panel D: endline choice behavior. See the description of the 2021 intervention in section B.3 for details.

Figure B.19: SMSs Intervention Warning Texts – 2016

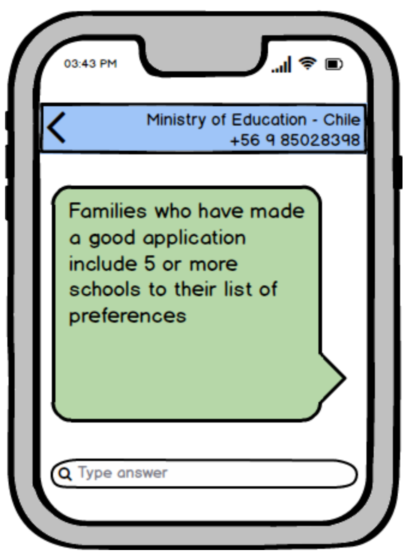
(a) General SMS - “More Schools, Higher Chances”



(b) General SMS - “Range Suggestion”



(c) General SMS - “Role Model”



(d) Personalized SMS

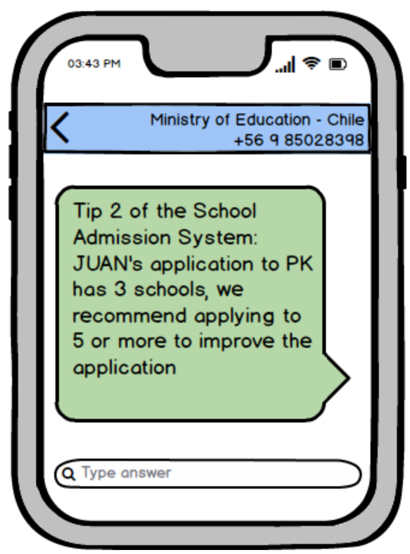
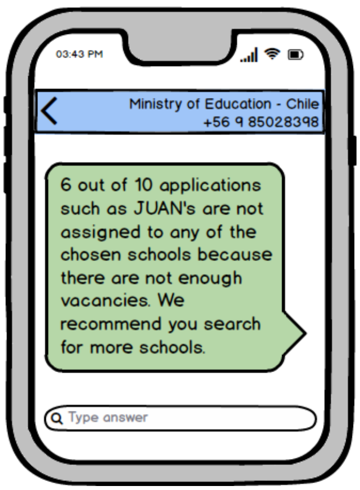
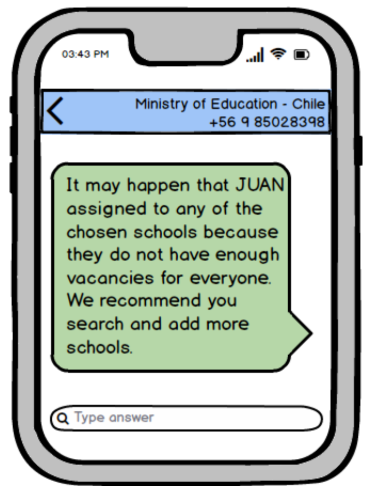


Figure B.20: SMSs Intervention Warning Texts – 2017

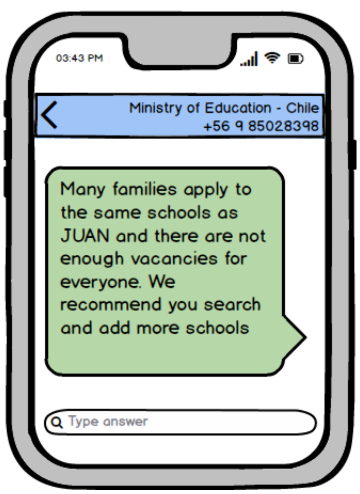
(a) Personalized Treatment - "Probability First"



(b) Personalized Treatment - "Consequences First"



(c) Personalized Treatment - "High Demand First"



(d) Personalized Treatment - "Recommendation First"

Figure B.21: Platform Pop-Ups

(a) Platform Pop-Up 2018

(b) Platform Pop-Up 2019 - 20

(c) Translation Pop-Up 2019 - 2020

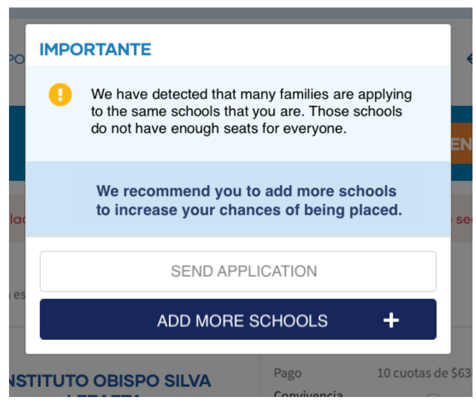
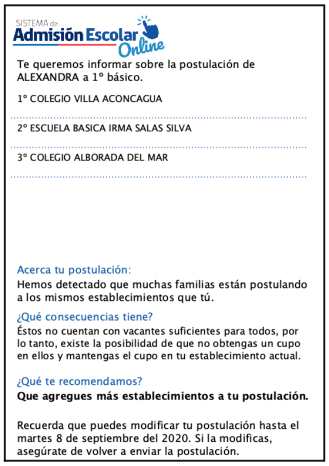
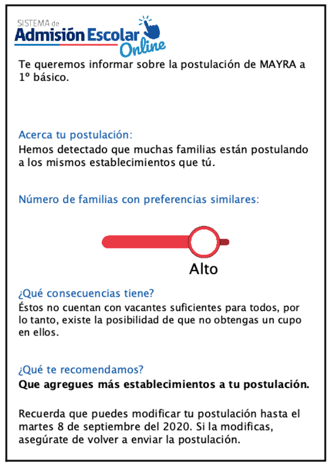


Figure B.22: WhatsApp Intervention Warning Images – 2020

(a) Simple Image



(b) Risk Warnings Bar



(c) School-Specific Vacancy Estimates

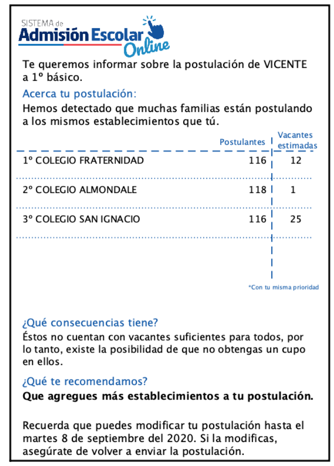
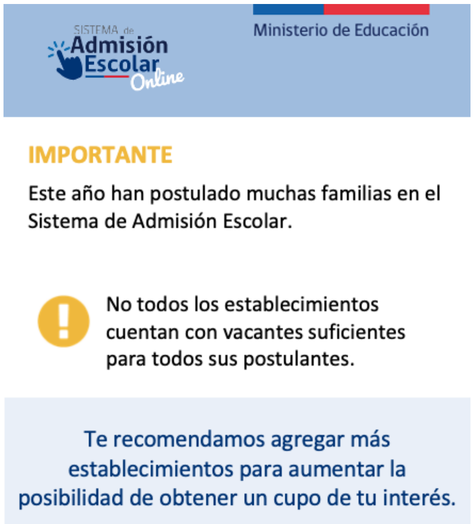
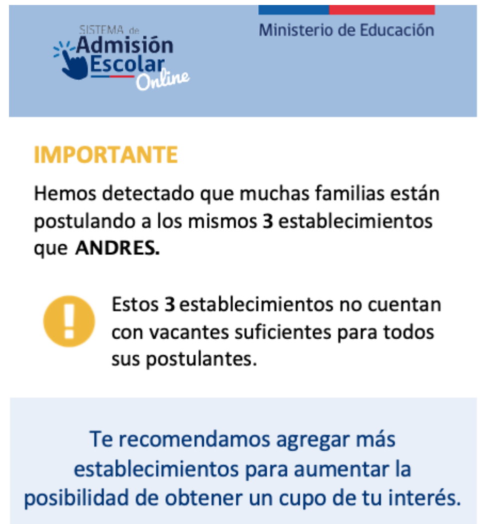


Figure B.23: WhatsApp Intervention Warning Images – 2021

(a) General Information Image



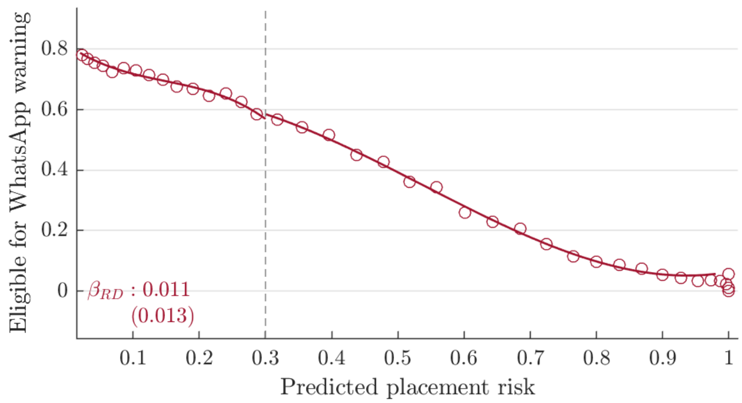
(b) Personalized Information Image



(c) Translation of General Information Image

(d) Translation of Personalized Information Image

Figure B.24: Eligibility for 2021 WhatsApp RCT



Notes. Share of 2021 applicants eligible for 2021 WhatsApp RCT by *updated* risk score. Only applicants who had not received the platform pop-up were eligible for the 2021 WhatsApp RCT. This graph plots of the share of such applicants in 50 quantile-spaced bins. Solid line shows a cubic fit. Reported coefficient and standard error are from local linear specifications using $+-0.1$ bandwidth. See section 2.5 for details. Some applicants below the 0.3 cutoff in the updated risk score received the platform pop-up, while some applicants above the 0.3 cutoff did not. This is because the platform pop-up assignment was based on initial risk scores, before the update. See the description of the 2021 intervention in section B.3 for details.

B.4 Datasets

Datasets come from three sources. The first is publicly available administrative data. These data include all the inputs necessary to replicate assignment outcomes (including application lists, priorities, lottery numbers, and seat counts) as well as historical enrollment records.[6] The second source is confidential administrative records on application submission and edit histories (i.e. the time path of submitted applications for a given applicant) and priority groups for each applicant at each school. We use this data to construct risk predictions. The third source is survey data collected after the application process in 2020. This section describes each of these sources.

Public Administrative Data

Public data comprises all the necessary inputs to compute the assignment. It includes rank order lists, priorities, vacancies, lottery numbers, and final assignments.

Confidential Administrative Data

In addition to the public data, we have access to applicants' priorities to every school, geocoding, and application edit histories. The details of data availability vary by year, particularly with regard to application edit histories and eligibility for the platform pop-up intervention. Looking across 2018, 2019, and 2020– the year range included in our main text analysis– we observe edit history data for 96.4% of applicants.

2016: We have daily copies of the application database. If an applicant files more than one application within a day, we observe only the last one.

2017: The NGO's risk classification web service stored the history of applications for students from the 20 most significant urban zones, covering 88% of the total. We have access to these data.

2018: The NGO's risk classification web service stored the history of applications for 84.2% of the applicants. We have access to these data. The initial intention here was to store application history for all applicants, but the remaining 15.8% of histories were not retained. The Mineduc IT team believes that the omission is because of a timeout they set to reach our web service. Applicants in this 15.8% did not receive the platform pop-up, and therefore are not included in our analysis of pop-up outcomes. Our platform pop-up analysis imposes two additional data restrictions, dictated by limits on the set of applicants eligible for the intervention. First, the NGO risk classification system considered only the 20 largest urban zones; students in other zones did not receive the pop-up. Second, the risk classification system excluded applicants that applied during

[6] These data are available on the Mineduc website, `https://centroestudios.mineduc.cl/`.

the first two days of the application process; the web service was not active until day three of the application cycle (see above). Overall, our pop-up analysis includes 44.3% of all 2018 choice applicants.

2019 The NGO's risk classification web service stored the history of applications for 99.9% of the applicants. Our analysis of the platform pop-up imposes the additional restriction that applicants must have applied after the first two days of the application cycle, because the web service providing the warnings was not active until day three (see above). Overall, our pop-up analysis includes 65.1% of all 2019 applicants.

2020 The risk classification web service stored the history of applications for all applicants. Our analysis of the platform pop-up excludes 9% of applicants applying to non-entry grades in the Metropolitan region, for whom there was no prior-year data on which to base initial risk predictions. These applicants did not receive the pop-up intervention. Overall, 91% of applicants 2020 are included in the pop-up analysis.

2021 The risk classification web service stored the history of applications for all applicants.

Survey Data

We sent an online survey to the parents of all applicants. The sample universe consisted of 373,710 households. This is slightly smaller than the number of applicants because we sent one email to each parent, and some households have multiple applicants. 66,282 (18%) started the survey, and 48,929 (13%) finished the survey. See section B.7 for survey text.

School Attribute Data

Tables 2.1 and 2.3 analyze the attributes of schools where students enroll. The variables we use are described in Neilson (2021). In particular, see Neilson (2021)'s school expenditure and value added estimation data supplements. A sketch of variables we use is as follows:

- Per teacher spending: The school's total wage bill for classroom teachers is divided by the number of classroom teachers.
- With copayment fee: Schools that charge an out-of-pocket fee beyond the government voucher are indicated with a 1 and schools with no additional fees are coded as a zero.
- School monthly fee (USD) is the average monthly out-of-pocket fee that families must pay at each school. This is the tuition net of the base voucher.
- Share of vulnerable students: The government of Chile identified students from disadvantaged families with a designation of "prioritario." This variable is share of students at a school that are "prioritario."

- Total enrollment per grade is the number of students enrolled by grade level. It is the cohort size.

- Per student spending: The school's total expenditures are divided by the number of students.

- Estimated value-added. This measure is estimated using student test scores in 4th grade and controls for measures of health at birth (such as birth weight, gestation), family demographics (parents' education, including mother's college entrance exams). Neilson (2021) shows that these measures are strongly correlated with alternate measures that control for baseline test scores, which can only be constructed for a few years of data. Because relatively few schools that enroll high school students (grades nine and up) also enroll the fourth graders whose scores form the basis for this measure, we used VA estimates only for students in grades eight and below.

In addition to the variables used in Neilson (2021), Table 2.3 also describes the effects of the warnings intervention on distance from home to school. We compute this as the Euclidean distance between the home and school, in kilometers.

B.5 Treatment First Stages

This section describes how warnings treatments of different types interact with each other. As stated in section 2.3, our broad goal is to make the point that warnings about application risk affect application behavior, not to disentangle the effects of warning timing or media. Because some readers may nevertheless be interested in the precise combinations of treatments that applicants received in different years, we describe the details here.

