IDB WORKING PAPER SERIES Nº IDB-WP-01420

The Welfare Effects of including Household Preferences in School Assignment Systems: Evidence from Ecuador

Gregory Elacqua
Isabel Jacas
Thomas Krussig
Carolina Méndez
Christopher A. Neilson

Inter-American Development Bank
Education Division

November 2022

The Welfare Effects of including Household Preferences in School Assignment Systems: Evidence from Ecuador

Gregory Elacqua
Isabel Jacas
Thomas Krussig
Carolina Méndez
Christopher A. Neilson

November 2022

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library
The welfare effects of including household preferences in school assignment systems: evidence from Ecuador / Gregory Elacqua, Isabel Jacas, Thomas Krussig, Carolina Méndez, Christopher A. Neilson.
p. cm. — (IDB Working Paper Series; 1420)
Includes bibliographic references.
1. School enrollment-Ecuador. 2. School choice-Ecuador. 3. Education-Effect of technological innovations on-Ecuador. 4. Social welfare-Ecuador. 5. Household surveys-Ecuador. I. Elacqua, Gregory M., 1972- II. Jacas, Isabel. III. Krussig, Thomas. IV. Méndez, Carolina. V. Neilson, Christopher. VI. Inter-American Development Bank. Education Division. VII. Series.
IDB-WP-1420

JEL Codes: I20, I21, I22
Keywords: Mechanism design, centralized student assignment, school choice, Ecuador

http://www.iadb.org

The Welfare Effects of including Household Preferences in School Assignment Systems: Evidence from Ecuador

Gregory Elacqua, Inter-American Development Bank
Isabel Jacas, ConsiliumBots
Thomas Krussig, UC Berkeley
Carolina Méndez, Inter-American Development Bank
Christopher A. Neilson, Yale University

November 17, 2022

Abstract

We study the welfare produced by a coordinated school assignment system that is based exclusively on minimizing distance to schools, comparing the matches it produces to a system that includes household preferences using a deferred acceptance algorithm. We leverage administrative data and a mechanism change implemented in the city of Manta, Ecuador in 2021 to estimate household preferences and show that considering applicant preferences produces large welfare gains. Our counterfactual exercises show that differences across alternative assignment mechanisms are small. Survey data on household beliefs and satisfaction support these conclusions. The evidence indicates that coordinated school choice and assignment systems can have large welfare effects in developing country contexts.

Key words: Mechanism design, centralized student assignment, school choice, Ecuador

1 Introduction

In this paper, we study the welfare effects of a significant policy change to the coordinated school assignment system in Manta, Ecuador. The assignment system in place before the policy change aimed to minimize the (linear) travel distance between homes and schools. Implementing this system proved challenging due to the considerable effort required to geo-reference all students and ensure that the assignment process results were consistent with actual transportation options as well as the existence of hills, rivers, etc. Given the costs and difficulties of reviewing linear distance-based assignments to correct for geographic accidents, the Ecuadorian Ministry of Education piloted an alternative assignment system in partnership with the Inter-American Development Bank (IADB) and the NGO ConsiliumBots in which applicants' preferences were the main driver. This was motivated by findings in the school choice literature on the benefits of coordinated assignment systems that take family preferences into account (Abdulkadiroğlu et al. (2017)). The new policy resulted in a system that followed standard best practices, including the use of the deferred acceptance algorithm, unlimited ranked ordered lists, and information provision systems (Abdulkadiroğlu & Sönmez, 2003; Pathak, 2011, 2017; Arteaga et al., 2021).

To compare the assignment alternatives, we take advantage of the fact that the system implemented in Manta elicited the true preferences and locations of all participating applicants. This allows us to compare the assignments made by the new centralized choice and assignment system (CCAS) with the simulated assignments of the prior alternative. Our methodology consists of using a counterfactual strategy, replicating the rules of the previous process, and simulating assignments with different lotteries following Abdulkadiroğlu et al. (2017).

Our main finding is that implementing a coordinated mechanism that incorporates the preferences of applicants has large welfare benefits. When compared to the previous system, using the deferred acceptance (DA) algorithm increases the portion of applicants assigned to any of their chosen schools from 49.96% to 78.44%, while the percentage of applicants assigned to their first choice increases from 42.42% to 69.76%. The main trade-off of implementing the DA alternative is that average linear distance to the school increases by 0.29km. Indeed, the DA algorithm results in distances between applicants' homes and schools that are 0.683km higher for Pre-School 1 applicants and 0.354km greater for those entering Pre-School 2, while the increase is 0.012km for Primary 1. Changes in distance to school are one indication of welfare effects, with higher distances being less desirable. The difference is smaller in the first year of primary school due to much greater congestion (many seats have already been taken by applicants enrolled during pre-school), and thus many primary school applicants are not assigned to any of their chosen schools under either alternative. If the sample is restricted to applicants who receive a different assignment under the distance and DA mechanisms (i.e., applicants who improve or worsen their utility when the mechanism is changed), the differences for each level increase to 1.481km for Pre-School 1, 0.548km for Pre-School 2 and 0.015km for Primary 1.

We focus our analysis on estimated welfare and on the share of applicants being assigned to more or less preferred alternatives based on their reported preferences as detailed below. School quality is not considered because: i) we do not aim to study whether families in Ecuador prefer higher-quality schools, but rather to assess the welfare consequences of the assignment system as it relates to applicants' valuation of different schools, and ii) we cannot (at least directly) observe school quality.[1]

These results contribute to a better understanding of the advantages of coordinated school choice and assignment systems. While several studies demonstrate the welfare benefits of using one mechanism compared to others, few have directly estimated the welfare benefits of a coordinated assignment system that takes household preferences into account. The most closely related paper is Abdulkadiroğlu et al. (2017), which examines the welfare effects of switching the previously uncoordinated New York City assignment system to a coordinated approach that incorporates family preferences. The authors find that most of the welfare gains are obtained from the coordination using the standard deferred acceptance algorithm, with only marginal gains when implementing alternatives. We find similar results in the context of Ecuador, though here we are comparing the DA algorithm to a coordinated one that centers on the distance of the household to the school.

2 Context and Algorithm Descriptions

We study school assignment in the coastal region of Manta, Ecuador. Specifically, we concentrate on the urban areas within and around the city of Manta,[2] including the geographic units ("*cantones*") of Manta, Montecristi, and Jaramijó.

Manta was selected as the result of a process that aimed to find a small but representative city to scale up the school assignment policy.[3] The selection process took into account students in the urban area, school coverage, distribution of school types (mainly public and private), as well as city size. Ultimately, Manta was chosen for its relative similarity to the alternatives. Table 5 of Appendix B compares the main characteristics of Manta and Guayaquil, another coastal city and the country's largest, using data from the 2010 Census and school transfer requests in the 2019-2020 school year.

The Ecuadorian educational system is organized into three levels: Pre-school (Educación Inicial), Primary School (Educación General Básica) and Secondary School (Bachillerato). In this paper, we focus on school assignments at the "entry level," a designation that encompasses enrollment in Pre-school 1, Pre-school 2, and Primary 1.[4,5]

[1] The Ecuadorian government does not currently apply census-based student learning assessments in primary grades. We also do not study the impact of the system on other measures of interest, such as educational segregation, as we lack socioeconomic data for participating applicants.

[2] The educational system in Ecuador is split into two educational regimes nationally, one for the coastal region and another for the country's interior. The academic year in the coastal regime (where Manta is located) begins in May and ends in January, while it begins in September and ends in June in the interior.

[3] There was a change in government in Ecuador in 2021, and the new administration recently decided to scale up the system in coastal districts beginning in 2023. Because of the COVID pandemic and the fact that the distance-centric alternative required several in-person interactions during the process, the Ministry is currently using a First-Come, First-Serve digital system.

[4] Pre-school is divided into two grades, called "*Inicial*" one and two. Primary school is divided into four levels. The first, which we call Primary 1, is for five-year-old children, while the other three levels cover children from six to eight years old, nine to eleven years old, and twelve to fourteen years old.

[5] The education system also offers different tracks: the regular track accounts for 98% of enrollment, with the remaining 2% distributed between schooling for students with special educational needs, artistic education, and

There are three types of schools in Ecuador: public ("*fiscal*" and municipal), "*fiscomisional*," and private. Public institutions are funded by the government, "*fiscomisionales*" receive mixed funding from the state and families, and private schools are fully funded by families. Nationwide, free public schools account for 73.8% of enrollment, split between the "*fiscales*" and municipal schools (the latter of which represent only 0.8% of total enrollment). *Fiscomisional* schools receive 6% of all enrollments, while private schools take the remaining 20%. In Manta, free public schools account for 66% of enrollment at the entry-level grades, while *fiscomisional* and private alternatives account for 4% and 30% respectively.

This pilot only included free public schools, meaning that private schools represent an outside option for families that is not explicitly incorporated into our model. This is important for our welfare comparisons, which might be overestimated considering that families can switch from a less desirable free public school to a private alternative. As such, our welfare comparisons should be interpreted as the difference between the utility offered to families by the free public system. It is nonetheless also relevant to note that, at least for the applicants that participated in the pilot, there seems to be only limited overlap between free public schools and private alternatives. This is illustrated by a survey conducted after the application period but before results were distributed.[6] Only 1% of respondents stated that their reason for not including more alternatives was because of the alternative of enrolling their child in a school outside the public system. This is a critical observation because, as we explain in Section 3, many applicants submitted only one or two preferences.[7] While the short lists may have been due to applicants' preferences for outside options, the survey results suggest that this was not a frequent consideration for participating families.

2.1 Distance-Centric Algorithm

Before the COVID-19 pandemic, the school assignment system in Ecuador was based mainly on the applicant's location, reported through the code on the family's electricity account (CUEN). In addition to the linear distance criterion, a prioritization criterion was also used to determine the order in which applicants were processed (being processed first was preferable). This prioritization criterion was randomly assigned to students applying to be enrolled in the system. In the school assignment literature, such a mechanism has been termed a "random serial dictatorship" (Abdulkadiroğlu & Sönmez, 1998).

The assignment system was part of a broader enrollment process comprising six phases, as described in Appendix C. The Assignment Phase was the third of these phases, and also operated in stages. In the first stage of the Assignment Phase, different types of enrollment in the system were identified, depending on whether students preferred to attend a regular program or a rural, bilingual, or special education program. Also, where possible, applicants with siblings already in the system were assigned to the same school. Students registering for non-regular programs (e.g., bilingual or special education) and those with already-schooled

adult schooling.

[6]The objectives of this survey were to gather information about parents' overall satisfaction level with the system, information sources they used to apply for schools, awareness of the school supply, among other aspects. The survey was completed by 1,484 parents.

[7]All applicants were ultimately assigned, though some to a school outside their reported list of preferences. In such cases, the assigned school was the closest possible alternative, as explained in subsection 2.1.

siblings were processed before the other applicants.

Once these groups were assigned via a process that was carried out directly at the district headquarters, the rest of the students were assigned using the distance-centric (DC) algorithm, which the Ministry called the "mathematical model." Legal guardians could complete an individual registration (of a single applicant) or one for a "group of siblings." While this latter option suggests that the system prioritized assigning groups of siblings to the same school over distance-based considerations, this was not confirmed by the Ministry experts with whom we interacted.

The processing of regular assignments was carried out as follows:

- At each level, random numbers were given to all applicants. These random numbers correspond to the prioritization criteria mentioned above and determined the processing order.
- Following this order, applicants were assigned to the closest school (linear distance) with vacancies, in an iterative process that used increasing distance radii from the applicant's home.[8]

This procedure can be conceptualized as an application of the "serial dictatorship" mechanism, in the sense that applicants select schools one after the other. Given that the order of choice has a random component, such assignment models have been termed "random serial dictatorships." The enrollment of groups of siblings can thus be considered a priority, since the system will try to assign these groups to the same schools over other individual applications.

As explained above, an applicant's home address was based on the legal guardians' electricity bill. Using the latter to identify family location has proven highly effective, but may also incentivize families to procure (and even buy) electricity bills closer to their schools of interest. Moreover, there are still areas where households do not have electricity meters. These facts were reported in a series of interviews carried out by the IADB in Quito and Guayaquil, where families and officials recounted different factors affecting the registration processes.[9] Given that we do not have precise estimates of location misreporting rates, we conduct a sensitivity analysis in Section 4 and simulate assignments under different levels of misreporting.

The Ecuadorian government's concern with minimizing the distance to school arises from public policy considerations, and not because this aspect affects other dimensions such as, for example, public expenditure on free busing to schools. The latter consideration is nevertheless relevant in other contexts (e.g., many US cities), meaning that analyses comparing assignment mechanisms in similar cases should consider inclusion of these budget factors.

[8]The schools available were evaluated at radii of 100m, 200m, 300m, and 500m, and then at increments of 250m up to 3.5 kilometers.

[9]For example, district officials commented: (1) *"In District 24, Durán, Guayaquil, families lend their electricity bills to each other so they can all have access to the education system. We estimate that more than 60% of families in this district do not use their own electricity bill, so they do not register their real geolocation."* (2) *"In District 8, Monte Sinaí, Guayaquil, families maintain that there are "illegal invasions" of other families in areas where popular schools are located, using electricity bills from that area to get a seat in these schools."*

2.2 Deferred Acceptance (DA) Mechanism

The pilot used the deferred acceptance mechanism (Gale & Shapley, 1962), following the best practices in school choice mechanism design (Pathak, 2011; Correa et al., 2019). The specification of the assignment algorithm included static and dynamic sibling priorities, family linking, and a multiple tie-breaking rule.[10]

The static and dynamic sibling priorities indicate that an applicant will be prioritized for assignment to a school/program if their sibling is already assigned to the school (static). If the applicant is applying at the same time with another sibling, and one of them is assigned to a school,[11] the applicant that has not been assigned yet will receive priority for being assigned to that same school (dynamic). The dynamic sibling priority is lower than the static sibling priority because the latter is already defined (the sibling is attending the school), while the former will depend on the answer from the applicant after the assignment.

The family linking feature consists of trying to assign all siblings applying together to the same schools. Following a descending order, where older applicants are assigned first, if an older sibling is assigned to school A, the applications of the younger siblings will be modified to put school A as the first-ranked school to improve the probability of being assigned together. Finally, a multiple tie-breaking rule gives each applicant a different lottery number for each school to which they apply. Lottery numbers are used to break ties within priority groups when a school receives more applications than spaces available.

3 Data

The data used in this paper come from the centralized choice and assignment system (CCAS) pilot web page created in 2021 in the region of Manta, Ecuador.[12] The first data set comprises the supply of vacancies for all schools and programs offered in the pilot, where an educational program consists of a combination of grade and school. The pilot was implemented for all students entering Pre-School 1, Pre-School 2, and the first year of primary school (i.e., ages 3 to 5) for the first time. Vacancies are presented in Panel A of Table 1.

Pre-School 1 has the most vacancies and is the least congested grade while Primary 1 is the

[10] The deferred acceptance algorithm was selected because it is both non-strategic and stable, and because it allows policy makers to implement desirable features such as dynamic sibling priority, family linked applications, and different priority-quota combinations. The only relevant drawback of the algorithm is that it is not Pareto efficient (i.e., it might be possible to assign an applicant to a higher priority without negatively affecting another one). The Stable Improvement Cycle (SIC) algorithm and the Top Trading Cycles (TTC) algorithms are more efficient, but at the cost of losing the strategy-proofness property in the case of the SIC mechanism, and the stability property in the case of the TTC mechanism.

Moreover, as shown in Section 4 of this paper, and in Abdulkadiroğlu et al. (2017) for the case of New York, the efficiency gains obtained from using these alternatives are marginal when compared to the gains due to a transition from an uncoordinated system or a centralized one that cannot be classified as a CCAS, which is the case in this analysis.

[11] This can happen if one sibling is older than the other and will depend on the order in which the algorithm is run. If it is descending, the older sibling will give dynamic priority to the younger sibling. If it is ascending, it will be the other way around.

[12] All PII data was eliminated for that purpose.

most congested.[13]

The second data set consists of student and legal guardian information, including geolocation, applicants' sibling relationships, special educational needs, and nationality.[14] For each applicant, we have their ranked ordered list (ROL) of reported preferences, which had no length limit, and the lotteries drawn up for each program. As mentioned, the applications were done by appending applicants' initial lists (containing their preferred programs) with all other alternatives sorted by distance. The latter alternatives were included in case the applicant was not assigned to one of their preferences and thus needed to be assigned to the nearest available program. The assigned lottery numbers were different for each of the programs listed by the student, but the same lottery number was drawn for all of the programs in the appended list.

Table 1: Vacancies and Applicants by Geographic Unit (*Cantón*) and Grade

Panel A: Vacancies

Cantón	Pre-School 1	Pre-School 2	Primary 1
Manta	1,830	1,394	425
Montecristi	905	668	654
Jaramijó	110	47	37
Total Grade	**2,845**	**2,109**	**1,116**
Total Global	**6,070**		

Panel B: Applicants

Cantón	Pre-School 1	Pre-School 2	Primary 1
Manta	1,101	1,143	338
Montecristi	481	437	124
Jaramijó	125	125	107
Other	2	0	1
Total Grade	**1,709**	**1,705**	**570**
Total Global	**3,984**		

The distribution of applicants by geographic unit and grade is presented in Panel B of table 1. Notably, at least in the case of the geographic unit of Manta (Cantón), the number of applicants in Pre-School 1 and Pre-School 2 is roughly equivalent. Although this poses a challenge from a public policy standpoint in that it is desirable to enroll students earlier, it is also an interesting dynamic for the application system since families' decision to postpone the enrollment of their child(ren) puts them at a strategic disadvantage. This is because there are fewer available seats in Pre-School 2, given that currently enrolled Pre-School 1 students move automatically to the next level.

[13] This is most likely explained by a combination of factors: i) students prefer schools closer to their homes (ceteris paribus) and establishments in more crowded areas have been filled by the previously implemented distance-based algorithm; ii) students that are not satisfied with their assigned school can ask for a transfer; iii) applicants strategically reported addresses close to the more preferred schools under the previous location-based assignment system.