Figure B.25 shows how treatment status varies with the value of the initial application risk prediction used in assignment of the platform pop-up in 2018 and 2019. In 2018, there were two kinds of interventions: the initial pop-up, and a followup SMS sent to still-risk applicants. Panel A shows that crossing the risk cutoff raises the count of applications students receive from zero to 1.07. The latter number is slightly above one because applicants who respond to the warning by submitting a revised application can see the warning a second time if the revised application is also risky. Panel B shows that there is a positive association between initial application risk and eventual receipt of the SMS reminder, but that there is not much of a discontinuity in SMS receipt at pop-up cutoff. This is because risk predictions change over time as more applications come in, so designation as risky later in the process is (relatively) smoothly distributed with respect to the initial risk score. Our evaluation of the pop-up RD in 2018 thus reflects the effects of receiving the pop-up warning for initial students, in a setting where the warning may interact with follow-on SMS interventions.

In 2019, the only intervention was the platform pop-up. Panel C of Figure B.25 shows how pop-up receipt varied across the 30 percent predicted risk cutoff. As in 2018, the count of pop-up warnings applicants receive rises by slightly more than one across the cutoff, because applicants who revise their rank lists may receive more than one. The interpretation here is straightforward, since there were no follow-on interventions in this year.

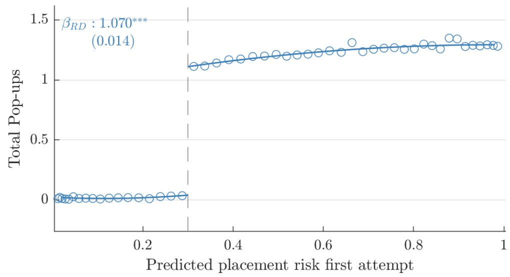
Figure B.26 shows how treatment outcomes for the platform pop-up, the two SMS warnings, and the RCT intervention vary with respect to the initial risk score cutoff for pop-up assignment. These graphs can help us understand how treatments change across the cutoff in our evaluation of the main pop-up RD in 2020. As shown in panel A, crossing the cutoff raises the count of platform pop-ups that applicants receive by a little more than one, just as in previous years. As shown in Panels B through D, receipt of follow-on SMS and WhatsApp interventions is postively correlated with initial risk, but discontinuities in follow-on treatments at the initial risk cutoff are not as sharp, because applicants respond and risk evaluations change. Relatively few students receive the WhatsApp RCT intervention, because this was conducted in a subsample. Our RD estimates of the effects of the 2020 platform pop-up intervention thus reflect the effect of the initial treatment, inclusive of interactions with SMS and WhatsApp follow-ons.

Figure B.27 shows how treatment outcomes for the platform pop-up, the two SMS warnings, and the RCT intervention vary with respect to the risk score cutoff used to assign treatment in the WhatsApp RCT. This score was computed just before the WhatsApp RCT, later in the process than the risk scores we use for assignment platform pop-ups. The sample in these graphs is students assigned to the treatment or control group in the WhatsApp RCT. Most people in both the treatment and control groups for the WhatsApp RCT intervention have already received a platform pop-up and an initial SMS warning (Panels A and B). Reciept of the RCT intervention itself jumps by 42% at the cutoff for treated students; the difference from one reflects the rate at which applicants randomized into the treated group open the message and view the warnings image. Most risky individuals in both the treatment and the control group also receive the follow-on SMS.

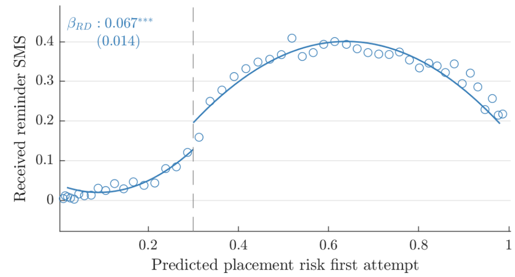
In the main text analysis of the WhatsApp intervention, we present both RCT comparisons of outcomes for treatment and control groups, and RD evaluations across the risk cutoff for both treatment and control. The detailed analysis of treatment receipt presented in Figure B.27 shows that the RCT estimates should be interpreted as intensive-margin effects of the additional WhatsApp intervention, in the context of the platform pop-up and SMS treatments. RD estimates of endline outcomes in the treatment group reflect the combined effect of the WhatsApp treatment, together with the effects of the discontinuous jump in platform pop-up and initial SMS warnings receipt at the cutoff. RD estimates of endline outcomes in the control group reflect similar discontinuous effects in the platform pop-up and initial SMS warning, but not the WhatsApp intervention itself.

Figure B.25: RD First Stages of 2018 and 2019 Interventions

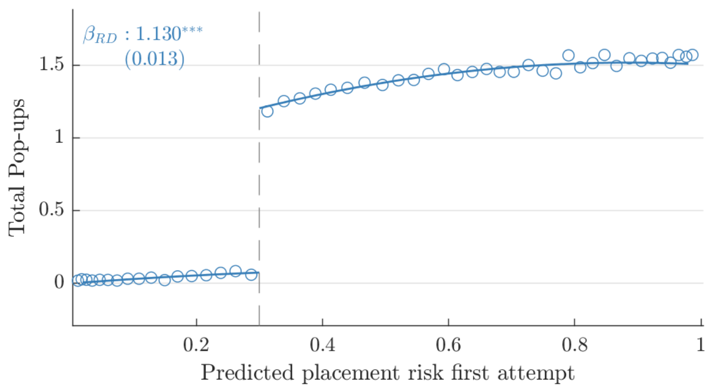
(a) First Stage Total Pop-Ups– 2018



(b) First Stage Reminder SMS– 2018



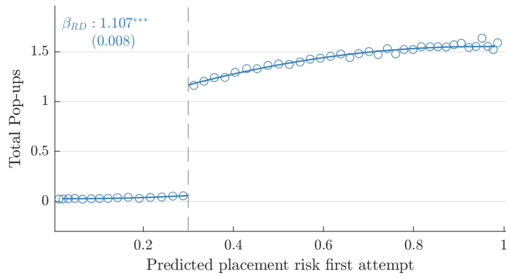
(c) First Stage Total Pop-Ups–2019



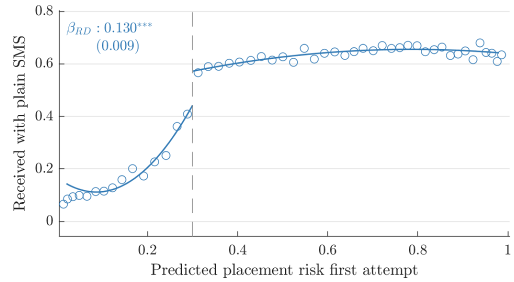
Notes. Warnings interventions in 2018 and 2019, by position relative to risk score prediction. Panel A: Count of platform pop-ups received in 2018 intervention. Panel B: count of reminder SMS messages received in 2018 intervention. Panel C: count of platform pop-ups received in 2019 intervention. 2019 intervention did not include SMS reminders. Points are centered means of 50 quantile-spaced bins of the support of the predicted placement risk $\in [0.02; 0.98]$. Solid line shows the quadratic fit. Reported coefficients and standard errors are from local linear specifications using $+-0.1$ bandwidth. See section 2.3 for intervention details.

Figure B.26: RD First Stages in 2020 Platform Popup Intervention

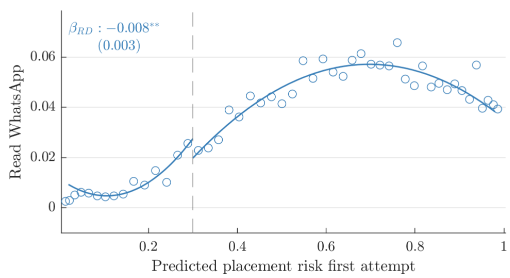
(a) Platform Pop-Up



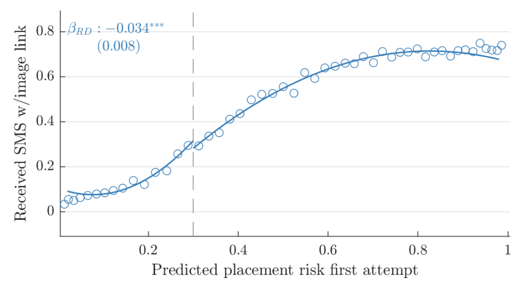
(b) SMS Warning (Pre-WhatsApp RCT)



(c) WhatsApp RCT Intervention

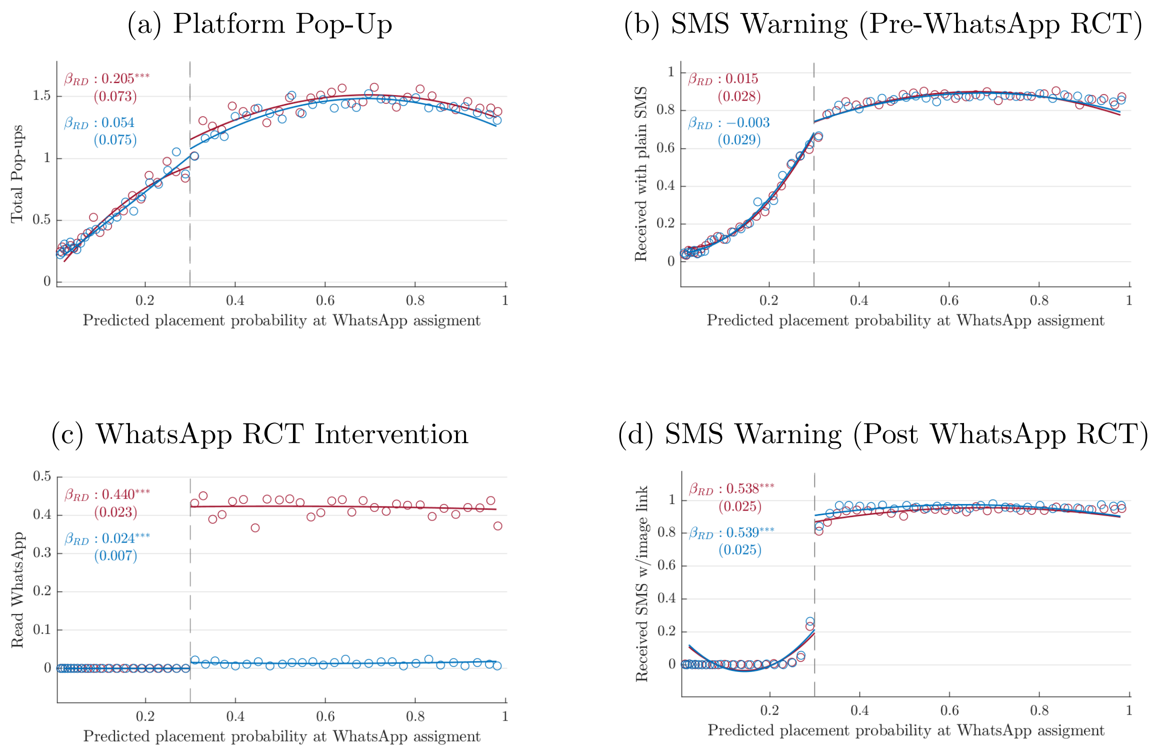


(d) SMS Warning (Post WhatsApp RCT)



Notes. Figure describes the receipt of various warnings interventions by risk prediction at the time of the on-platform pop-up intervention in 2020. For all graphs, risk predictions are as computed at the time of the initial platform pop-up. Panel A: count of platform pop-ups. Panel B: count of pre-WhatsApp reminder SMS interventions. Panel C: share of individuals reading the randomized WhatsApp intervention. Panel D: share of individuals receiving follow-up SMS, after WhatsApp. Points are centered means of 50 quantile-spaced bins of the support of the predicted placement risk $\in [0.02; 0.98]$. Solid line shows the quadratic fit. Reported coefficients and standard errors are from local linear specifications using $+-0.1$ bandwidth. See section 2.3 for intervention details.

Figure B.27: First Stages of 2020 WhatsApp Intervention



Notes. Figure describes the receipt of various warnings interventions by risk prediction at the time of the WhatsApp RCT intervention in 2020. Sample is the treatment group in the RCT intervention only. For all graphs, risk predictions are as computed at the time of RCT intervention, not the initial platform pop-up. Panel A: count of platform pop-ups. Panel B: count of pre-WhatsApp reminder SMS interventions. Panel C: share of individuals reading the randomized WhatsApp intervention. Panel D: share of individuals receiving follow-up SMS, after WhatsApp. Points are centered means of 50 quantile-spaced bins of the support of the predicted placement risk $\in [0.02; 0.98]$. Solid line shows the quadratic fit. Reported coefficients and standard errors are from local linear specifications using $+-0.1$ bandwidth. See section 2.3 for intervention details.

B.6 Risk Prediction

Obtaining Admissions Chances from Application Data

We define risk as nonplacement probability. In this section we describe how the NGO computed the simulated values of application risk that form the basis for the information interventions.

The NGO took a simulation-based approach. This required three inputs for each market-year: a seat count for each school and grade, a projected total number of applicants N, and a joint distribution of applications and student types (i.e., priorities). At each iteration of the simulation, we resample N applications from this joint distribution, and redraw the lottery tiebreakers. We then run the deferred-acceptance algorithm and determine the probability of admission for applicants in each school-grade-priority group combination conditional on not being admitted to a more-preferred option. For example, if 50 people in priority group p apply to school-grade j and did not get into an option they prefer, and 10 of these people are admitted to j, then the simulated placement probability for this school-grade-priority group cell is $10/50 = .2$. We repeat this process 500 times, and then compute the average admissions probability for each school-grade-priority group over all the simulations. This procedure is similar in spirit to Agarwal and Somaini (2018), but differs in that we average over probabilities within priority groups rather than cutoff scores. Given this set of (conditional) admissions probabilities for each school-grade-priority cell, applications are identified as risky if nonplacement risk is above 30%.

Simulated Demand

We now describe how the NGO approximated the count and joint distribution of applications and priorities. The empirical distribution of applications and priorities is not observed until after the application deadline. In practice, the NGO had to project this distribution based on a combination of past data and early applications to the centralized system.

We begin the demand prediction procedure by identifying urban zones that correspond to education markets. We do this using the urban polygons of the census of 2017, similar to Neilson (2021). We merged two polygons if they were close enough or if we observed a substantial flow of students from one market to another. Each market is then defined by a set of polygons and a grade. A school belongs to a market if its location lies inside one of the polygons and offers the grade that defines the market. We consider schools located outside of every polygon as part of a "rural market" defined by region and grade.

Given these market definitions, the NGO took two different approaches to simulate the distribution of applications and priorities. The first approach was used for 2020 and 2021 simulations. This approach uses the previous year's applications to estimate this year's congestion, given the current-year supply of seats in each academic program by grade combination. A challenge for this approach is that the menu of available programs is defined not just by school and grade, but also in some cases by gender, "shift" (morning, afternoon, or both), campus, and specialization (only for grades 11 and 12). In some cases, the menu of available options along these additional dimensions shifts over time. The approach the NGO took was to first match on all of these attributes, and keep available one-to-one matches. Then, for programs without a one-to-one match, they matched on school, grade, and gender only, and manually inspected cases to determine the best of multiple options when more than one was available, allowing for one-to-one or one-to-many matches. Completely new academic programs were assigned values of zero risk; applicants listing new programs did not receive risk warnings. This approach followed from the idea that the goal was to identify known risks, rather than speculate on what options might prove to be risky, and with the knowledge that the there would be later followup based on observed same-year application counts. The advantage of this approach is that it can be applied from the beginning of the choice process, before any applications are submitted.

The second, which was used for the 2018 and 2019 processes, and beginning in the sixth day of the 2020 choice process, resamples from current applications to meet the expected count of applicants within each market. i.e., we estimate the count of applicants N in a given market, then resample N applications from the $M < N$ applicants who have already applied in the market. This method cannot be used for applicants at the very beginning of the application process; this is why students applying in the first few days of the 2018 and 2019 processes are omitted from the platform pop-up intervention (see above).

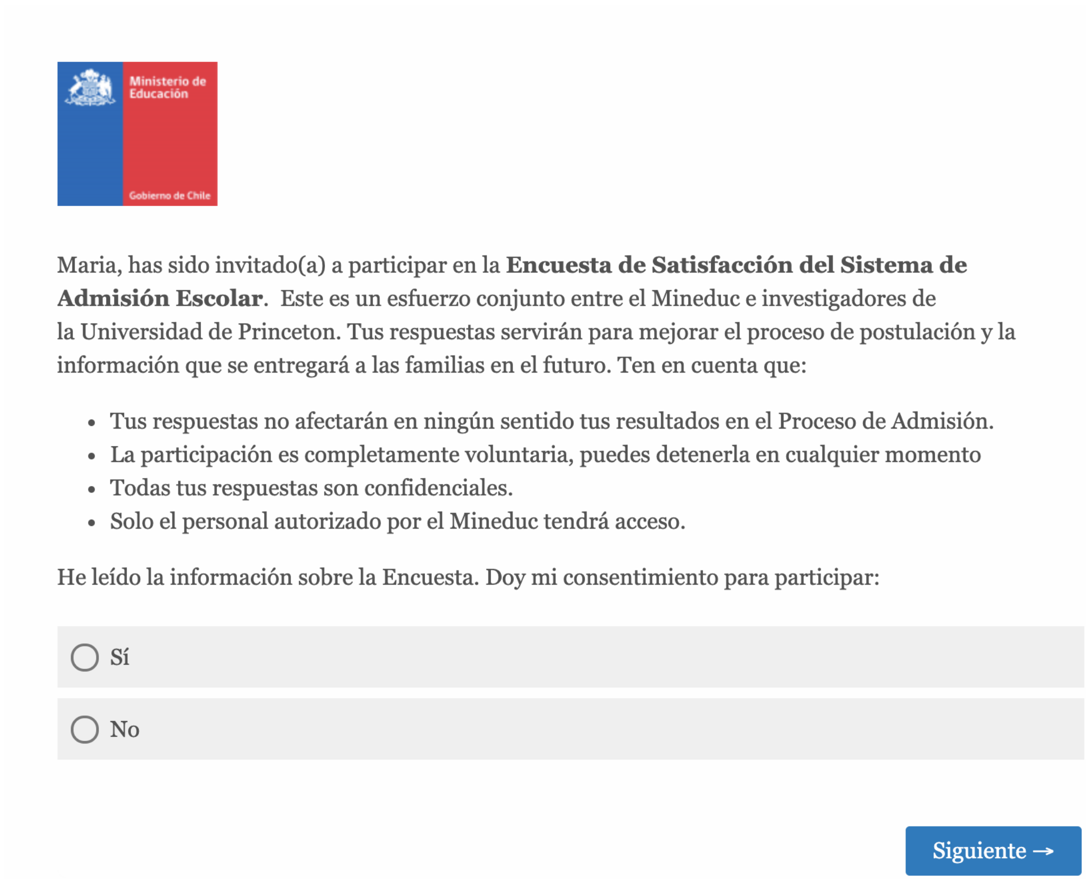
We compute the expected count of applicants, N, in each market as the number of students who participated in choice in the market in the previous year, or, in markets where there is no previous-year data, as the number of students who entered a new school at the start of the most recent school year (i.e. who switch schools between December and March).

Changes in Risk Predictions within the Application Cycle

Risk predictions change during the application process as more applications are filed. Applicants may have different values of predicted risk at the moment of application and when the SMS reminder is sent, even if they do not change their rank-order lists. Risk predictions presented on the platform pop-up and in the various SMS and WhatsApp interventions reflect live updates about the count and distribution of applications conducted every 3 days.

B.7 2020 Survey Translation

Figure B.28: 2020 Survey Landing Page



Notes. This is the website displayed after applicants clicked the invitation link to participate in the 2020 survey. 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
    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 didn'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 didn't know it, Only by name, I know it well)*]

    a) [School not considered in application 1]

    b) [School not considered in application 2]

    c) [School not considered in application 3]

d) [School not considered in application 4]

e) [School not considered in application 5]

19. From 1 to 10,How easy it is to find information on the academic performance of schools?
[*Slider 1 to 10*]

20. Imagine that you spend time researching all schools that you do not know well.After you know them well, do you think you would add at least one of these schools to your application?
[*Select one*]

    a) No

    b) Yes

21. From 0 to 100, how likely would you add it as your first preference?
[*Slider 0 to 100*]

22. From 0 to 100, how likely would you add it below your last choice?
[*Slider 0 to 100*]

23. During the application process, did you get any recommendations about adding more schools to your list?
[*Select one*]

    a) No

    b) Yes

24. By what method did you receive the recommendation to add more schools?(Select all that apply)
[*Select multiple*]

    a) SMS

    b) WhatsApp

    c) E-mail

    d) Web page

    e) Other

25. By what method did you receive the recommendation to add more schools?- Other
[*Open text*]

26. If [applicant name] get a seat in the following schools, from 1 to 7, how satisfied would you be?
[*Slider 1 to 7*]

a) First preference: [Name preference 1]
b) Last Preference: [Name Last preference]
c) If you are not in any school in the regular period

27. Would you like to have had the following information on schools that did not have at the time of application?
[*Yes or No*]

    a) Information about your chances of being accepted
    b) Standarized test score
    c) Performance category
    d) Price
    e) Priority for economically-vulnerable students
    f) SAT scores
    g) Seats available

28. What is your preferred method of contact during the application process?
[*Select one*]

    a) E-mail
    b) Other
    c) SMS
    d) Telephone
    e) WhatsApp

29. What is your preferred method of contact during the application process? - Other
[*Open text*]

30. For registration purposes only, what is the highest educational level of the Mother (or Stepmother) of [applicant name]?
[*Select one*]

    a) Educación Básica Completa
    b) Educación Básica Incompleta
    c) Educación Media Completa
    d) Educación Media Incompleta
    e) Educación incompleta en una Universidad
    f) Grado de magíster universitario

g) No estudió

h) Titulada de un Centro de Formación Técnica o Instituto Profesional

i) Titulada de una Universidad

31. Do you know if [Field-nomPostulante] is a priority student (SEP)?
[*Select one*]

 a) He/she is not a beneficiary of the preferential subsidy

 b) I do not know

 c) He/she is a beneficiary of the preferential subsidy

32. Do you have any other comments, complaints or suggestions to make us?
[*Open text*]

B.8 Additional School Quality Results

This appendix provides more detail on the effects of warnings on school quality reported in section 2.5 and Table 2.3 of the main text. The main finding is that the overall gain in school value added arises from both a shift towards oversubscribed schools, which are higher value added on average, and from shifts in value added within oversubscription status.

Appendix Figure B.29 plots the distribution of school value added and per-teacher spending in schools that are oversubscribed and schools that are not. Because our goal is to understand how these measures vary by oversubscription status for a set of schools that a particular student might choose between, we restrict the sample schools offering pre-kindergarten in Santiago (the largest single "market" in our data).

On average, oversubscribed schools have higher value added and higher per-teacher spending than undersubscribed schools, but there is much dispersion within each category and the distributions overlap substantially. The implication is that the search treatment may in principle raise value added or teacher spending at the schools students attend by shifting students toward over-subscribed schools or by improving value added within oversubscription category.

Appendix Table B.10 decomposes the overall gain in value added reported in Table 2.3 into within- and between- oversubscription status channels. Results are imprecise in some cases but nevertheless provide insight into the channels through which value added rises across the cutoff.

The first two rows repeat our main first stage (i.e., add any school) and value added results. The next two rows report that the warnings intervention pushed students to enroll in oversubscribed schools. This holds both overall and in the sample of students who enroll in schools where VA measures are available, though effects are larger in the full sample. Note that these effects on *enrollment* are not the same as results for *placement* in undersubscribed schools reported in Table 2.5. This is as expected given the imperfect compliance with placed outcomes reported in Table 2.1.

Rows five and six show the within vs. between decomposition results. $E[VA|type]$ is the mean value added for students given the type of school they attend, where type is either oversubscribed or undersubscribed. This rises across the cutoff because students are more likely to enroll in oversubscribed schools, where mean value added is higher. These between type gains account for about 20% (0.023/0.103) of the overall gains in value added we observe for compliers with the warnings treatment. $VA - E[VA|type]$ is the difference between the value added of the school where the student enrolls and the type specific mean. This rises across the cutoff because students attend higher value added schools within oversubscription type. Within-type shifts account for about 80% (0.080/0.103) of value added gains for compliers. The within and between effects mechanically add up to the full effect.

The final two rows report shifts in value added across the cutoff conditional on enrollment in either an over- or undersubscribed school. The goal of these specifications is to understand whether the within-type shifts come from over- or under-subscribed schools. We interpret the results with caution because, as documented in the upper rows of the table, the share of students enrolling in oversubscribed schools rises across the cutoff. The (imprecise) results from these specifications suggest that the within-type value added gains come from both over- and undersubscribed schools.

Gains from the shift towards oversubscribed schools and the shift within oversubscribed schools are consistent with the observations that a) most of the highest quality schools are obsersubscribed, b) all spots at these schools are by definition allocated in the centralized match, and c) changes to the application list induced by the warnings intervention raise the probability of being assigned to any school in the centralized match and also shift the distribution of schools to which students are assigned.

Gains from shifts within undersubscribed schools (where spots remain open after the match) suggest the possibility that, even absent capacity constraints, finding a good school may be easier within the centralized process than in the "scramble" that follows it. This pattern could arise if, for example, families searching for schools in the scramble feel pressure to accept the first offer they receive, because they are concerned the school will fill up. Understanding how search plays out in the scramble is a topic of possible interest for future work.

Table B.10: RD Estimates of Platform Pop-Up Effects

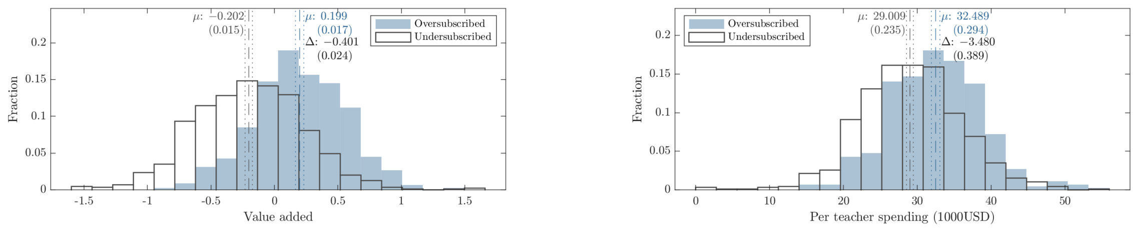
	(1)	(2)
	All	
		IV
Add any	0.216	
	(0.010)	
Value added (VA)	0.022	0.103
	(0.011)	(0.051)
Enrolled in oversubscribed ($type = over$)	0.039	0.178
	(0.010)	(0.048)
Enrolled in oversubscribed ($type = over$)|not missing VA	0.022	0.103
	(0.014)	(0.065)
$E[VA\|type]$	0.005	0.023
	(0.003)	(0.014)
$VA - E[VA\|type]$	0.017	0.080
	(0.011)	(0.050)
$(VA - E[VA\|type])\|type = under$	0.029	0.131
	(0.020)	(0.091)
$(VA - E[VA\|type])\|type = over$	0.013	0.063
	(0.012)	(0.059)
NL	10,782	10,782
NR	11,285	11,285

Notes. Local linear RD estimates of pop-up effects from warning pop-up on application platform. Computed using triangular kernel with bandwidth 0.1. Heteroskedasticity-robust nearest neighbor variance estimator with minimum of 3 neighbors reported in parentheses; computed as in Calonico et al. (2014). We report estimates in the pooled sample across years 2018-2020. IV (column 2) shows the instrumental variable specifications, where the endogenous regressor is the add any school indicator. "Add any" is the first stage indicator for adding at least one school to the choice application. "Value added (VA)" repeats main text results on value added at the enrolled school. "Enrolled in oversubscribed" is an indicator for enrolling in a school with a binding capacity constraint. "Enrolled in oversubscribed | not missing VA" restricts the sample to students enrolling in schools where value added measures are available. "$E[VA|type]$" is the mean value added for the enrolled school given over/underscription status, and "$VA - E[VA|type]$" is the deviation of value added at the enrolled school from the type-specific mean. "$VA - E[VA|type]|type = over/under$" additionally conditions on the sample on students enrolling in over- or under-subscribed schools, respectively.

Figure B.29: Distributions of School Value Added and Spending by Oversubscription Status

(a) Value Added

(b) Spending Per Teacher



Notes. Histograms of value added (panel A) and spending per teacher (panel B) for schools that are over- and under-subscribed. Means by over/undersubscription status and the difference between means reported in each panel and indicated by vertical lines. Sample is schools offering pre-Kindergarten in Santiago (the largest single market in our dataset).

B.9 RDs at Multiple Cutoffs from the 2017 Pilot

A 2017 pilot of the platform pop-up intervention provides additional evidence on the effects of warnings across the risk distribution. The pilot was essentially identical to the 2018-2020 intervention, but a) was limited to markets that had implemented centralized choice by 2017, and b) varied the cutoff across cities, with some cities having cutoffs of 30%, others 50%, and others 70%. Appendix Table B.11 reports results from the pilot for the pooled sample, and split by risk cutoff. The sample size is roughly 3% as large as in our main analysis, so estimates are noisy. However, the pooled sample effects on the key add any school and change in application risk outcomes are quite similar to what we see in the main intervention. Splitting across cutoff values, point estimates are largest at the 50% cutoff, and smaller for the 30% cutoff than for the 70% cutoff. These results provide further support for the idea that warnings interventions have large effects across the distribution of application risk.