[14] The preferences of applicants who have a sibling already enrolled at their school of interest are specified in Panel A of Table 15 in Appendix B. We do not have information on cases in which an applicant's siblings are enrolled in schools not included in their reported preferences.

Figure 2 of Appendix A provides an overview of applicant priorities and the lengths of ranked ordered lists.Note that most applicants declared only a single preference despite there being no limits placed on the length of the preference list. This may be a legacy of the previous system in which applicants did not choose a portfolio of schools and in which it was implied that applicants were largely assigned to a school based on distance (walking or driving, as obtained from Google Maps) rather than their preferences.

A complementary explanation for the large number of short application lists is that, as observed in Figure 3 of Appendix A, applicants who replied to the survey distributed after the application period had ended indicated high expectations that they would be assigned to their top choice. These responses were obtained before the results were published to avoid bias. In the same survey, when asked why they did not add more programs to their ROL, 56% of respondents replied that they had no information on alternatives close to their home, 33% said that they were sure that they would be assigned to their reported preference, 6% said that it was difficult to find more schools, 4% declared that they preferred receiving no assignment to adding more alternatives, and 1% declared that their preferred option was a non-public school (i.e., an outside option).

In any case, the fact that the CCAS was new to families in Manta likely also resulted in them not fully adapting their behavior to the new system and rules, meaning that they may not have taken full advantage of the introduction of parental choice and preference reporting. If this is the case, our findings on the welfare gains obtained with the introduction of the CCAS system are probably downward biased when compared with the longer-term results that will eventually be obtained once families are fully accustomed to the new system.

4 Mechanism Result and Welfare Comparison

Our analysis in this section is based on the fact that applicants' reported preference orderings are an accurate representation of true family preferences. This assertion is supported by the non-strategic nature of the DA mechanism, which was furthermore emphasized in the pilot program's communication strategy. We thus compare the share of applicants who were assigned to one of their preferred options under both alternatives, and then use reported preferences to estimate welfare differences. Our welfare analysis forms part of a broader body of literature that uses structural models to study family preferences over school attributes and school choice policy counterfactuals (e.g., Neilson (2021); Kapor et al. (2020); Abdulkadiroğlu et al. (2017); Idoux (2022)).

To estimate preferences, we follow the utility model of Abdulkadiroğlu et al. (2017) and their Markov Chain Monte Carlo (MCMC) estimation methodology through Gibbs sampling (see Rossi et al. (1996) for a more detailed description and Kapor et al. (2020); Idoux (2022) for other recent implementations of this method in a school choice context).[15] Our strategy faces two major challenges: first, we have a very limited set of school covariates; and second, we assume – as is standard in the literature – that applicants lack full knowledge of the schools in their choice set (i.e., all available schools in their grade). To address the first challenge, we

[15] In the case of Kapor et al. (2020), to compare the Immediate Acceptance and the DA algorithms, and in the case of Idoux (2022), to weight the effect of policy changes, family preferences, and residential sorting.

estimate a model without any covariates. The differences in the appeal that different schools have for each family are thus fully determined by a school-specific effect that is unobservable to the econometrician, but which we assume families are aware of when comparing schools.[16] We further include a random coefficient on the distance to school parameter to allow for heterogeneity across families in the relative value of the school's mean utility and the importance of distance. With regard to the second challenge, we consider a relatively small set of alternatives that are less geographically spread than other studies, as well as assume that families know and consider all available schools within that geographic region. Tackling the consideration set formation problem is beyond the scope of this paper.

We compare the distance-centric (DC) algorithm described in subsection 2.1 with the DA algorithm using the Stable Improvements Cycles (SIC) (Erdil & Ergin, 2008) and the Top Trading Cycles (TTC) algorithms (see Abdulkadiroğlu & Sönmez (2003)) as benchmarks. The TTC algorithm is our welfare benchmark since it delivers a student-optimal assignment, and thus higher welfare than the SIC algorithm, given that the latter delivers a stable student-optimal assignment (and the stability restriction reduces attainable welfare). The TTC algorithm also results in a higher welfare than the DA algorithm, which delivers a stable but not necessarily student-optimal assignment. Given the potential multiplicity of the DA, SIC and TTC assignments, we follow the procedure described in Abdulkadiroğlu et al. (2017) to ensure a monotonic welfare comparison across all simulations. This implies that we first run the SIC algorithm over each DA assignment (as described in Erdil & Ergin (2008)), and then the TTC algorithm over the obtained DA-SIC result.[17] Following Abdulkadiroğlu et al. (2017), we run 100 lottery simulations of the DA and DC algorithms to get our welfare calculations.

With regard to the DC algorithm, one relevant point is that parents could strategically report a different address, using someone else's electricity bill (CUEN) in order to be placed at a preferred school. To include this possibility in the analysis, we run counterfactual assignments in which a random proportion of the applicants strategically choose an address close to their most preferred program. We use different random proportions as we do not have a good estimate of CUEN misreporting under the previous system.

To compare mechanisms, we first re-run the DA algorithm used in the pilot. We use the same inputs, except that we do not include students with special needs in order to make the assignment comparable to that of the DC algorithm.[18] In the implemented DA, the reported preference rankings were appended to all non-ranked programs using a linear distance sorting criterion. Applicants received a lower priority in the distance-imputed preferences to maximize assignment to the reported preferences.[19] We define assignments to imputed preference as non-preference assignments to distinguish them from the overall assignment obtained with the DC algorithm.

[16] We might include the main infrastructure of schools as a covariate, given that families could review a list of the school infrastructure in the application interface. However, we prefer a simpler model that directly focuses on the differences in mean utilities, or that which matters for our welfare comparison. The estimated infrastructure parameters would not serve any purpose in our context.

[17] In this case, the the SIC and TTC algorithms obtain the same assignment. In other words, the are no attainable Pareto efficiency improvements from relaxing stability constraints in this particular context.

[18] We eliminate both students and vacancies related to special needs, which account for only 0.23% of applicants. This decision was made because students with special needs had a special assignment round before the regular one.

[19] When referring to applicant preferences, we intend reported preferences without the distance-imputed preferences.

To replicate Ecuador's previous system (described in subsection 2.1), we consider all available programs and rank them using linear distance sorting. Students with siblings in the system were assigned (if possible) to their sibling's school before the main process was initiated. To this end, we create a priority group for these students that only applies at the schools in which their siblings are enrolled. This priority is followed by a priority for groups of siblings applying together, as these groups were processed before individual applicants in the main process. This priority is thus applied to all available programs. Finally, given that applicants were processed sequentially, we run a single tie-breaking lottery to break ties.

4.1 Utility Estimation

To estimate welfare, we first need to estimate the parameters determining the utility that families would receive from an assignment to a particular school. To this end, equation 1 presents our utility model.[20] This approach has been increasingly adopted in the literature (Abdulkadiroğlu et al., 2017; Kapor et al., 2020; Idoux, 2022), mainly because of its relative ease of implementation in ranked-ordered data contexts.

$$\begin{aligned} u_{ij} &= S_{ij}\lambda + \delta_j + (-1+\gamma_i)d_{ij} + \epsilon_{ij} \\ \delta_j &= \overline{\delta} + \xi_j \\ \overline{\delta} &\equiv 0 \end{aligned} \tag{1}$$

$$\begin{aligned} \gamma_i &\sim \mathcal{N}\left(0, \sigma_\gamma\right) \\ \xi_j &\sim \mathcal{N}\left(0, \sigma_\xi\right) \\ \epsilon_{i,j} &\sim \mathcal{N}\left(0, \sigma_\epsilon\right) \end{aligned}$$

Here, S_{ij} is a dummy variable equal to 1 if applicant i has a sibling in school j, d_{ij} is the distance between applicant i and school j, and ξ_j is a school-specific preference that is unobservable to the econometrician but that families do observe when comparing alternatives. To identify the parameters, we need to determine a scale normalization for the utility, which we do by setting $\overline{\delta} = 0$ as in Abdulkadiroğlu et al. (2017). We also include random coefficients over distance to school, represented by γ_i, in order to consider the heterogeneity in the relative importance of school attributes and travel distance for different applicants.[21] Given that in this

[20]Here, we basically adapt the model used by Abdulkadiroğlu et al. (2017), estimated through Gibbs sampling (Rossi et al., 1996). We estimate utility using the Markov-Chain Monte Carlo (MCMC), a Bayesian estimation procedure. We therefore use the same conjugate priors, specifically the Inverse-Wishart distribution. The full utility specification, including priors, is provided in Appendix D.

[21]Conceptually, in each iteration of the Gibbs sampler, utilities are drawn using the estimated parameters of the previous iteration, using reported preference rankings to restrict possible values. Specifically, assuming that i's ranking is of size R (1 being the most preferred alternative and R the least preferred), utilities are drawn iteratively using a truncated normal distribution so that:

$$u_{i,j(r=1)} > u_{i,j(r=2)} > \ldots > u_{i,j(r=R)} > u_{i,j(r=\hat{r})},\ \forall \hat{r} > R$$

To do this iterative sampling, $u_{i,j(r=1)}$ is drawn from $(u_{i,j(r=2)}, \infty)$ if $R > 1$ and using $u_{i,j(r=2)}$ from the previous iteration, and from $(-\infty, \infty)$ when $R = 1$.

utility specification, units of utility are expressed in distance units (km) – which is a result of imposing a -1 parameter on the average dis-utility of linear distance to school – using a random coefficient on distance is quite similar to using a random coefficient for school attributes as in Abdulkadiroğlu et al. (2017), in that doing so ends up affecting the relative importance of distance or school attributes. As explained above, by specifying utility in this way, we change the relative relevance of the distance to school and the school-specific unobservable.

The specification in equation 1 assumes that an applicant's utility increases when a sibling is already enrolled in a school. This, however, does not take into account that the reason for the sibling being enrolled in that school may be because the family liked the school when the sibling enrolled (or transferred) in the first place. This implies that ϵ_{ij} is not random for such cases, highlighting the bias in the estimation. In Appendix B.1 we therefore present our preference parameter estimates and welfare calculations with no sibling-related considerations. For robustness, we further present our results without including the random coefficient on distance. Overall, the findings and conclusions remain the same across these alternate specifications.

As in Abdulkadiroğlu et al. (2017), identification relies on the assumption that families report their preferences truthfully and consider all the alternatives within their geographic unit. Likewise, the key conditional independence assumption is that

$$(\gamma_i, \epsilon_{ij}) \perp d_{ij} | \xi_j$$

which, in our case, implies that conditional on the vertical school-specific parameter, unobserved tastes for programs are independent of linear distance to school. The previous system, in which families could borrow or buy an electricity bill near their preferred school instead of actually changing their residence, aligns with this conditional independence assumption.

Table 6 in Appendix B presents the estimates from equation 1, and the potential scale reduction factors (Gelman et al., 1992) to assess mixing and convergence of the Gibbs sampling procedure (values close to one imply convergence). We also present the trace plots of the estimated σ_ϵ in each iteration of the Gibbs sampling in Figure 6 of Appendix A. We discarded the initial 50,000 iterations of the Gibbs sampler as a burn period and used the following 100,000 to compute the mean parameters and standard deviations. The trace plots show that the values of σ_ϵ remained stable. Estimates eliminating random coefficients from equation 1 are presented in Table 11 of Appendix B.1, and estimates of the main specification without siblings are presented in Table 12 of the same appendix section. We observe that estimated parameters are very similar across the three alternative models, consistent with the similarity of the welfare estimates using the different specifications.

As in Abdulkadiroğlu et al. (2017), we estimate utilities conditional on the estimated parameters and, importantly, also conditional on the reported preference rankings:

$$\mathbb{E}\left[u_{i,j} | r_i, \xi, \lambda, \sigma_\epsilon, \sigma_\xi, \Sigma_\gamma, d_i\right]$$

Here, r_i represents i's reported preference ranking, and to compute i's expected utility if assigned to school j we directly average over the iterations of the Gibbs sampler procedure, which allows us to easily condition on reported preference rankings. Estimated average utilities are measured in kilometers (km), which is a feature of using a scale normalization of -1 on the linear distance parameter.

5 Results

In this section, we begin by describing the differences between systems in terms of assignment to preferences and linear distance to home. We then present our welfare comparison using the utility model introduced above.

Notably, the reported preferences of 55.5% of the applicants (n.2,206) match the ranking used to simulate the distance assignment mechanism, reflecting the importance of distance between home and school to families. Specifically, this means that the first preference of over half of the applicants was the closest school to their homes (or the closest one where a sibling is enrolled). Table 2 compares the results of a single simulation of the DA and DC algorithms. In the distance-centric alternative, a significantly lower percentage of applicants are assigned to their preferred school. However, the percentages are quite similar for applicants with a strong preference for distance, as can be observed in rows 1 and 2 of Panel B. This outcome highlights that using a coordinated alternative that considers preferences does not harm (at least on average) applicants worried mainly about distance to school. In terms of the average linear distance of the assignment, we see that the DA algorithm assigned students to schools an average of 0.29 km farther away than the DC algorithm.

Table 2: Mechanism Comparison - Results

	DA	Distance
Panel A: Applicants assigned in:		
Any preference	3,118	1,986
	(78.44%)	(49.96%)
First preference	2,773	1,686
	(69.76%)	(42.42%)
Average assignment distance	1.30km	1.01km
Panel B: Applicants with the same 1st preference (2,206) assigned in:		
Any preference	1,779	1,537
	(81.16%)	(70.12%)
First preference	1,644	1,472
	(75.00%)	(67.15%)

Figure 4 in Appendix A shows the assignment to different declared preference rankings for both systems. As we can see, the DA algorithm assigns more students to their first preference than the DC algorithm (70% to 42%) and much fewer students to an alternative outside of their reported preference list (22% vs 50%). Tables 7 to 9 in Appendix B and Figure 5 in Appendix A display these results for the different grades. Greater congestion leads to smaller differences between the two mechanisms in terms of applicants being assigned to their preferred options. However, there are two forces at play. On the one hand, more congestion implies that fewer applicants are assigned to a reported preference when using the DA alternative. On the other hand, under the distance-based alternative, more congestion increases the probability that one applicant who is placed in a closer school displaces another who would have ranked that school at the top of their list (particularly in the cases where the latter applicant's first preference and closest school coincide).

To evaluate the effect of location misreporting, we compute counterfactual assignments in which a random sample of applicants report the location of their most preferred school as their address instead of their true residence. The exercise simulates cases in which families submit another household's electricity bill to maximize the likelihood of being assigned to their most preferred school. We compute assignments with misreporting levels of 10%, 30%, 50%, 70%, and 90%.

Table 3: Mechanism Comparison with Location Misreporting

	Applicants assigned to any preference	*Average Distance*
Distance Mech without misreporting	1,986 (49.96%)	1.01km
Distance Mech + 10% misreporting	2,055 (51.70%)	1.04km
Distance Mech + 30% misreporting	2,142 (53.89%)	1.09km
Distance Mech + 50% misreporting	2,230 (56.10%)	1.18km
Distance Mech + 70% misreporting	2,332 (58.67%)	1.21km
Distance Mech + 90% misreporting	2,403 (60.45%)	1.29km
DA Algorithm	3,118 (78.44%)	1.30km

The results of this exercise show that, as the percentage of applicants who change their location increases, the percentage of applicants assigned to one of their preferences rises as well (from 50% to 59%). Nevertheless, the rates of assignment to a preferred option does not reach the level of the DA algorithm, since applicants who misreport their location can only signal a preference for a single alternative. If, however, they are not assigned to that alternative, they may end up in a school farther from home and less preferred to other options (some of which may be closer to their true location). When misreporting is greater, the true average distance to school (i.e., using the real and not the reported location) increases as well. At 90% misreporting, the distance to home reaches the same level as in the DA alternative. This implies that misreporting can close the gap to the DA mechanism only partially and at the cost of rapidly increasing the (true) distance between applicants' home and school.

Table 4 presents the estimated differences in mean utilities (both in km), as well as in standard deviations with respect to a student-optimal (TTC) assignment benchmark.[22] In Panel A, we can see that differences between the DA and TTC algorithms are small in terms of welfare (less than 80 meters) at the pre-school levels, and significantly smaller than the welfare loss under the DC (distance) alternative (0.689km and 0.430km on average, respectively). The difference is larger when we consider only applicants assigned to different schools under the different algorithms, as shown in Panel B.

In Primary 1, the difference between the DA and the DC algorithms is smaller because more applicants are assigned to a non-preferred alternative due to increased congestion, as shown in Figure 5 of Appendix A. Moreover, in Table 10 of Appendix B, we can see that the share of applicants assigned to the same non-preferred school in both algorithms increases significantly in Primary 1 (56.7% of applicants assigned to the same school, compared to 2.69% and 21.35%

[22]SIC and TTC actually have the exact same assignment in all 100 simulations, as explained in Appendix E.