Table B.11: RD Estimates of Platform Pop-Up Effects - 2017

	(1)	(2)	(3)	(4)	(5)
Risky cutoff	Pooled		0.3	0.5	0.7
		IV			
Any modification	0.237		0.107	0.352	0.137
	(0.059)		(0.112)	(0.086)	(0.140)
Add any	0.192		0.024	0.321	0.132
	(0.056)		(0.101)	(0.084)	(0.135)
Schools Added	0.428	2.225	0.109	0.651	0.394
	(0.174)	(0.739)	(0.350)	(0.214)	(0.322)
Δ Risk	-0.060	-0.313	-0.003	-0.103	-0.046
	(0.024)	(0.099)	(0.039)	(0.038)	(0.051)
Placed to preference	0.120	0.624	-0.014	0.210	-0.003
	(0.058)	(0.314)	(0.098)	(0.077)	(0.155)
NL	671	671	187	354	130
NR	647	647	194	334	119

Notes. Local linear RD estimates of pop-up effects from warning pop-up on application platform. Computed using triangular kernel with bandwidth 0.1. Heteroskedasticity-robust nearest neighbor variance estimator with minimum of 3 neighbors reported in parentheses; computed as in Calonico et al. (2014). We report estimates in the pooled sample and for each different risky cutoff definition. IV (column 2) shows the instrumental variable specifications, where the endogenous regressor is the add any school indicator.

B.10 2020 WhatsApp Treatment Arms

The 2020 WhatsApp RCT tested the efficacy of different forms of the warnings intervention. We presented the personalized risk warning in three ways: 1) a text-only warning that nonplacement risk was high, 2) a visual risk display, with a red bar indicating high risk, and 3) a list of the schools the student had applied to displaying the count of applications and estimated places available at those schools. See Appendix Figure B.22 for images of each arm.

Our main analysis in Table 2.4 pools across arms. Appendix Table B.12 separates estimates by arm. All three treatments caused students to add schools to their application. Effects in the visual display and application list arms were 25-35% larger than for the text only arm. Making warnings more salient or informative improves performance, but most of the gains are from the 2.

Table B.12: RCT Estimates Split by WhatsApp Message

	(1)	(2)	(3)	(4)
	Personalized	Personalized - By image type		
	Pooled	Simple	Warning bar	Vacancy
Add any (clean)	0.033	0.028	0.035	0.036
	(0.002)	(0.003)	(0.004)	(0.004)
Add any (endline)	0.044	0.036	0.049	0.046
	(0.003)	(0.005)	(0.006)	(0.005)
N Treatment	8,995	3,009	2,971	3,015
N Control	8,975	8,975	8,975	8,975

Notes. Evaluation of 2020 WhatsApp RCT, splitting out by message type. Treatment arms are as described in Appendix B.10. Appendix Figure B.22 shows example images. Estimates are differences in the share of students adding any school to their baseline application between the treatment group and a control group that did not get any message. Column 1 contains pooled estimates of the treatments from columns 2-4. "Clean window" corresponds to outcomes measured 44 hrs after WhatsApp messages were sent. "Endline" outcomes were measured at the end of the application process (75 hrs after WhatsApp message). See section 2.3 for a description of the WhatsApp RCT.

B.11 New Haven Warnings Policies and Data

School Choice Institutions

New Haven, Connecticut uses a centralized mechanism to assign students to schools in all grades. As reported in Akbarpour et al. (2020), New Haven used a Boston- or Boston-like mechanism to assign students to schools from the the 1990s through 2018. In 2019, New Haven switched to a truncated deferred acceptance mechanism (DA-MTB), which allowed applicants to list a maximum of four schools. In 2020, the district raised the maximum application length to six. The school choice process in New Haven takes place in the winter and early spring of each year, for enrollment the following fall. Following a series of informational events in January, the choice application opens in early February, with final applications due in early March. Applicants receive placement outcomes in early April.

Students from outside New Haven may attend NHPS schools through the district's interdistrict choice program. Choice "markets" in the the NHPS school choice system are defined by grade and by residency. Counts of available seats are determined separately for each school-grade combination, and then are further split by whether the applicant is from New Haven or a nearby town.

The first column of Table B.13 reports descriptive statistics for all students applying through the choice system in 2020. 58% of applicants apply to a major transition grade– either pre-kindergarten, kindergarten, or grade 9. We focus on these transition grades in our analysis risk warnings. This is because we rely on prior-year risk predictions, and these are more stable in larger markets.

Warnings Intervention

NHPS policymakers conducted two information intervention policies as part of the 2020 choice process for PreK, Kindergarten, and high school grades. The first was a warnings intervention applied to all risky applicants. In this intervention, students submitting applications with an estimated nonplacement risk of 50% or higher received an email warning one week before the application deadline. The email suggested that the applicant might want to add more schools to their rank list. The email also provided a link to an online risk simulator tool, where applicants could input hypothetical choice applications and learn about the chances of placement for those applications. Panel A of Figure B.30 displays an example of the email sent to risky applicants.

The second intervention consisted of an email sent to a randomly chosen group of nonrisky applicants. This email was identical to the warnings treatment email, but did not contain the line about high nonplacement risk. In the main text, we refer to this email as the "encouragement nudge" intervention.

Panel B of Figure B.30 displays the email sent as part of this second intervention. Randomization was stratified by market (grade by residency status). All school choice applicants could view the application simulator using their NHPS username and login, once they arrived at the simulator page. As we show below, simulator use by untreated individuals was rare. This makes sense because control group applicants did not receive information about the simulator's web address.

The application simulator website used in both interventions worked as follows. Applicants where first asked to state their beliefs about the admission chance for each of their choices, which were pre-loaded (see Figure B.31a). Afterwards, their predicted admission chances were displayed to them as shown in Figure B.31b. Users then had the opportunity to add, remove or change schools and received immediate feedback on their changed portfolio risk. The schools available to them were shown both on a map and an alphabetical list (see figure B.31c and B.31d).

In contrast to the risk predictions we constructed for the Chilean choice intervention, the predictions in the New Haven setting relied only on prior-year applications. Specifically, NHPS computed portfolio risk predictions based on the admission chances that the same application would have had in the previous application year.[7]

Figure B.32 describes the distribution of predicted placement probabilities for different values of realized placement probabilities. As with our predictions in the Chilean setting, risk predictions do not perfectly match ex post values, but do closely track them. One point of contrast with the Chilean setting is that our New Haven placement probability predictions tend to somewhat overstate true placement chances— lottery odds became somewhat worse for applicants in 2020, relative to submitting the same application in 2019. In practice, this meant that the risk warnings went to a *riskier* set of applicants than would have been the case had placement chances remained steady.

Two new schools entered the NHPS system in 2020. NHPS did not compute risk predictions for applications including these schools, and excluded applicants from the information intervention. Stepping down from the full sample to the sample of intervention students reduced the sample size as follows. 58% of all choice applicants applying in intervention-eligible grades. 46% submit applications by seven days before the application deadline and are included in the intervention procedure. 36% (of the full sample) applied to simulator-eligible grades, did so in time to be included in the intervention, and applied only to schools included in the simulator. These 36% of applicants formed the universe of applicants potentially subject to the warnings and simulator interventions.

Column 2 of Table B.13 describes the 36% of students in the intervention-eligible universe. These students are less likely to be African American and more likely to be Hispanic than the full sample.[8] Columns 3 and 4 of Table B.13 split the sample by treatment assignment

[7]To make sure applicants understood this, the text of the intervention stated that the warning was based on past data. See Figure B.30.

[8]A few additional features of the data are worth noting. Nine students who have been assigned treatment later change the grade they are applying to or delete their profile altogether. These would be counted in the column 2 sample but are not considered to have an eligible grade. Thus the share of applicants to an eligible grade is not exactly one. In addition, 14 students apply on February 26 after the application portfolio snapshot was taken at 7.00pm but before the last wave of treatment assignments is made. We classify these

of either the warnings intervention or the simulation intervention, which corresponds to predicted risk levels above and below 50%. The mean risk score in the former group is 89%, in the latter group it is 5.4%. 65% of all applicants in the intervention-eligible sample received an email. The remaining 35% where either assigned to the control group (33%) which did not receive any emails or the email could not be delivered (3%). 98.2% of students in the warnings group and 96% of students in the simulator intervention group successfully received an email corresponding to their treatment group.

We also construct a comparison sample of choice applicants in 2019, consisting of all choice applicants applying in the major choice grades in that year. We construct this comparison sample to resemble as closely as possible the set of students who would have been included in a 2019 warnings intervention, had one taken place. We do this by considering only students in eligible grades who had submitted their application at least seven days in advance of the application deadline. We compute risk predictions for this group using the 2019 application data based on the state of their application as of seven days before the deadline. The students in this sample form the basis for the 2019 comparison group plotted in Figure 2.8. Column 5 of Table B.13 displays descriptive statistics for this comparison sample. This group closely resembles the 2020 eligible sample on demographic characteristics and choice outcomes.

Intervention Results

Figure 2.8 displays our main results for the RD and DD analysis of the warnings intervention in 2020. We discuss these findings in section 2.7 of the main text. Appendix Figure B.33 displays RD-DD graphs for balance and first stage outcomes, parallelling the main results in Figure 2.8.

Appendix Table B.14 provides additional detail on the effects of the warnings policy beyond what is reported in Figure 2.8. We report two kinds of effect estimates. The first are RD estimates using only the 2020 data. The RD specifications allow for separate slope terms above and below the cutoff value, and include all data except for mass points at risk values of zero and one. The second are difference-in-difference estimates where the first difference is 2020 vs. 2019 and the second difference is above vs. below the warnings threshold. The difference-in-difference specifications control for risk-group fixed effects in ten percentage point bins. Both RD and DD estimates pool across the encouragement nudge and no-contact control group among non-risky applicants. We do this because average outcomes for these groups are essentially the same.

individuals as not having applied in time. This is why the share of applicants to an eligible grade that apply in time is smaller than one. As discussed in Kapor et al. (2020), New Haven residents applying to ninth grade have undersubscribed neighborhood high schools as their default placements. When computing application risk, the district defined these outcomes as placements, meaning that no in-district ninth-grade applicants were classified as risky for the purposes of the warnings intervention. The ninth grade applicants who did receive the warning were those applying through interdistrict choice programs. This is why the share of ninth graders in the warnings sample is smaller than the share in the eligible sample.

Panel A of Appendix Table B.14 shows that predetermined characteristics are balanced across the cutoff, although estimates of changes in female and Black share are imprecise. Panel B shows that while nearly all above-threshold students received a warnings email, relatively few logged into the online simulator or ran a simulation. This suggests that the behavioral effects we observe come mostly from the warning and not from the simulator availability. This is consistent with the large effects in the Chilean setting, which did not include a simulator. The implication is not that risk simulators cannot form part of an effective intervention, but that effective interventions do not require simulation.

Panel C shows estimates of effects on different choice outcomes. The RD estimates indicate that crossing the threshold and receiving the warning causes 13.8 percent of applicants to add at least one school to their application. These are the compliers with the information treatment. Ex post realized application risk falls by 3.2 percentage points across the cutoff. This means that compliers with the policy reduce their application risk by 23.2 percentage points (= 0.032/0.138), or 42% of the below-threshold mean ex post risk of the initial application. Compared to the Chilean setting, the complier population is somewhat smaller, while risk falls by more per complier in absolute terms, and the reduction as a share of baseline risk levels is similar.

Comparing the RD and DD estimates confirms the visual impression from Figure 2.8 that behavioral changes are larger for less risky students in the risky group, though estimates are imprecise. A possible explanation is that the highest-risk applicants are those applying to a small number of highly desirable schools. These applicants may have outside options beyond the public system and be uninterested in additional inside options (Akbarpour et al., 2020).

Appendix Table B.15 reports our findings from the encouragement nudge RCT delivered to randomly chosen non-risky students. Panel I shows that standard balance tests pass. Panel II shows that treatment and control groups are balanced on the initial risk prediction ("risk score") as well as on the ex post risk associated with their initial application ("initial realized risk"). It also shows that there is no difference in *final* realized risk; i.e. the ex post risk of the final submitted application. The implication is that simulator access did not affect application risk.

Panel III shows that many applicants who receive the treatment email granting simulator access do in fact click the link and interact with the simulator. Treatment rasies the likelihood of simulator login by 23 percentage points and the share of applicants conducting simulator runs by 11 percentage points. In practice, the requirement that applicants state their beliefs about admissions chances at different schools prior to each use of the simulator may have placed a substantial burden on prospective simulation users. NHPS eliminated this requirement from subsequent implementations. Panel IV shows how treatment changes choice behavior. As with the headline risk values reported in panel III, we see no evidence of effects here.

Table B.13: Sample Descriptives New Haven

	2020				2019
	All grades	Eligible	Warnings	Simulator	Comparison group
I. Demographics					
Female	0.513	0.539	0.510	0.530	0.547
African American	0.432	0.338	0.366	0.330	0.380
Hispanic	0.395	0.468	0.389	0.489	0.432
White	0.125	0.147	0.203	0.133	0.139
NH Resident	0.725	0.674	0.429	0.759	0.719
II. Simulator Eligiblity					
PreK3	0.075	0.105	0.268	0.046	0.129
PreK4	0.105	0.147	0.294	0.094	0.190
K	0.163	0.288	0.229	0.314	0.291
Grade 9	0.242	0.460	0.208	0.547	0.390
Apply to eligible grade	0.577	0.996	0.992	0.999	1.000
+ in time	0.458	0.991	0.988	0.994	1.000
+ only to simulator schools	0.363	1.000	1.000	1.000	1.000
III. Interactions with simulator					
Risk score	0.294	0.294	0.889	0.054	0.339
Warnings email	0.105	0.285	0.982	0.000	0.000
Received email	0.238	0.649	0.982	0.960	0.000
IV. Placements					
Placed 1^{st}	0.259	0.320	0.132	0.399	0.304
Placed other	0.337	0.369	0.097	0.477	0.375
Unplaced	0.403	0.311	0.771	0.124	0.321
V. Choice outcomes					
Change length or school		0.056	0.100	0.033	0.030
Lengthen app.		0.042	0.092	0.020	0.019
Insert new school		0.018	0.025	0.014	0.011
Append new school		0.024	0.065	0.008	0.010
Change school		0.029	0.031	0.023	0.022
Shorten app.		0.011	0.007	0.013	0.006
Difference in realized risk		-0.006	-0.015	-0.002	-0.003
Difference in simulated risk		-0.006	-0.019	-0.000	-0.003
N	7027	2551	740	967	3150

Notes. Samples vary by column. The first column consists of all applicants in 2020. The second column consists of the sample that was eligible for treatment in 2020. Columns 3 and 4 represent the subsamples that have been assigned to either treatment group. The fifth column consists of those applicants in 2019 that would have been eligible for treatment had their been any. Statistics reported represent shares of applicants in the respective sample or the mean difference in the last two rows of panel V.

Table B.14: RD and DD Estimates of Warnings Effects in New Haven

	RD		Diff. in Diff.	
Outcome	β	SE	β	SE
A. Demographics				
Female	0.113	(0.083)	0.001	(0.033)
African American	0.077	(0.072)	0.040	(0.031)
Hispanic	0.033	(0.082)	-0.003	(0.032)
White	-0.029	(0.073)	0.006	(0.025)
N		740		3918
B. Interaction with Simulator				
Warnings email	1.001	(0.010)		
Pr(Any login)	0.133	(0.074)		
Number of Logins	0.138	(0.090)		
Pr(Any sim. run)	0.068	(0.063)		
N		740		
C. Choice Outcomes				
Change length or school	0.146	(0.047)	0.042	(0.015)
Lengthen app.	0.138	(0.046)	0.053	(0.013)
Insert new school	0.053	(0.028)	0.012	(0.007)
Append new school	0.089	(0.038)	0.039	(0.011)
Change school	0.050	(0.028)	0.002	(0.009)
Shorten app.	0.004	(0.012)	-0.008	(0.005)
Diff. in realized risk	-0.032	(0.012)	-0.007	(0.004)
Diff. in simulated risk	-0.033	(0.016)	-0.014	(0.004)
Any realized risk reduction	0.156	(0.044)	0.036	(0.010)
Any simulated risk reduction	0.135	(0.045)	0.046	(0.011)
N		740		3918

Notes. RD and difference-in-differences estimates of the effects of the New Haven, CT warnings intervention. The samples for these regressions consist of the universe of applicants to grades PreK, and K in the NHPS simulator study i.e. that have been randomized into either control or one of the two treatment groups or the equivalent comparison group in the 2019 application process. RD specifications are based on local linear fit, dropping observations with predicted portfolio risk of of less than 1% or more than 99%. For the difference-in-differences panel, no observations are dropped based on their risk score. Robust SEs in parentheses. See section 2.7 for details.

Table B.15: Treatment Balance RCT

	Control	Treatment	Difference	
	Mean	Mean	β	SE
I. Demographics				
Female	0.574	0.530	-0.043	(0.024)
African American	0.323	0.330	0.010	(0.022)
Hispanic	0.512	0.489	-0.026	(0.023)
White	0.114	0.133	0.018	(0.015)
Repeat grade	0.063	0.057	-0.003	(0.011)
II. Risk				
Risk score	0.046	0.054	0.002	(0.005)
Initial realized risk	0.124	0.136	-0.002	(0.009)
Final realized risk	0.123	0.134	-0.003	(0.009)
III. Interaction with simulator				
Received email	0.000	0.960	0.960	(0.006)
Warnings email	0.000	0.000	0.000	(0.000)
Pr(Any login)	0.018	0.224	0.198	(0.014)
Number of Logins	0.023	0.260	0.227	(0.017)
Pr(Any sim. run)	0.012	0.126	0.109	(0.011)
IV. Choice outcomes				
Change length or school	0.043	0.033	-0.005	(0.009)
Lengthen app.	0.022	0.020	0.002	(0.007)
Insert new school	0.016	0.014	0.002	(0.006)
Append new school	0.006	0.008	0.001	(0.004)
Change school	0.035	0.023	-0.007	(0.008)
Shorten app.	0.012	0.013	0.001	(0.005)
Difference in realized risk	-0.001	-0.002	-0.002	(0.001)
Difference in simulated risk	-0.000	-0.000	-0.001	(0.001)
Any placement	0.878	0.876	0.016	(0.012)
N	844	967	1811	1811

Notes. Statistics in this table are estimated from the sample of individuals in the control group and the Simulator-only (no warnings) treatment group. The column panels distinguish between these two subsamples. The reported coefficients β reflect regression estimates of the treatment indicator on outcomes, controlling for fixed effects of randomization time blocks and markets, i.e. resident status by grade.

Figure B.30: Email Communication with Parents

Office of Choice & Enrollment
54 Meadow Street, New Haven, CT 06519
NHPS School Choice Simulator Letter

NEW HAVEN
PUBLIC SCHOOLS OF CHOICE

Date: 2020/02/25

Dear Parent or Guardian of Jimmy Heckman,

Thank you for your participation in the New Haven Public Schools of Choice Program. You have successfully submitted an application for Jimmy!

Remember that you can make changes to your child's application until **March 2nd, 2020.**

We are contacting you because wanted to flag that the schools listed on the application you submitted for Jimmy have in the past have had many more applicants than slots. This means that the chances of being placed at those schools are often low. Our calculations suggests had you submitted this same application **last year**, there would have been **at least a 50% chance you would have not been placed at any of the choices listed on your application.**

This does not mean that you have to change your application or that you will not be placed anywhere this year. This message is meant only to be a timely alert. We advise you to apply to your favorite schools and rank them in the order that you like them. To increase your chances of being admitted to some school, what you may want to do is look into additional schools and add them to your application.

To learn more about your **chances of being placed in the schools on your application**, we strongly encourage you to use the **New Haven Public Schools Application Simulator**.

To use the Application Simulator:

1. Visit our simulator website.
(Please use a web browser different from Internet Explorer, such as Google Chrome or Microsoft Edge).
2. Log in using the **same** user name and password used in the NHPS School Choice website.

The **Application Simulator** can help you see what the chances of being placed in each school in your application would have been last year. You can also search, explore, and add new schools.

Note: The Application Simulator shows the chances of school placement based on data from last year, 2019. **The simulator does not show placement results for the 2020 School Choice Process**. The number of applications and available seats do not repeat exactly from year to year, so placement chances this year will differ somewhat from chances last year. Information from last year may still be a useful guide for families submitting choice applications.

For more information, please contact us:

- Email us at simulator@new-haven.k12.ct.us
- Visit the **Choice & Enrollment Office** at 54 Meadow Street, New Haven, CT 06519.
- Contact us by phone: **(475) 220-1438**.

Again, thank you for choosing New Haven Public Schools of Choice!

Sincerely,
Choice & Enrollment Office

(a) Risky Applicants

Office of Choice & Enrollment
54 Meadow Street, New Haven, CT 06519
NHPS School Choice Simulator Letter

NEW HAVEN
PUBLIC SCHOOLS OF CHOICE

Date: 2020/01/30

Dear Parent or Guardian of Jimmy Heckman,

Thank you for your participation in the New Haven Public Schools of Choice Program. You have successfully submitted an application for Jimmy!

To learn more about your **chances of being assigned to the schools on your application**, we strongly encourage you to use the **New Haven Public Schools Application Simulator**.

To use the Application Simulator:

1. Visit our simulator website.
2. Log in using the **same** user name and password used in the NHPS School Choice website.

The **Application Simulator** can help you discover the chances of being assigned to each school in your application and your chances of not being assigned to any school in your application. You can also search, explore, and add new schools.

Please note that the Application Simulator predicts the chances of school assignment **based on data from last year**. The results show the chance of school assignment if you had submitted an application last year. The simulator does not show placement results for the 2020 School Choice Process.

Remember that you can make changes to your child's application until **March 2nd, 2020.**

For more information, please contact us:

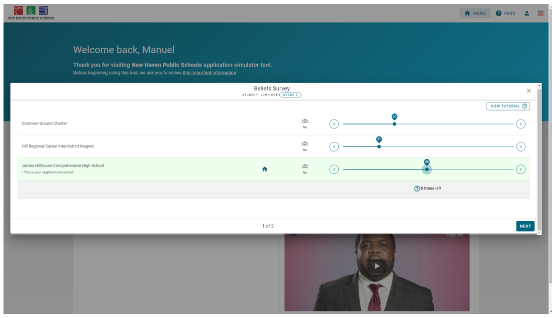
- Email us at simulator@new-haven.k12.ct.us
- Visit the **Choice & Enrollment Office** at 54 Meadow Street, New Haven, CT 06519.
- Contact us by phone: **(475) 220-1430** or **(475) 220-1431**. *Please note phone lines are extremely busy during this time of the year.*

Again, thank you for choosing New Haven Public Schools of Choice!

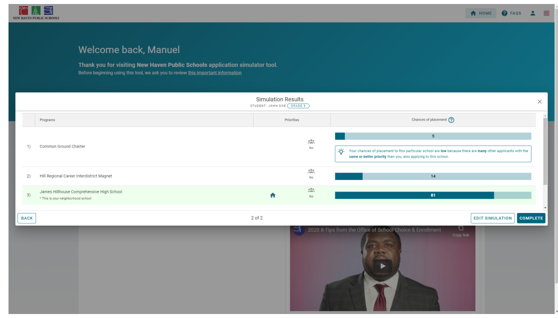
Sincerely,
Choice & Enrollment Office

(b) Non-risky Applicants

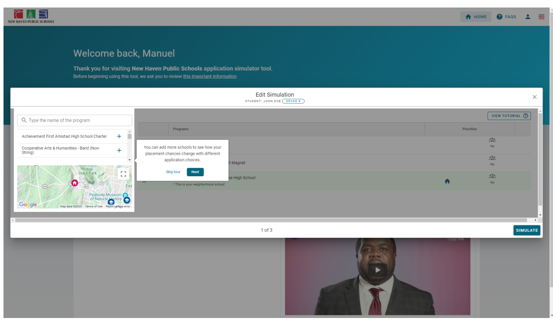
Figure B.31: New Haven Simulator



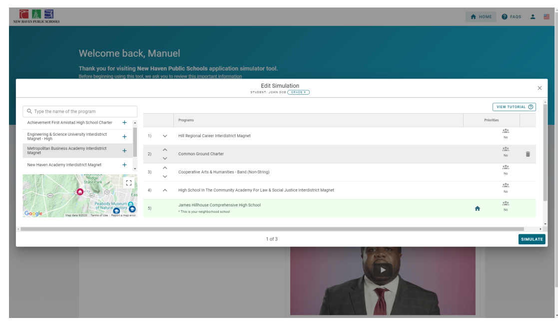
(a) Beliefs Survey Page



(b) Predicted Admission Chances

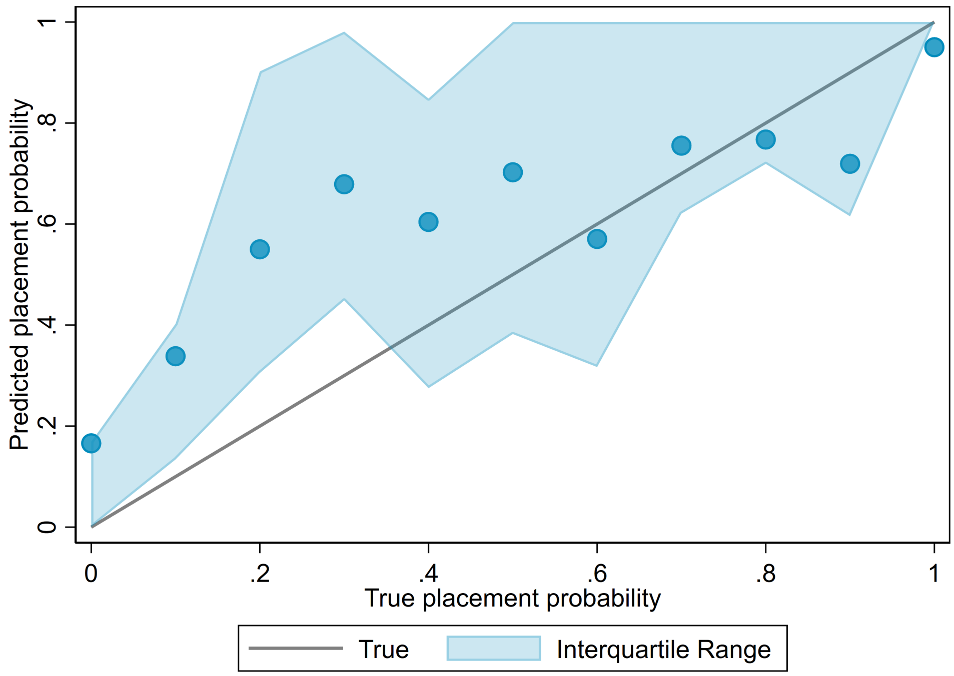


(c) Simulating Portfolio Changes



(d) New Predicted Admission Chances

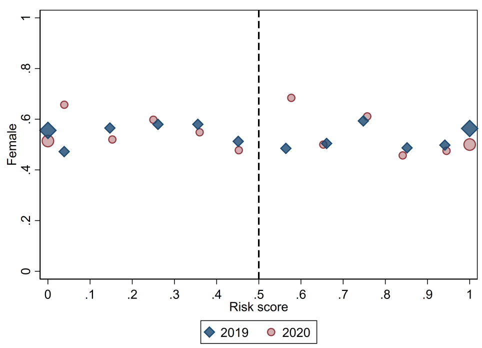
Figure B.32: Observed vs. Predicted Placement Probability in New Haven



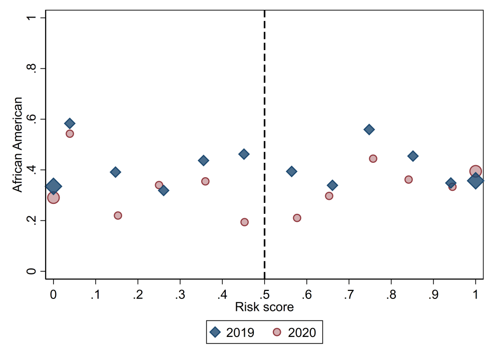
Notes. Distribution of predicted placement probability by value of ex post observed placement probability. For each bin of observed placement, we display the mean and IQR of predicted values. 45-degree line displayed for reference. See section B.11 for details.

Figure B.33: Treatment Balance and First Stage Outcomes

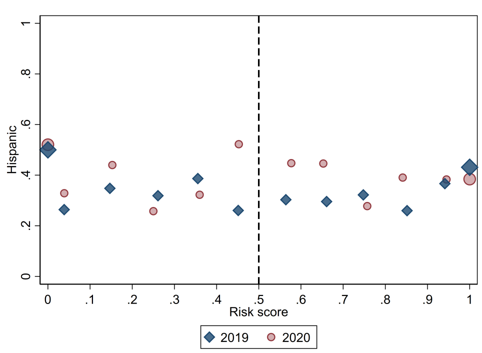
(a) Female



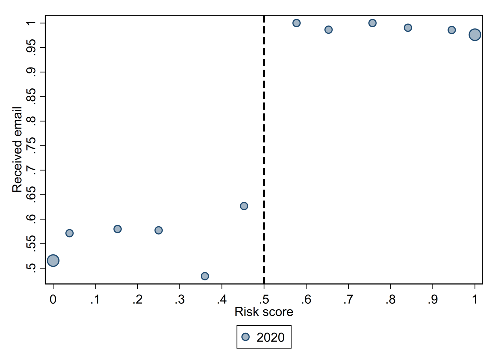
(b) Black



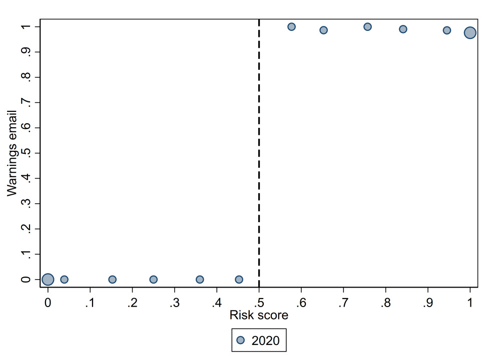
(c) Hispanic



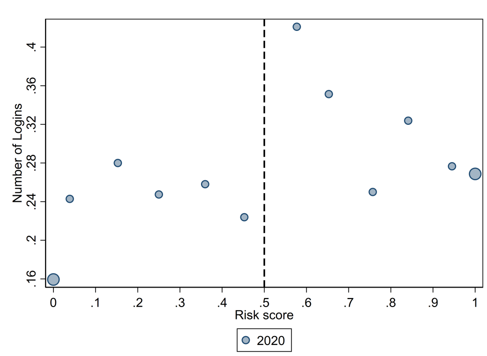
(d) Received Email



(e) Received Warnings Email



(f) Number of Logins



Notes. Balance and first stage effects for warnings intervention in New Haven, CT centralized choice. Figures show predetermined covariates and treatment receipt by risk score as of 7 days prior to application deadline in 2019 and 2020. Points are centered binned means within intervals of width 0.1, except for top- and bottom-most points, which are for students with risk scores of 1 and 0, respectively. Panels A-C display values for both 2020 application cohort and 2019 comparison group. Panels D-F display treatment receipt for 2020 cohort only; no warnings treatment or simulator intervention took place in 2019. See section B.11 for details.