Table 4: Differences in Welfare: Student-Optimal vs. DC and DA Algorithms

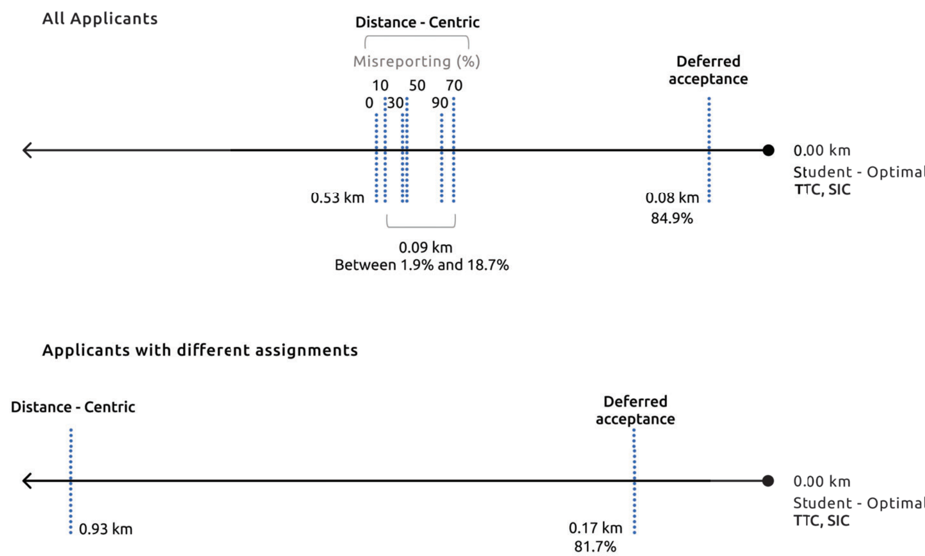
Measure	Pre-School 1 DC	Pre-School 1 DA	Pre-School 2 DC	Pre-School 2 DA	Primary 1 DC	Primary 1 DA
Panel A: All simulated applicants						
Δ Mean utility (km)	-0.689	-0.003	-0.430	-0.076	-0.318	-0.306
$\frac{\Delta Mean\ utility}{\sigma_{Ut.\ FB}}$	-0.699	-0.003	-0.158	-0.028	-0.094	-0.090
Panel B: Applicants with different assignments across algorithms						
Δ Mean utility (km)	-1.486	-0.005	-0.666	-0.118	-0.417	-0.402
$\frac{\Delta Mean\ utility}{\sigma_{Ut.\ FB}}$	-1.466	-0.005	-0.263	-0.046	-0.127	-0.121

Δ Mean utility (km) is measured computing $u_{i,j(\mu)} - u_{i,j(TTC)}$, where $j(\mu)$ represents the school to which individual i is assigned under mechanism μ. We then compute average utilities for each algorithm and simulation and finally compute the average for each algorithm across simulations. $\frac{\Delta Mean\ utility}{\sigma_{Ut.\ FB}}$ simply uses the utility variance under the TTC mechanism to scale this difference in each simulation. This is done to facilitate extrapolations to other contexts. This same table is presented in Appendix B.1 for the specification without siblings.

in Pre-School 1 and 2 respectively). Furthermore, conditional on having a different assignment in the DC and DA algorithms, the share of applicants who move from a non-preferred to a preferred assignment under the DA algorithm is 40.85% in Primary 1, compared to 77% in Pre-School 1 and 53% in Pre-School 2 (i.e., the share of applicants with improved outcomes is smaller in later years). Finally, the DC algorithm finds on average schools that are closer to home, which is a feature of not prioritizing reported preferences. Given that utility is on average greater for applicants with a lower home-to-school distance, this leads to a lower average difference in utility between mechanisms.

The distribution of estimated welfare overall and in each grade is presented in Figure 7 of Appendix A. Here, we observe that the phenomenon described in the above paragraph occurs in all grades, with two peaks in utility in each figure: one among applicants assigned to a preferred option and another for those assigned to a non-preferred option that is close to home. The DA and TTC (and SIC) algorithms have very similar distributions. However, the TTC algorith does improve the assignment relative to the DA algorithm in Primary 1, which is explained by the fact that, with higher congestion, stability constraints imposed by tie-breaking lotteries are more restrictive. By eliminating them, TTC (and SIC) achieve a significant improvement (0.318 km overall over the DA assignment and 0.417 km if restricted to applicants with different assignments), as shown in Table 4.

Figure 1: Welfare Differences between Algorithms (km)



Finally, Figure 1 compares welfare gains under different mechanisms, similar to Figure 5 in Abdulkadiroğlu et al. (2017). The main takeaway is that, while the differences in magnitudes between our paper and that of Abdulkadiroğlu et al. (2017) are large, the proportions are actually very similar. Thus, the coordinated mechanisms that include applicant preferences and the alternatives considered in each case are roughly the same. Meanwhile, the differences in magnitudes are explained by the different settings of New York and Manta. In addition, Abdulkadiroğlu et al. (2017) study assignment to secondary school (where applicants are willing to travel more), while we assess assignment to pre-school and early primary school (where families place a greater value on distance from home).

With regard to improvements over the DA algorithm, the potential is context-dependent, as shown by the differences observed in the various grades. The margin, therefore, is not irrelevant, but likely to be more important in more congested grades (i.e., post-entry-level grades). It would arguably be best to focus on implementing a CCAS first, then use the reported preferences to study the potential of the SIC and TTC algorithms (and possibly others), before weighing the trade-offs between improving Pareto efficiency and losing stability or the strategy-proof properties.

6 Discussion

In this paper, we document and study the effects of the Centralized Choice and Assignment System (CCAS) pilot implemented in early 2021 in the Ecuadorian city of Manta, where the previously existing system assigned students exclusively by the linear distance between their home and school. We contrast these systems using a sudden change in policy and the data generated by the new non-strategic deferred acceptance (DA) algorithm (Gale & Shapley, 1962). We estimate preferences closely following Abdulkadiroğlu et al. (2017) and use these to quantitatively study the welfare effects of the policy change.

Our main result is that implementing a coordinated mechanism that incorporates applicants' preferences has relatively large welfare benefits. This finding echoes that observed for New York City in Abdulkadiroğlu et al. (2017), though here, our setting is a developing country. The extent of the differences depends on different factors such as the congestion of schools, the grades considered, the characteristics of the city (e.g., residential segregation), and family preferences.

Specifically, we document that when compared to the previous distance mechanism used by the government, the DA algorithm increases the percentage of applicants assigned to a preferred school from 49.96% to 78.44%, while the percentage of applicants assigned to their first preference rises from 42.42% to 69.76%. The main trade-off of implementing the DA alternative is that the average linear distance between applicants' home and school increases by 0.29km. To assess the overall effect of the policy, we turn to our estimated welfare comparisons and show that the welfare gains are between 0.683km and 0.012km higher when the DA algorithm is used. Meanwhile, if the analysis is restricted only to applicants who are assigned to a different school under each alternative mechanism, these gaps more or less double in magnitude.

The difference in welfare gains is smaller in the first year of primary school than in either year of pre-school because congestion is much more significant and many applicants cannot be assigned to any of their preferred options under either mechanism.

The results showing that coordinated school choice systems are beneficial for families in developing country contexts is important in that more and more countries are adopting similar systems worldwide.[23] These findings are also timely since the COVID-19 pandemic has accelerated the adoption of online application and enrollment systems, facilitated by the fact that an increasing number of countries now have the pre-conditions to realize them. While various aspects of these policies beg further study, the growing body of evidence that preferences and coordination are a central driver of welfare gains is one important step in better understanding how to implement this type of market design in practice.

[23] See Neilson (2021) and www.ccas-project.org.

References

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

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

Abdulkadiroğlu, A. & Sönmez, T. (1998). Random serial dictatorship and the core from random endowments in house allocation problems. *Econometrica*, 66(3), 689–701.

Arteaga, F., Kapor, A. J., Neilson, C. A., & Zimmerman, S. D. (2021). *Smart Matching Platforms and Heterogeneous Beliefs in Centralized School Choice*. Working Paper 28946, National Bureau of Economic Research.

Correa, J., Epstein, R., Escobar, J., Rios, I., Bahamondes, B., Bonet, C., Epstein, N., Aramayo, N., Castillo, M., Cristi, A., & et al. (2019). School choice in chile. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, EC '19 (pp. 325–343). New York, NY, USA: Association for Computing Machinery.

Erdil, A. & Ergin, H. (2008). What's the matter with tie-breaking? improving efficiency in school choice. *American Economic Review*, 98(3), 669–89.

Gale, D. & Shapley, L. S. (1962). College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1), 9–15.

Gelman, A., Rubin, D. B., et al. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4), 457–472.

Idoux, C. (2022). Integrating new york city schools: The role of admission criteria and family preferences. *Job Market Paper, MIT Economics Department*.

Kapor, A. J., Neilson, C. A., & Zimmerman, S. D. (2020). Heterogeneous beliefs and school choice mechanisms. *American Economic Review*, 110(5), 1274–1315.

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

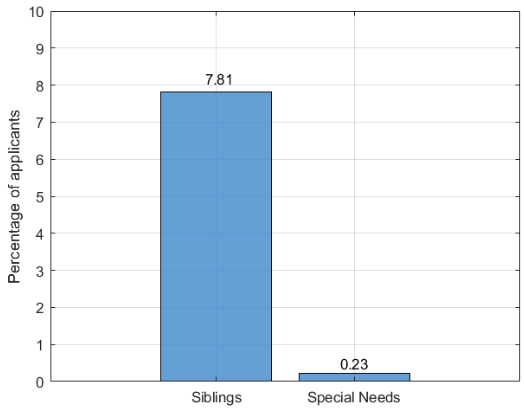
Pathak, P. A. (2011). The mechanism design approach to student assignment. *Annual Review of Economics*, (4), 513–36.

Pathak, P. A. (2017). What really matters in designing school choice mechanisms.

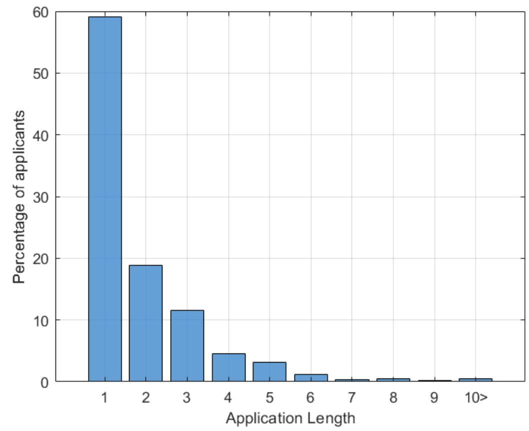
Rossi, P. E., McCulloch, R. E., & Allenby, G. M. (1996). The value of purchase history data in target marketing. *Marketing Science*, 15(4), 321–340.

A Figures

Figure 2: Distribution of Declared Applicant Priorities and Ranked Ordered List Size

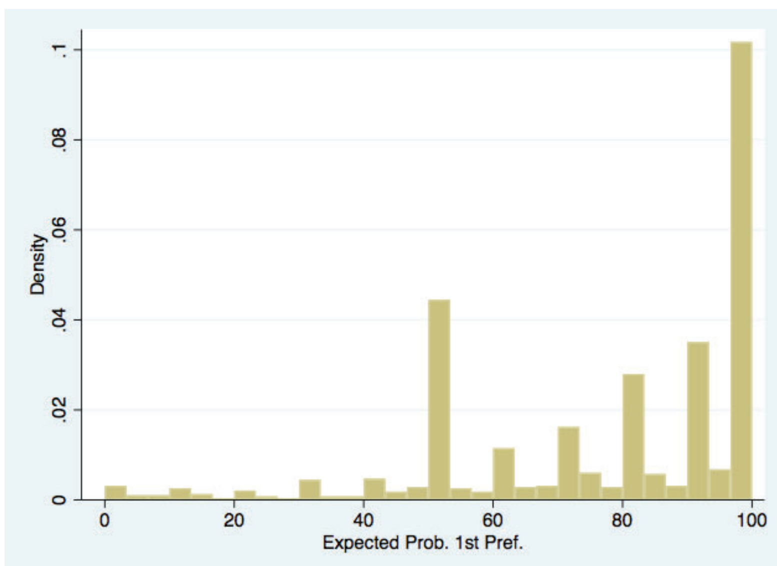


Priorities declared by applicants



Application length

Figure 3: Perceived Probability of Admission to 1st Preference



These responses were obtained in an online survey carried out after the end of the application period but before assignment results were communicated (to avoid biasing responses).

Figure 4: Ranking Assigned: DA and Distance Mechanism

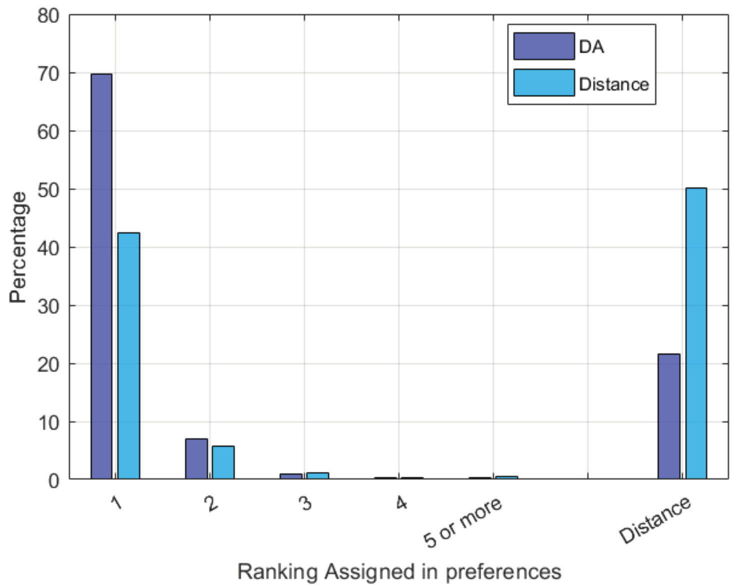


Figure 5: Ranking Assigned by Grade: DA and Distance Mechanism

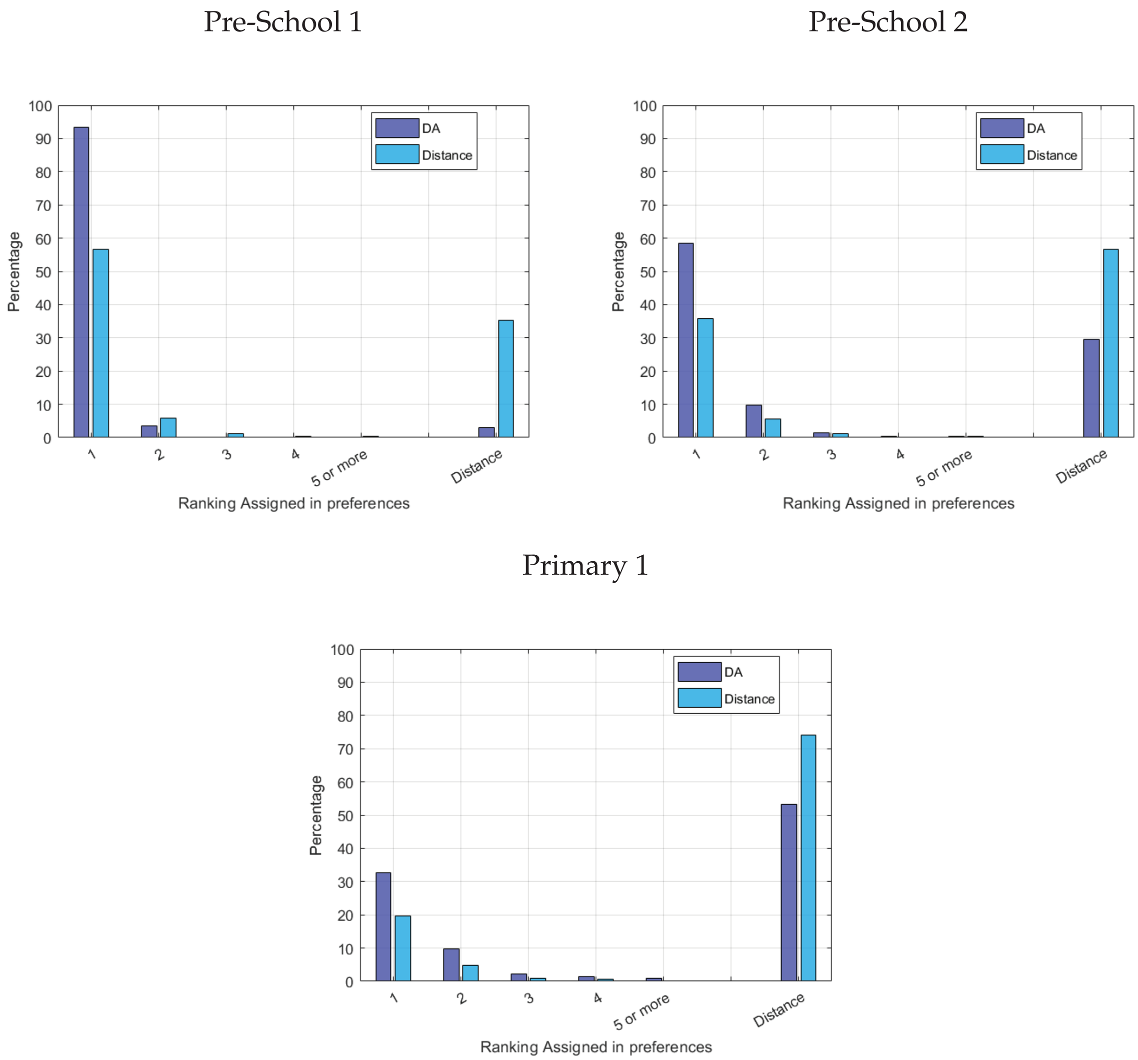
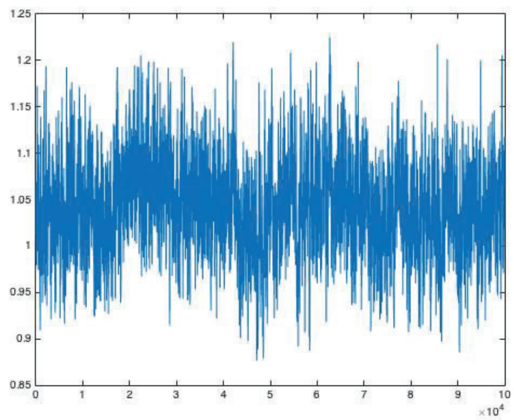
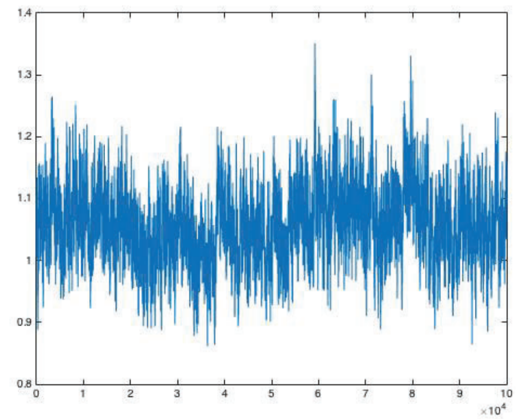