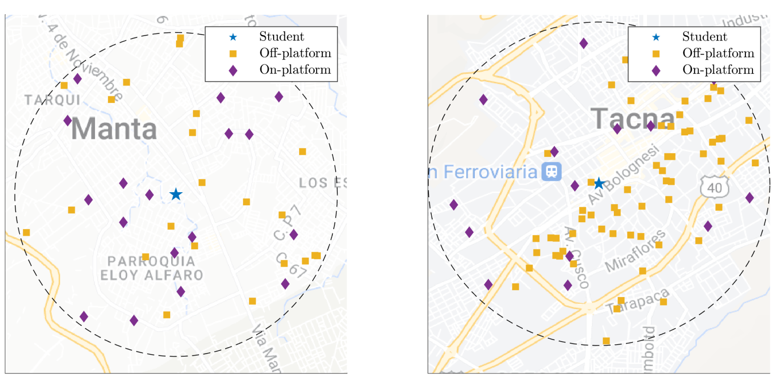
Appendix C

Appendix of Can Information on School Attributes and Placement Probabilities Direct Search and Choice?

C.1 Additional Figures

Figure C.1: Schooling options in a 2km radius from the city center

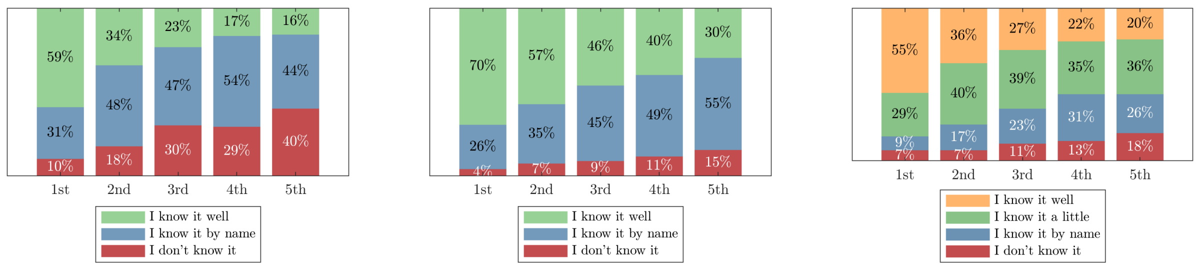
(a) Manta, Ecuador (b) Tacna, Peru



Notes: The star in Panels (a) and (b) reflect the location of the median student, around which the 2 km radius is drawn. The squares and diamonds represent the on-platform and off-platform schooling options within that radius.

Figure C.2: Knowledge About Options in Application

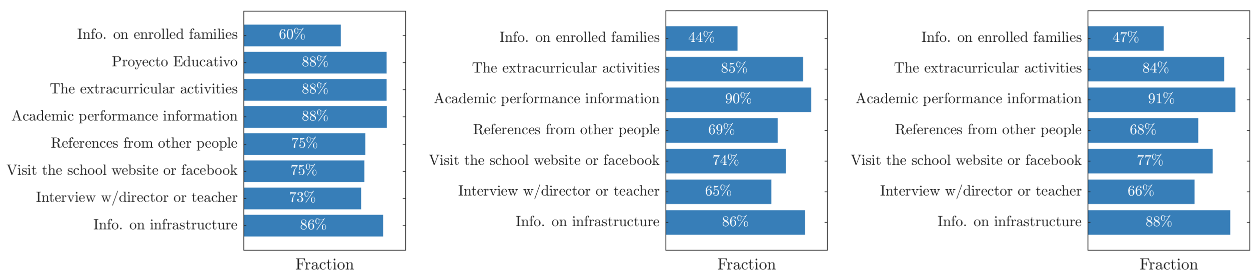
(a) Ecuador (b) Peru 2021 (c) Peru 2022



Notes: Panels (a) to (c) show the level of knowledge of the schools in the application (question 6 in appendix C.4.)

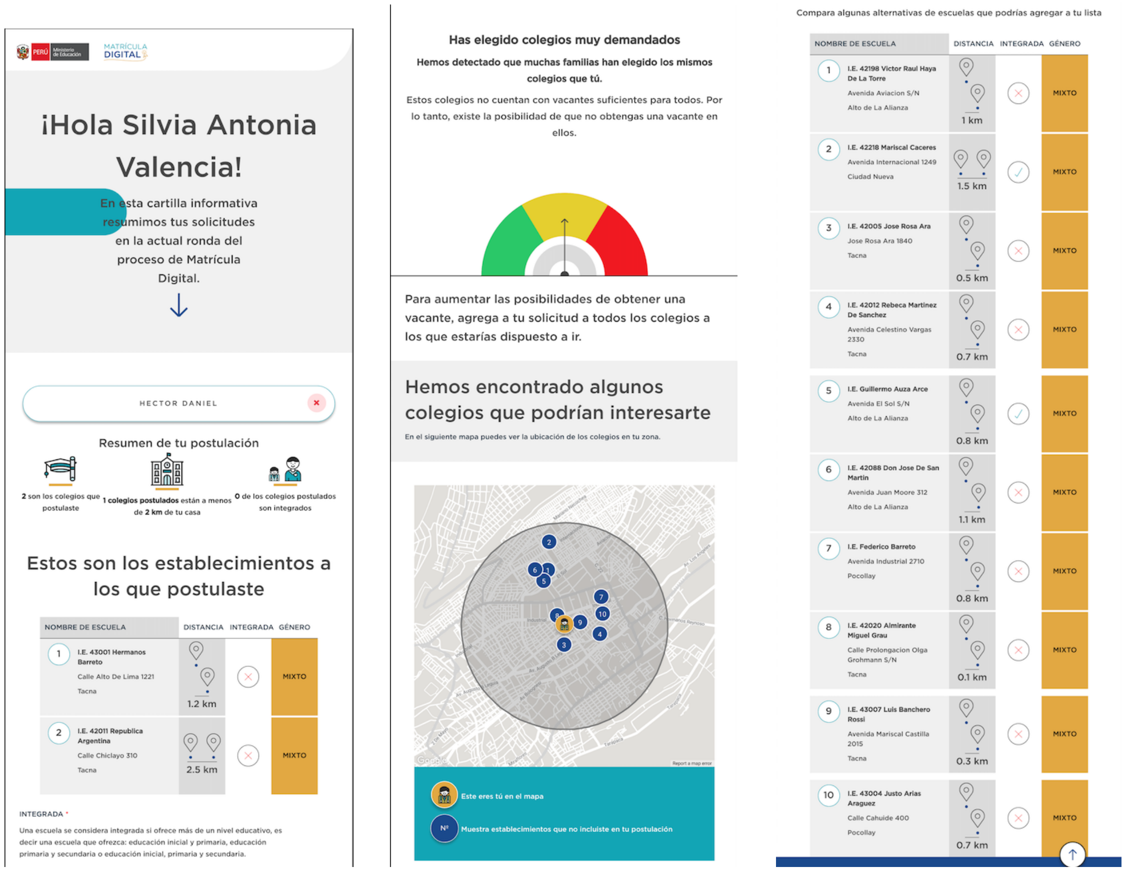
Figure C.3: Necessary Steps for Learning about a School

(a) Ecuador (b) Peru 2021 (c) Peru 2022



Notes: Panels (a) to (c) show the answer to the question of the necessary steps for learning about a school (question 4 in appendix C.4.)

Figure C.4: Report card example of T_2 for Peru 2022



Notes: The panels illustrate a report card sent to an applicant in the 2022 Peru application process who was assigned to T_2 (warning + suggestion list) based on their positive placement risk.

Figure C.5: Information on the Number of Schools Nearby (Ecuador/Peru)

Figure C.6: Treatments for Classifying Schools (Ecuador)

C.2 Additional Tables

Table C.1: Summary of Information Interventions.

	Ecuador 2021			Peru 2021			Peru 2022		
	T_1	T_2	T_{3-2021}	T_1	T_2	T_{3-2021}	T_1	T_2	T_{3-2022}
Target population	Applicants with predicted placement risk greater than 0%			Applicants with predicted placement risk greater than 0%			Applicants with predicted placement risk greater than 30%		
A. Sent by email									
Link to report card	x	x	x	x	x	x			
B. Sent by WhatsApp									
Link to report card	x	x	x	x	x	x	x	x	x
Non-placement warning				x	x	x	x	x	x
C. Included on the report card									
Non-placement warning	x	x	x	x	x	x		x	x
Suggestion list of 10 schools		x			x			x	
Suggestion list of 10 schools w/info on popularity and congestion			x			x			
Suggestion list of 3 schools									x

Notes. This table shows the target populations and different contents of our information intervention depending on the context. "Placement risk" is equivalent to non-placement probability.

Table C.2: RCT Results: Differences Between Treatment 2 and 3.

	(1)	(2)	(3)
Country	Ecuador 2021	Peru 2021	Peru 2022
Intervention difference	$T_{3-2021} - T_2$ info pop/cong (diff.)	$T_{3-2021} - T_2$ info pop/cong (diff.)	$T_{3-2022} - T_2$ less schools (diff.)
A. Choice behavior			
Add any school	-0.018	-0.012	0.045
	(0.023)	(0.035)	(0.041)
Number of schools added	-0.225**	-0.116	0.023
	(0.112)	(0.121)	(0.184)
Add popular	-0.023	-0.076**	0.037
	(0.016)	(0.033)	(0.037)
Add congested	-0.017	-0.018	0.037
	(0.021)	(0.035)	(0.039)
B. Add schools from list			
Add from list (10)	-0.016	-0.028	-0.027
	(0.022)	(0.033)	(0.028)
Add outside list (10)	-0.025	-0.019	
	(0.016)	(0.032)	
Add popular from list (10)	-0.017	-0.026	
	(0.014)	(0.021)	
Add congested from list (10)	-0.001	-0.014	
	(0.017)	(0.031)	
C. Assignment and Enrollment Outcomes			
Placed in list (10)	-0.003	-0.020	
	(0.032)	(0.019)	
Enrolled in list (10)	-0.005	-0.017	
	(0.037)	(0.026)	

Notes. *** p<0.01, ** p<0.05, * p<0.1. This table shows the differential effect of treatments T_3 versus T_2 of Ecuador 2021 (columns 1) and Peru (columns 2 and 3). For Ecuador 2021 and Peru 2021 (columns 1 and 2) the difference between T_{3-2021} and T_2 is that the former included information in popularity and congestion for each school in the list. ("info pop/cong"). For Peru 2022 the difference between T_{3-2022} and T_2 is that the former included 3 schools and the latter 10 schools. The sample considers only applicants that opened the link to the report card.

Table C.3: RCT Results: Effect of Suggestions on Subjective Outcomes.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Country	Ecuador 2021			Peru 2021			Peru 2022		
Intervention	T_1 Warning	T_2 + list (10)	T_3 + list (10) + info pop/cong	T_1 Warning	T_2 + list (10)	T_3 + list (10) + info pop/cong	T_1 Warning	T_2 + list (10)	T_3 + list (3)
	(base)	(diff.)	(diff.)	(base)	(diff.)	(diff.)	(base)	(diff.)	(diff.)
A. Application process									
Was it difficult to search for schools?	0.279	-0.056	-0.003	0.408	-0.155**	-0.182**	0.353	-0.031	0.001
	(0.041)	(0.056)	(0.054)	(0.057)	(0.076)	(0.073)	(0.067)	(0.091)	(0.090)
Evaluation of school info on platform [1 to 20]	17.913	-1.292***	-0.059	13.277	0.944	0.930	14.556	0.401	0.376
	(0.277)	(0.496)	(0.381)	(0.560)	(0.726)	(0.801)	(0.631)	(0.827)	(0.862)
General evaluation of process [1 to 20]	18.396	-0.534	-0.217	13.584	-0.336	0.199	14.143	0.973	-0.198
	(0.239)	(0.417)	(0.351)	(0.542)	(0.756)	(0.774)	(0.658)	(0.837)	(0.905)
Received warning	0.567	0.100	0.057	0.808	0.053	-0.008	0.451	0.216**	0.231**
	(0.044)	(0.061)	(0.059)	(0.045)	(0.060)	(0.063)	(0.070)	(0.093)	(0.091)
Do not plan to apply to private	0.548	0.040	0.045	0.772	0.009	0.001	0.745	-0.078	0.103
	(0.043)	(0.061)	(0.057)	(0.047)	(0.065)	(0.065)	(0.062)	(0.087)	(0.076)
B. Satisfacation									
Satisfaction if placed in 1st [1 to 20]	19.524	-0.260	-0.287	19.224	0.013	-0.306	19.137	-0.239	-0.106
	(0.194)	(0.315)	(0.297)	(0.187)	(0.340)	(0.379)	(0.320)	(0.445)	(0.462)
Satisfaction if placed in last [1 to 20]	13.729	-1.244	0.298	12.882	-0.763	0.142	14.000	-0.610	-1.431
	(0.706)	(1.086)	(0.952)	(0.713)	(0.980)	(0.981)	(0.784)	(1.041)	(1.050)
Satisfaction if no placement [1 to 20]	4.008	-0.438	0.479	2.053	0.816	0.289	3.137	-0.578	-0.953
	(0.449)	(0.579)	(0.604)	(0.276)	(0.559)	(0.469)	(0.602)	(0.754)	(0.695)
C. Knowledge and beliefs									
Proportion of schools from list (10) that know	0.355	-0.074**	-0.044	0.361	0.143**	0.075	0.208	0.022	
	(0.024)	(0.032)	(0.031)	(0.041)	(0.056)	(0.057)	(0.035)	(0.048)	
Proportion of schools from list (3) card that know							0.270		0.079
							(0.051)		(0.074)
Private schools that don't know (out of 5)	2.762	0.152	0.222	2.082	0.062	0.137	2.406	0.234	0.421
	(0.145)	(0.207)	(0.193)	(0.187)	(0.260)	(0.263)	(0.237)	(0.329)	(0.311)
Subjective risk	0.081	-0.010	0.010	0.120	-0.010	0.007	0.107	0.006	0.001
	(0.018)	(0.023)	(0.023)	(0.016)	(0.022)	(0.025)	(0.020)	(0.028)	(0.029)

Notes. *** p<0.01, ** p<0.05, * p<0.1. This table shows the effect of the information intervention on the applicants of Ecuador 2021 (columns 1 to 3) and Peru (columns 4 to 9). Columns 1, 4 and 7 compare the portfolios before and after treatment for all applicants that were assigned to T_1: warning message but no suggestion's list. Columns 2, 5 and 8 show estimates of the differential effect of showing a list of 10 suggested schools in addition to the warning (T_2) compared to only showing the warning (T_1). Columns 3 and 6 show estimates of the differential effect of showing a list of suggested schools with information on pupularity and congestion in addition to the warning (T_{3-2021}) compared to only showing the warning (T_1). Column 9 show estimates of the differential effect of showing a list of 3 suggested schools in addition to the warning (T_{3-2022}) compared to only showing the warning (T_1). "list (10)" is the list of 10 suggestions, while "list (3)" is the list of 3 suggestions. "info pop/cong" refers to the additional information on popularity and congestion showed for each school in the list. The sample considers only applicants that opened the link to the report card.