Figure 6: Trace Plots σ_ϵ in Main Specification

Pre-School 1



Pre-School 2



Primary 1

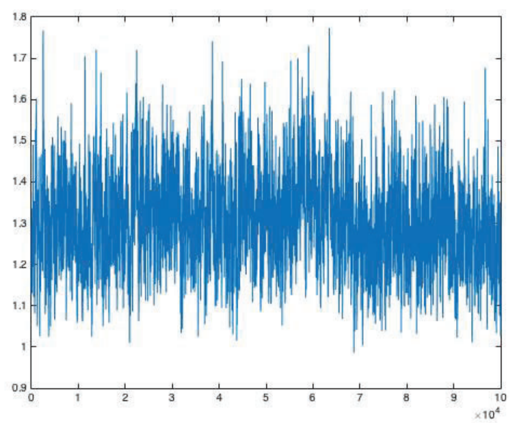
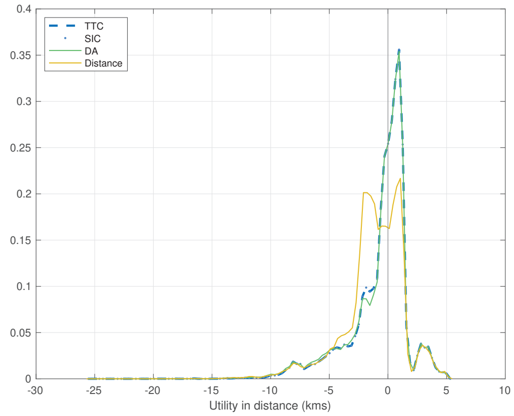
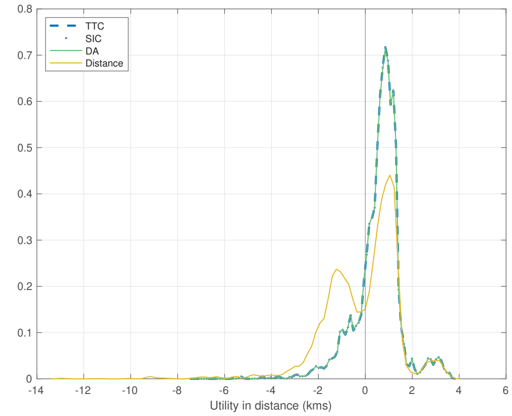


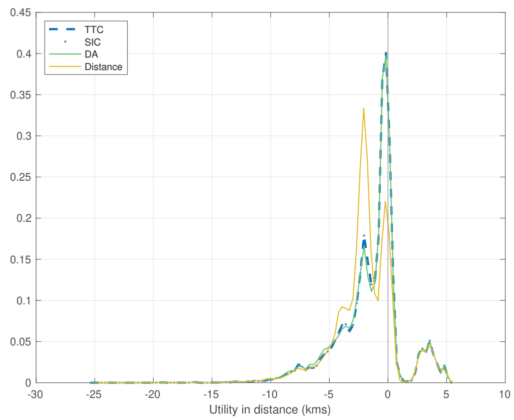
Figure 7: Welfare Distribution



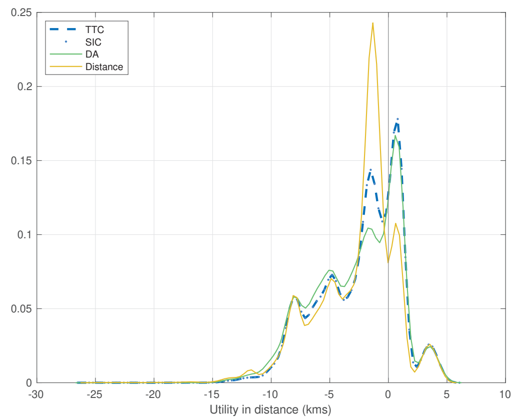
All applicants



Pre-School 1



Pre-School 2



Primary 1

In this figure, we plot the utilities obtained with our model when using the scale normalizations $\overline{\delta} \equiv 0$ and -1 as the average disutility from each linear km of distance between the school and the reported location of the family. The level of utility is not relevant, as it depends on the normalization. However, the mass from the utility distribution when using the distance-centric algorithm being shifted to the left is relevant, as it indicates how the relative distributions of utilities compare, and lead to the average differences presented in Table 4.

Figure 8: Figure 5 of Abdulkadiroğlu et al. (2017)

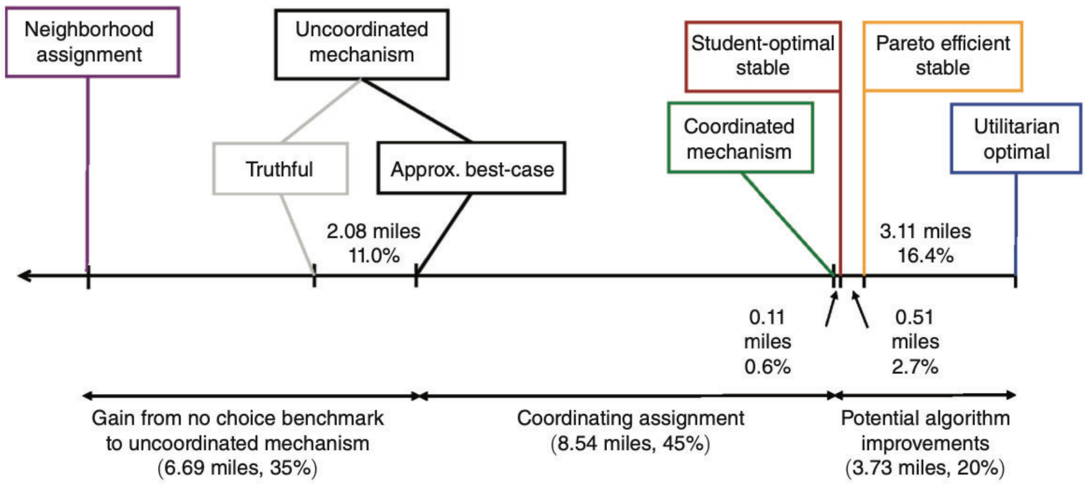


Figure 5. Coordinating Assignments versus Algorithm Improvements

B Tables

Table 5: Comparative Statistics for Guayaquil and Manta

	Guayaquil	**Manta**
Total population	2,291,158	221,122
Population 3-5 years old (% of population 3-17 years old)	19.7	19.1
Minors in the school system (% of population 3-17 years old)	78.9	80.2
Average Mother's Education (of minors 3-17 years old)	11.3 years	10.8 years
Total schools	885	153
Share of public schools	54%	43%
Share of private schools	44%	54%
Share of "fiscomisional" schools	2%	3%
Total enrollment	687,046	86,455
Share of enrollment in public schools	57%	67%
Share of enrollment in private schools	40%	25%
Share of enrollment in "fiscomisional" schools	4%	8%

Table 6: Estimates and Potential Scale Reduction Factors: Main Specification

	Pre-School 1		Pre-School 2		Primary 1	
Estimate	Mean (SD)	PSRF	Mean (SD)	PSRF	Mean (SD)	PSRF
ξ_j	0.073		-0.918		0.048	
	(0.492)		(0.559)		(0.648)	
λ	3.055	1.003	4.732	1.007	3.790	1.001
	(0.365)		(0.562)		(0.842)	
σ_ξ	0.291	1.001	1.251	1.017	0.540	1.004
	(0.066)		(0.453)		(0.143)	
σ_ϵ	1.042	1.060	1.057	1.025	1.303	1.020
	(0.048)		(0.060)		(0.102)	
σ_γ	1.253		1.690		1.079	
Tot. schools	55		57		54	
Tot. students	1,098		885		389	

Table 7: Mechanism Comparison - Results Pre-School 1

	DA	DC
Panel A: Applicants assigned in:		
Any preference	1,654	1,102
	(96.95%)	(64.60%)
First preference	1,592	966
	(93.32%)	(56.62%)
Average assignment distance	0.87km	0.52km
Panel B: Applicants with the same 1st preference (985) assigned in:		
Any preference	965	955
	(97.97%)	(96.95%)
First preference	942	946
	(95.63%)	(96.04%)

Table 8: Mechanism Comparison - Results Pre-School 2

	DA	DC
Panel A: Applicants assigned in:		
Any preference	1,199	737
	(70.45%)	(43.30%)
First preference	93.32	609
	(58.52%)	(35.78%)
Average assignment distance	1.32km	1.08km
Panel B: Applicants with the same 1st preference (959) assigned in:		
Any preference	708	607
	(73.83%)	(63.30%)
First preference	618	569
	(64.44%)	(59.33%)

Table 9: Mechanism Comparison - Results Primary 1

	DA	DC
Panel A: Applicants assigned in:		
Any preference	265	147
	(46.74%)	(25.93%)
First preference	185	111
	(32.63%)	(19.58%)
Average assignment distance	2.56km	2.29km
Panel B: Applicants with the same 1st preference (281) assigned in:		
Any preference	144	112
	(51.25%)	(39.86%)
First preference	116	103
	(41.28%)	(36.65%)