C.3 Outreach and Treatment Details

The main channel of communication with families was through the messaging app WhatsApp. We used the cellphone numbers reported by the applicants in the registration process.

The messages we sent through the WhatsApp messaging app differed between the two contexts. In Peru, we sent a warning message about the possibility of not being assigned (WhatsApp warning), while in Ecuador we did not. In both settings, we sent reminders to check the link with the report card. An example of the messages we sent is displayed in Table C.4.

Table C.4: Example of WhatsApp Conversation from Peru 2022 - Translation to English

#	Name	Content
1	WhatsApp introduction	Hello [guardian name], we are writing to you since you are registered in the 2022 Digital School Enrollment System. We would like to share information with you regarding your application. Answer "Yes" to review it.
2	Whatsapp disclaimer	The answers you give us in this conversation are confidential and will not affect your application. Our aim is to help you to have more information so that you can submit a good application
3	WhatsApp warning	**We have detected that many families have chosen the same schools as you!** Many families are applying to the same schools as you, so there is a chance that you may not be placed at those schools. To increase your chances of being placed, **add all the schools you would be willing to attend to your application.**
4	WhatsApp with link to report card	In the following link, you will find **important information regarding your application.** [link to report card]
5	WhatsApp reminder	Remember that you can make changes to your application **until December 26 at [link to application platform**] The last application you send will be the valid application. If you change your mind, feel free to reflect this in your application.
6	WhatsApp closing	See you soon [guardian name], have a nice day

Notes. This table shows an example of the messages sent to applicants with an elevated placement risk. Messages 2 to 6 were sent only if the guardian answered positively to the first message (WhatsApp introduction). Applicants without an elevated placement risk received the same messages but no WhatsApp warning (message 2).

Table C.5: Example of WhatsApp Conversation from Peru 2022 - Original Spanish

#	Name	Content
1	WhatsApp introduction	Hola [nombre apoderado], te escribimos dado que estás registrado en el Sistema de Matrícula Escolar Digital 2022. Quisiéramos compartir contigo información respecto a tu postulación. Contesta "Sí" para revisarla.
2	Whatsapp disclaimer	Las respuestas que nos entregues en esta conversación son confidenciales y no afectarán tu postulación. Buscamos ayudarte a que tengas más información para que realices una buena postulación
3	WhatsApp warning	**Hemos detectado que muchas familias han elegido los mismos colegios que tú!** Muchas familias estánn postulando a los mismos colegios que tu, por lo que existe la posibilidad de que no obtengas una vacante en ellos. Para aumentar las posibilidades de obtener una vacante, **agrega a tu postulación todos los colegios a los que estarías dispuesto a ir.**
4	WhatsApp with link to report card	En el siguiente enlace encontrarás **información importante respecto a tu postulación.** [link a cartilla]
5	WhatsApp reminder	Recuerda que puedes hacer cambios a tu postulación **hasta el 26 de diciembre en [link plataforma de postulación**] La última postulación que envíes será la postulación válida. Si cambias de opinión, no dudes en reflejarlo en tu postulación.
6	WhatsApp closing	Hasta pronto[nombre apoderado], que tengas un buen día

Notes. This table shows an example of the messages sent to applicants with an elevated placement risk. Messages 2 to 6 were sent only if the guardian answered positively to the first message (WhatsApp introduction). Applicants without an elevated placement risk received the same messages but no WhatsApp warning (message 3).

C.4 Survey Details

We distributed the survey by WhatsApp, but before families knew their placement results and after the application process was over. The WhatsApp message included a link to the Qualtrics platform. Each survey was personalized with information about the applicant that included their name, the rank-order list, and schools in the neighborhood that were not included in the application.

The translated and original surveys are provided below.

Survey's translation

1. Which score would you give the following aspects of the application process?
   [*Slider 1 to 20*]
   a) Information about schools available on the platform
   b) Information on the online appliacation process (relevant dates, application's steps-, etc).
   c) Ease to use the online application platform
   d) In general, which score would you give the online application process?
2. How did you get information about the school choice process? Select all those that correspond
   [*Select multiple*]
   a) Through the UGEL
   b) Through the Municipality
   c) Through the current school (or the initial)
   d) Through the newspaper or radio
   e) Through social networks (Facebook, Instagram, Twitter, YouTube)
   f) Through friends or family
   g) Through the Minedu website
   h) I did not use any of the above
3. Through which social network?
   [*Select multiple*]
   a) Facebook
   b) Twitter
   c) Instagram

d) Youtube
e) Snapchat
f) Tiktok

4. In the process of creating your school preferences list. Which steps do you consider necessary in order to get to know a school well before adding it?
[*Select multiple*]

    a) The infrastructure
    b) Interview with the director or a teacher
    c) Visit the school website or facebook
    d) Obtain references from other people
    e) Obtain academic performance information
    f) The extracurricular activities that it offers
    g) The set of prioritized values
    h) The Institutional Educational Project (PEI)
    i) Know the families that go to the school

5. Is there any other relevant step for you that we have not included in the previous question?
[*Open text*]

6. How well do you know the schools you chose on the online platform? [One question for each school ranked]
[*Select one*]

    a) I know it well
    b) I know it a little (this option was only available in Perú 2022)
    c) I do not know it
    d) I know it by name

7. We notice that during the process you added schools to your initial list. Did you know these schools the application process began?
[*Select one*]

    a) I knew it well before applying
    b) I knew it by name only before applying
    c) I didn't even know it by name before applying.

8. To convince yourself to add these additional schools, Did you look for more information? [One question for each school that added]
   [*Select one*]

   a) It was not necessary to look for more information
   b) Yes it was necessary to find out more about them

9. You chose the school [first preference] as the first preference for [applicant name]: on a scale from 0 to 100, with what probability do you think you will get a seat in that option?
   [*Slider 0 to 100*]

10. Imagine that he would have selected your second option ([Second Colegio Preference]) as the first preference. On a scale from 0 to 100, with what probability do you think would get a seat in that option?
    [*Slider 0 to 100*]

11. Imagine that you would have selected your third option [Third Preference]) as the first preference. On a scale from 0 to 100, with what probability do you think would get a seat in that option?
    [*Slider 0 to 100*]

12. Some families fail to obtain a seat in any of the options they chose because there are not enough vacancies. Using the same range from 0 to 100, with what probability do you think that [applicant name] will not obtain a seat in any of the [number of schools in ranking] schools in the ranking?
    [*Slider 0 to 100*]

13. Why didn't you add more schools to your application? (select the main reason)
    [*Select one*]

    a) I know the other schools well and I prefer to finish without a vacancy before adding those alternatives
    b) I think I will get a vacancy for sure in one of the schools I chose
    c) It is very difficult to find more schools
    d) There are no more public schools close enough
    e) If I don't get a vacancy I enroll in a private school

14. If you add more schools to your application, do you think any of these two things (or both) would happen?
    [*Yes or No*]

    a) Decreases the overall probability of not being assigned to a school

b) Decrease the possibility of obtaining a seat in my first preferences

15. Next we show you 5 public schools to which you did not apply. How well do you think you know these schools? [One question for each school]
[*Select one*]

    a) I know it well
    b) I know it a little (this option was only available in Perú 2022)
    c) I do not know it
    d) I know it by name

16. Here are 5 private schools. How well do you think you know these schools? [One question for each school]
[*Select one*]

    a) I know it well
    b) I know it a little (this option was only available in Perú 2022)
    c) I do not know it
    d) I know it by name

17. Did you apply or plan to apply to private schools?
[*Select one*]

    a) No
    b) I haven't decided
    c) Yes

18. Imagine that the platform also had private schools, how many private schools you know would have added to your list?
[*Select one*]

    a) 1
    b) 2
    c) 3 or more
    d) I don't know any private school

19. What would be the first private school that would add to your list?
[*Open text*]

20. What would be the second private school that would add to your lis?
[*Open text*]

21. What would be the third private school that would add to your list?
[*Open text*]

22. We present below the list of public schools that you included in the application and the other schools you mentioned. Please order them by reflecting your preferences: above the most preferred and below the least preferred. (Drag schools to modify or confirm order)
[*Rank alternatives*]

23. During the application process, did you receive any recommendation about adding more schools to your list?
[*Select one*]

    a) No
    b) Yes

24. Through which channel did you receive the recommendation to add more schools?(Select all those who apply)
[*Select multiple*]

    a) SMS
    b) WhatsApp
    c) Email
    d) Web page
    e) Phone call
    f) Other

25. If you find out that there is a school that many other families are applying to, but that you have not added it to your list, you would say that:
[*Select one*]

    a) Doesn't tell me anything about the quality of the school
    b) I don't know
    c) It must be a good school.
    d) I would have to know it more, but I think it's good

26. Imagine that you are still looking for schools and find a new one that you like it a lot, even more than your first preference, but it has 100 applicants and 30 vacancies, what would you do?
[*Select one*]

    a) I add it to my list in 1st preference

b) I add it to my list but a preference lower than the 1st
c) I don't add it and I keep looking
d) I don't know

27. If [applicant name] gets a seat in the following schools, from 1 to 20, how satisfied would you be?
[*Slider 1 to 20*]
    a) First preference: [first preference name]
    b) Last preference: [last name preference]
    c) If you don't get a seat at any school

28. Would you like to have had the following information about the schools that you did not have at the time of applying?
[*Select multiple*]
    a) Information about your probability of obtaining a seat
    b) Academic performance
    c) Number of applicants
    d) Seats available
    e) Shift of the school

29. Select the contact channels that you have used to communicate with the MINEDU during the application process
[*Select multiple*]
    a) SMS
    b) Email
    c) WhatsApp
    d) Telephone
    e) In-person
    f) Other

30. Which channel do you prefer?
[*Select one*]
    a) In-person
    b) Email
    c) Other

d) SMS
e) Telephone
f) WhatsApp

31. What steps of the application process were difficult? (You can select more than one option)
[*Select multiple*]

a) Creation of account
b) Filling guardian's personal information
c) Filling student's personal information
d) Search for schools
e) Registering siblings
f) Registration of special eduacation certificate
g) Choice of schools for preferences
h) Postulation type selection
i) Filing the application
j) None

32. Only for registration purposes, what is the highest educational level attained by the mother of the applicant?
[*Select one*]

a) Complete non-university tertiary education
b) Incomplete non-university tertiary education
c) I did not study
d) Postgraduate (master's or doctorate)
e) Complete primary education
f) Incomplete primary education
g) Completed secondary education
h) Incomplete secondary education
i) Complete non-university tertiary education
j) Incomplete non-university tertiary education
k) Complete university tertiary education
l) Incomplete university tertiary education

33. Do you have any other comment, claim or suggestion?
[*Open text*]

Original survey

1. ¿Qué nota le pondría a los siguientes aspectos del proceso de postulación?
   [*Slider 1 to 20*]

   a) Información sobre los colegios disponibles en la plataforma
   b) Información sobre el proceso de Matrícula Digital (fechas relevantes, pasos para postular, etc).
   c) Facilidad para usar la plataforma de Matrícula Digital
   d) En general, ¿qué nota le pondría al proceso de Matrícula Digital a través de la plataforma de matrícula digital?

2. ¿Cómo se informó sobre el proceso de Matrícula Digital?Selecciona todas las que correspondan
   [*Select multiple*]

   a) A través de la UGEL
   b) A través de la Municipalidad
   c) A través del colegio actual (o la inicial)
   d) A través del periódico o radio
   e) A través de redes sociales (Facebook, Instagram, Twitter, Youtube)
   f) A través de amigos o familiares
   g) A través del sitio web del Minedu
   h) No utilicé ninguna de las anteriores

3. ¿A través de qué red social se informó respecto de Matrícula Digital?
   [*Select multiple*]

   a) Facebook
   b) Twitter
   c) Instagram
   d) Youtube
   e) Snapchat
   f) TikTok

4. A la hora de armar su lista de preferencias de colegios en la plataforma de Matrícula Digital ¿Qué pasos considera necesarios para conocer bien un colegio antes de agregarlo?
   [*Select multiple*]

a) Conocer su infraestructura
b) Entrevistarte con el director o algún profesor
c) Visitar la página web o facebook del colegio
d) Obtener referencias de algún conocido
e) Obtener información de rendimiento académico
f) Conocer las actividades extracurriculares que ofrece
g) Conocer el conjunto de valores priorizados
h) Conocer el proyecto educativo institucional (PEI)
i) Conocer sobre las familias que van al colegio

5. ¿Hay algún otro paso relevante para usted que no hayamos incluido en la pregunta anterior?
[*Open text*]

6. ¿Qué tan bien conoce a los colegios que eligió en la plataforma de Matrícula Digital?
[Una pregunta por cada colegios del ranking]
[*Select one*]

a) Lo conozco bien
b) Lo conozco un poco
c) No lo conozco
d) Solo de nombre

7. Notamos que durante el proceso de Matrícula Digital agregó colegios a su listado inicial ¿Conocía estos colegios desde antes de que comenzara el proceso de matrícula?
[*Select one*]

a) Lo conocía bien de antes de postular
b) Lo conocía sólo de nombre antes de postular
c) No lo conocía ni de nombre antes de postular

8. ¿Para convencerse a agregar estos colegios adicionales tuvo que buscar más información? [Una preguna por cada colegio que agregó]
[*Select one*]

a) No fue necesario buscar más información
b) Sí fue necesario averiguar más de ellos

9. Usted eligió al colegio [colegio primera preferencia] en primera preferencia para [nombre postulante]: En una escala del 0 a 100, ¿con qué probabilidad o seguridad cree que va a obtener una vacante en esa opción?
   [*Slider 0 to 100*]

10. Imagine que hubiese puesto su segunda opción ([colegio segunda preferencia]) en su primera preferencia: En una escala del 0 a 100, ¿con qué probabilidad o seguridad cree que obtendría una vacante en esa opción?
    [*Slider 0 to 100*]

11. Imagine que hubiese puesto su tercera opción [colegio tercera preferencia]) en su primera preferencia: En una escala del 0 a 100, ¿con qué probabilidad o seguridad cree que obtendría una vacante en esa opción?
    [*Slider 0 to 100*]

12. Algunas familias no logran obtener una vacante en ninguna de las opciones que eligieron debido a que no hay vacantes suficientes. Usando el mismo rango de 0 a 100, ¿con qué probabilidad o seguridad cree que [nombre postulante] NO va a obtener una vacante en ninguno de los [numero de colegios en ranking] colegios a los que postuló?
    [*Slider 0 to 100*]

13. ¿Por qué no agregó más colegios a su postulación?(Marque la razón principal)
    [*Select one*]

    a) Conozco bien los otros colegios y prefiero terminar sin vacante antes de agregar esas alternativas
    b) Creo que voy a obtener una vacante con toda seguridad en alguno de los colegios que elegí
    c) Es muy difícil encontrar más colegios
    d) No hay más colegios públicos lo suficientemente cerca
    e) Si no obtengo una vacante me matriculo en un colegio privado

14. Disminuye la posibilidad de quedarme sin vacante
    [*Select one*]

    a) No
    b) Sí

15. Disminuya la posibilidad de obtener una vacante en mis primeras preferencias
    [*Select one*]

    a) No
    b) Sí

16. A continuación le mostramos 5 colegios públicos a los que no postuló. ¿Qué tan bien cree que conoce a estos colegios? [Una pregunta por cada colegio que le preguntamos] [*Select one*]

    a) Lo conozco bien
    b) Lo conozco un poco
    c) No lo conozco
    d) Solo de nombre

17. A continuación le mostramos 5 colegios privados. ¿Qué tan bien cree que conoce a estos colegios? [Una pregunta por cada colegio que le preguntamos] [*Select one*]

    a) Lo conozco bien
    b) Lo conozco un poco
    c) No lo conozco
    d) Solo de nombre

18. ¿Postuló o tiene pensado postular a colegios privados? [*Select one*]

    a) No
    b) No lo he decidido
    c) Sí

19. Imagine que la plataforma tuviera también colegios privados, ¿cuántos colegios privados que conoce hubiera agregado a su lista? [*Select one*]

    a) 1
    b) 2
    c) 3 o más
    d) No conozco ningún colegio privado

20. ¿Cuál sería el primer colegio privado que agregaría a su lista en la plataforma de Matrícula Digital? [*Open text*]

21. ¿Cuál sería el segundo colegio privado que agregaría a su lista en la plataforma de Matrícula Digital? [*Open text*]

22. ¿Cuál sería el tercer colegio privado que agregaría a su lista en la plataforma de Matrícula Digital?
[*Open text*]

23. Le presentamos a continuación la lista de colegios públicos que registró en su postulación y los privados que nos mencionó. Por favor ordénelos reflejando su preferencia: arriba el más preferido y abajo el menos preferido.(Arrastre los colegios para modificar o confirmar el orden)
[*Rank alternatives*]

24. Durante el proceso de postulación, ¿recibió alguna recomendación sobre agregar más colegios a su lista por parte del Minedu?
[*Select one*]

    a) No
    b) Sí

25. ¿A través de qué medio recibió la recomendación de agregar más colegios?(Seleccione todos los que apliquen) - Selected Choice
[*Select multiple*]

    a) SMS
    b) Whatsapp
    c) Correo Electrónico
    d) Pagina web
    e) Llamada telefónica
    f) Otro

26. Si se enterara que hay un colegio al que muchas otras familias están postulando, pero que usted no lo ha agregado a su lista, diría que:
[*Select one*]

    a) No me dice nada sobre la calidad del colegio
    b) No sé
    c) Seguramente es un buen colegio
    d) Tendría que conocerlo más, pero creo que es bueno

27. Imagine que sigue buscando colegios y encuentra uno nuevo que le gusta mucho, incluso más que su primera preferencia, pero tiene 100 postulantes y 30 vacantes, ¿qué haría?
[*Select one*]

    a) Lo agrego a mi lista en 1ra preferencia

b) Lo agrego a mi lista pero una preferencia menor a la 1ra
c) No lo agrego y sigo buscando
d) No sé

28. Si [nombre postulante] obtiene una vacante en los siguientes colegios, del 1 al 20, ¿qué tan satisfecho(a) estaría?
[*Slider 1 to 20*]
    a) Primera preferencia: [nombre primera preferencia]
    b) Última Preferencia: [nombre última preferencia]
    c) Si no obtiene una vacante en ningún colegio

29. ¿Le gustaría haber tenido la siguiente información sobre los colegios que NO tuvo al momento de postular?
[*Select multiple*]
    a) Información sobre tu probabilidad de obtener una vacante
    b) Rendimiento académico
    c) Cantidad de postulantes
    d) Vacantes disponibles
    e) Turno

30. Marque los medios de contacto ha utilizado para comunicarse con el Minedu durante el proceso de postulación: - Selected Choice
[*Select multiple*]
    a) SMS
    b) Correo electrónico
    c) Whatsapp
    d) Teléfono
    e) Atención presencial
    f) Otro

31. Selected Choice
[*Select one*]
    a) Atención presencial
    b) Correo electrónico
    c) Otro

d) SMS
e) Teléfono
f) Whatsapp

32. ¿Qué pasos del proceso de postulación le resultaron difíciles de realizar? (Puede marcar más de una opción)
[*Select multiple*]

a) Creación de cuenta
b) Registro de datos de apoderado
c) Registro de datos del postulante
d) Búsqueda de colegios
e) Registro de datos de hermano
f) Registro de NEE
g) Elección de colegios para lista de preferencias
h) Selección de tipo de postulación
i) Envío de ficha de postulación
j) Ninguno

33. Solo con fines de registro, ¿hasta qué nivel educativo llegó la madre (o apoderada) del postulante?
[*Select one*]

a) Educación ocupacional completa
b) Educación ocupacional incompleta
c) No estudió
d) Posgrado (maestría o doctorado)
e) Primaria Completa
f) Primaria Incompleta
g) Secundaria Completa
h) Secundaria Incompleta
i) Superior no universitaria completa
j) Superior no universitaria incompleta
k) Superior universitaria completa
l) Superior universitaria incompleta

34. ¿Tienes algún otro comentario, reclamo o sugerencia que nos quieras hacer?
[*Open text*]

Bibliography

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

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

Abdulkadiroğlu, A., N. Agarwal, and P. A. Pathak (2017). The welfare effects of coordinated assignment: Evidence from the new york city high school match. *American Economic Review 107*(12), 3635–89.

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

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

Abdulkadiroğlu, A., Y.-K. Che, and Y. Yasuda (2011). Resolving conflicting preferences in school choice: The "boston mechanism" reconsidered. *American Economic Review 101*(1), 399–410.

Abdulkadiroğlu, A., P. A. Pathak, and A. E. Roth (2005). The new york city high school match. *American Economic Review 95*(2), 364–367.

Abdulkadiroğlu, A., P. A. Pathak, A. E. Roth, and T. Sönmez (2005). The boston public school match. *American Economic Review 95*(2), 368–371.

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

Abdulkadiroğlu, A., P. A. Pathak, and C. R. Walters (2018). Free to choose: Can school choice reduce student achievement? *American Economic Journal: Applied Economics 10*(1), 175–206.

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

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

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

Agarwal, N. and P. Somaini (2018). Demand Analysis Using Strategic Reports: An Application to a School Choice Mechanism. *Econometrica 86*(2), 391–444.

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

Agrawal, A., J. Gans, and A. Goldfarb (2018). *Prediction Machines: The Simple Economics of Artificial Intelligence*. Harvard Business Press.

Ainsworth, R., R. Dehejia, C. Pop-Eleches, and M. Urquiola (2023, April). 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. and M. Sidibe (2020). School Choice Under Imperfect Information. *SSRN Electronic Journal*.

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

Akbarpour, M., C. Neilson, A. Kapor, W. van Dijk, and S. Zimmerman (2020). Centralized school choice with unequal outside options. *Industrial Relations Section Working Paper*.

Allende, C. (2019). Competition Under Social Interactions and the Design of Education Policies. Job Market Paper.

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

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

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association 91*(434), 444–455.

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

Arteaga, F., G. Elacqua, C. Méndez, T. Krussig, and C. Neilson (2022). Can Information on School Attributes and Placement Probabilities Direct Search and Choice? Evidence from Choice Platforms in Ecuador and Peru. Working Paper.

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

Ashlagi, I. and A. Nikzad (2020). What matters in school choice tie-breaking? how competition guides design. *Journal of Economic Theory 190*, 105–120.

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

Balarin, M., C. Fontdevila, P. Marius, and M. F. Rodríguez (2019, October). Educating on a budget: The subsistence model of low-fee private schooling in Peru. *Education Policy Analysis Archives 27*, 132–132.

Barahona, N., S. Otero, and C. Dobbin (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., F. Molinari, and M. Thirkettle (2021, June). Discrete Choice under Risk with Limited Consideration. *American Economic Review 111*(6), 1972–2006.

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

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

Bobba, M. and V. Frisancho (2020). Self-perceptions about academic achievement: Evidence from mexico city. *Journal of Econometrics*.

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

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

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica 82*(6), 2295–2326.

Calsamiglia, C., C. Fu, and M. Güell (2020, February). Structural Estimation of a Model of School Choices: The Boston Mechanism versus Its Alternatives. *Journal of Political Economy 128*(2), 642–680.

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

Chen, Y. and Y. He (2020). Information acquisition and provision in school choice: an experimental study. Technical report.

Chen, Y. and Y. He (2021). Information acquisition and provision in school choice: a theoretical investigation. *Economic Theory*, 1–35.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014, September). Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *American Economic Review 104*(9), 2593–2632.

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

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

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

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

DellaVigna, S. and E. Linos (Forthcoming). Rcts to scale: Comprehensive evidence from two nudge units. *Econometrica.*

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

Doraszelski, U., K. Seim, M. Sinkinson, and P. Wang (2017). Ownership concentration and strategic supply reduction. Technical report, National Bureau of Economic Research.

Dynarski, S., C. Libassi, K. Michelmore, and S. Owen (2021). Closing the gap: The effect of reducing complexity and uncertainty in college pricing on the choices of low-income students. *American Economic Review 111*(6), 1721–56.

Elacqua, G., I. Jacas, T. Krussig, C. Méndez, and C. Neilson (2022). What they want or where they live: A welfare comparison of coordinated assignments focused on location vs choice in the case of Ecuador. Technical report.

Elacqua, G., I. Jacas, T. Krussig, C. Méndez, C. Neilson, A. Roman, and S. Soares (2022). Centralized Student Assignment Systems: Implementation Guide. Technical report, Inter-American Development Bank.

Enke, B. and F. Zimmermann (2019). Correlation neglect in belief formation. *The Review of Economic Studies 86*(1), 313–332.

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

Fernandes, D., J. G. Lynch, Jr., and R. G. Netemeyer (2014). Financial literacy, financial education, and downstream financial behaviors. *Management Science 60*(8), 1861–1883.

Gabaix, X. (2019). Behavioral inattention. In *Handbook of Behavioral Economics: Applications and Foundations 1*, Volume 2, pp. 261–343. Elsevier.

Gobierno de Chile Ministerio de Educación (2017, September). *El primer gran debate de la Reforma Educacional: Ley de Inclusión Escolar.*

Grenet, J., Y. He, and D. Kübler (2021). Decentralizing centralized matching markets: Implications from early offers in university admissions. *arXiv preprint arXiv:2107.01532*.

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

Gruber, J., B. Handel, S. Kina, and J. Kolstad (2020, April). Managing Intelligence: Skilled Experts and AI in Markets for Complex Products. Technical Report w27038, National Bureau of Economic Research, Cambridge, MA.

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

Gurantz, O., J. Howell, M. Hurwitz, C. Larson, M. Pender, and B. White (2021). A national-level informational experiment to promote enrollment in selective colleges. *Journal of Policy Analysis and Management 40*(2), 453–479.

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

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

Hakimov, R., D. Kübler, S. Pan, et al. (2021). Costly information acquisition in centralized matching markets. Technical report, CRC TRR 190 Rationality and Competition.

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

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

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

Hsieh, C.-T. and M. Urquiola (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., J. Leshno, I. Lo, and B. Lucier (2020). Information Acquisition in Matching Markets: The Role of Price Discovery. SSRN Working Paper.

Jin, G. Z. and P. Leslie (2003). The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards. *The Quarterly Journal of Economics 118*(2), 409–451.

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

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

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

Lavecchia, A. M., H. Liu, and P. Oreopoulos (2016). Behavioral economics of education: Progress and possibilities. In *Handbook of the Economics of Education*, Volume 5, pp. 1–74. Elsevier.

Madrian, B. C. (2014). Applying insights from behavioral economics to policy design. *Annu. Rev. Econ. 6*(1), 663–688.

Mani, A., S. Mullainathan, E. Shafir, and J. Zhao (2013). Poverty impedes cognitive function. *Science 341*(6149), 976–980.

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

Manzini, P. and M. Mariotto (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 P. Zarembka (Ed.), *Frontiers in Econometrics*, New York, pp. 105–142. Academic Press.

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

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

Milgrom, P. and I. Segal (2020). Clock auctions and radio spectrum reallocation. *Journal of Political Economy 128*(1), 1–31.

Narita, Y. (2018, June). 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.

Neilson, C. (2021). Targeted vouchers, competition among schools, and the academic achievement of poor students.

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

Pathak, P. A. and T. Sönmez (2013, February). School Admissions Reform in Chicago and England: Comparing Mechanisms by Their Vulnerability to Manipulation. *American Economic Review 103*(1), 80–106.

Rees-Jones, A., R. Shorrer, and C. J. Tergiman (2020). Correlation neglect in student-to-school matching. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pp. 467–468.

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

Son, S. J. (2020). Distributional impacts of centralized school choice. *Yale University Working Paper*.

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

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

Thaler, R. H. and W. Tucker (2013). Smarter information, smarter consumers. *Harvard Business Review 91*(1), 44–54.

Tincani, M. M., F. Kosse, and E. Miglino (2021). Subjective beliefs and inclusion policies: Evidence from college admissions.

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

Tversky, A. and D. Kahneman (1974, September). Judgment under Uncertainty: Heuristics and Biases. *Science 185*(4157), 1124–1131.

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

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

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