Table 10: Assignment In and Out Preferences under DA and DC Algorithms

	Pre-School 1	Pre-School 2	Primary 1
Applicants assigned to same schools under DA and DC	967	801	261
	(56.68%)	(47.06%)	(46.03%)
Applicants assigned to different schools under DA and DC	739	901	306
	(43.32%)	(52.94%)	(53.97%)
Applicants assigned to same schools under DA and DC			
Both DA and DC in preferences	941	630	113
	(97.31%)	(78.65%)	(43.30%)
Both DA and DC out of preferences	26	171	148
	(2.69%)	(21.35%)	(56.70%)
Applicants assigned to different schools under DA and DC			
DA in preferences and DC out of preferences	569	478	125
	(77.00%)	(53.05%)	(40.85%)
DA out of preferences and DC in preferences	17	16	7
	(2.30%)	(1.78%)	(2.29%)
Both DA and DC in preferences	144	91	27
	(19.49%)	(10.10%)	(8.82%)
Both DA and DC out of preferences	9	316	147
	(1.22%)	(35.07%)	(48.04%)

B.1 Appendix Robustness Checks

Table 11: Estimates and Potential Scale Reduction Factors. Main Specification without Random Coefficients

	Pre-School 1		Pre-School 2		Primary 1	
Estimate	Mean (SD)	PSRF	Mean (SD)	PSRF	Mean (SD)	PSRF
ξ_j	0.131		-1.480		0.075	
	(0.532)		(0.693)		(0.606)	
λ	3.486	1	4.719	1	4.301	1
	(0.396)		(0.561)		(0.919)	
σ_ξ	0.350	1.001	2.766	1	0.497	1
	(0.082)		(0.747)		(0.131)	
σ_ϵ	1.403	1	1.314	1	1.702	1
	(0.054)		(0.061)		(0.121)	
Tot. schools	55		57		54	
Tot. students	1,098		885		389	

Table 12: Estimates and Potential Scale Reduction Factors. Main Specification without Siblings

	Pre-School 1		Pre-School 2		Primary 1	
Estimate	Mean (SD)	PSRF	Mean (SD)	PSRF	Mean (SD)	PSRF
ξ_j	0.070		-1.121		0.044	
	(0.500)		(0.559)		(0.643)	
σ_ξ	0.300	1.005	1.661	1	0.536	1
	(0.068)		(0.534)		(0.142)	
σ_ϵ	1.089	1.025	0.968	1.038	1.306	1.001
	(0.062)		(0.054)		(0.101)	
σ_γ	1.462		1.376		0.896	
Tot. schools	55		57		54	
Tot. students	1,021		839		345	

Table 13: Differences in Welfare: Student-Optimal vs. DC and DA algorithms. Specification without Random Coefficients

	Pre-School 1		Pre-School 2		Primary 1	
Measure	Dist	DA	Dist	DA	Dist	DA
Panel A: All simulated applicants						
Δ Mean utility (km)	-0.773	-0.003	-0.456	-0.071	-0.317	-0.296
$\frac{\Delta \textit{Mean utility}}{\sigma_{Ut.\ FB}}$	-0.750	-0.003	-0.177	-0.027	-0.091	-0.085
Panel B: Applicants with different assignments across algorithms						
Δ Mean utility (km)	-1.667	-0.006	-0.707	-0.109	-0.415	-0.388
$\frac{\Delta \textit{Mean utility}}{\sigma_{Ut.\ FB}}$	-1.631	-0.006	-0.302	-0.047	-0.124	-0.116

Δ Mean utility (km) is measured computing $u_{i,j(\mu)} - u_{i,j(TTC)}$, where $j(\mu)$ represents the school to which individual i is assigned under mechanism μ. We then compute average utilities for each algorithm and simulation and finally compute the average for each algorithm across simulations. $\frac{\Delta \textit{Mean utility}}{\sigma_{Ut.\ FB}}$ simply uses the utility variance under the TTC mechanism to scale this difference in each simulation. This is done to facilitate extrapolations to other contexts.

Table 14: Differences in welfare: Student-optimal vs DC and DA algorithms. Specification without Siblings

	Pre-School 1		Pre-School 2		Primary 1	
Measure	Dist	DA	Dist	DA	Dist	DA
Panel A: All simulated applicants						
Δ Mean utility (km)	-0.717	-0.004	-0.348	-0.083	-0.176	-0.361
$\frac{\Delta \textit{Mean utility}}{\sigma_{Ut.\ FB}}$	-0.756	-0.004	-0.154	-0.037	-0.052	-0.104
Panel B: Applicants with different assignments across algorithms						
Δ Mean utility (km)	-1.560	-0.009	-0.543	-0.130	-0.228	-0.470
$\frac{\Delta \textit{Mean utility}}{\sigma_{Ut.\ FB}}$	-1.442	-0.008	-0.218	-0.052	-0.066	-0.133

Δ Mean utility (km) is measured computing $u_{i,j(\mu)} - u_{i,j(TTC)}$, where $j(\mu)$ represents the school to which individual i is assigned under mechanism μ. We then compute average utilities for each algorithm and simulation and finally compute the average for each algorithm across simulations. $\frac{\Delta \textit{Mean utility}}{\sigma_{Ut.\ FB}}$ simply uses the utility variance under the TTC mechanism to scale this difference in each simulation. This is done to facilitate extrapolations to other contexts.

Table 15: Priorities and Assignments in DA: Potential Improvements for SIC and TTC

	Pre-School 1	Pre-School 2	Primary 1
Panel A:	*Ranking of schools where an applicant has sibling priority(*)*		
1st preference	100	157	42
2nd preference	7	3	1
3rd preference	1	2	0
Panel B:	*Ranking of DA assignments for applicants with sibling priority below 1st preference*		
1st preference	8	3	0
2nd preference	0	2(**)	1(***)
3rd preference	0	0	0

Note: None of the potential applicants that could participate in an improvement cycle (Panel B) coincide in the programs to which they were applying, such that no cycles were attainable.
(*) Panel A shows the highest ranked program where applicants have a sibling priority. If an applicant has priority in both the 1st and 2nd preference, they will only appear in the 1st preference in this table.
(**) One of these two applicants had sibling priority in their second preference, and the other had sibling priority in their third preference.
(***) This applicant had sibling priority in their second preference.

C Phases of the Distance-Centric Algorithm Implementation Process

The overall process started with the Preparation Phase, in which the Ministry of Education updated all school supply information (i.e., location, available spaces, closure or opening of educational programs, etc.).

In the second, or Registration Phase, families registered their children on a website in order to be granted a spot in a public school. Legal guardians needed to indicate the type of registration (individual or sibling group), the grade level to be attended, any older siblings already enrolled in the public school system, special educational needs, and nationality. They also provided their electricity bill number so as to be geo-located.

This was followed by the Assignment Phase and then the Consultation Phase, during which time families could enter the website to see their school assignments. Finally, the fifth and sixth phases consisted of the School Change Petitions Phase and Continuous Enrollment. Applicants could ask to change schools if there were spaces available, and they could also enroll in a given school once the academic year had already started.

D Full Utility Specification

$$
\begin{aligned}
u_{ij} =& S_{ij}\lambda + \delta_j - d_{ij} + \gamma_i d_{ij} + \epsilon_{ij} \\
\delta_j =& \bar{\delta} + \xi_j \\
\bar{\delta} \equiv& 0
\end{aligned}
$$

$$
\begin{aligned}
\lambda \sim& \mathcal{N}\left(0, \sigma_\lambda\right) \\
\gamma_i \sim& \mathcal{N}\left(0, \sigma_\gamma\right) \\
\xi_j \sim& \mathcal{N}\left(0, \sigma_\xi\right) \\
\epsilon_{i,j} \sim& \mathcal{N}\left(0, \sigma_\epsilon\right) \\
\sigma_\gamma \sim& IW(\tau_\gamma, df_\gamma) \\
\sigma_\xi \sim& IW(\tau_\xi, df_\xi) \\
\sigma_\epsilon \sim& IW(\tau_\epsilon, df_\epsilon)
\end{aligned}
$$

We follow Rossi et al. (1996) and Abdulkadiroğlu et al. (2017) in using disperse priors. The only exception is the use of a smaller τ_γ, given that in this context it is reasonable to impose a smaller prior on the mean variance of the parameter, considering that $\gamma_i > 1$ would imply that a family actually prefers schools farther away from home. Specifically, we use $\sigma_\lambda = 100$, $\tau_\gamma = 0 + size(\gamma_i) = 1$, $df_\gamma = 3 + size(\gamma_i) = 1$,[24] $\tau_\xi = 1$, $_\xi = 2$, $\tau_\epsilon = 3 + n_{schools}$, and $_\epsilon = 3 + n_{schools}$.

[24]This implies that the mean of the σ_γ prior is 0.5.

E DA-SIC and TTC Equivalence in our Context

As shown in Table 15, there is no potential for priority trading cycles.

The Top Trading Cycles (TTC) algorithm includes the possibility of trading priorities between applicants, which happens when they prefer the alternatives in which they do not have the priority more than ones in which they do, and are thus "willing to trade" the priority. In other words, TTC has the potential to provide improvements over SIC, when there is not a complete correlation between priorities and preferences. In our case, for the priority at declared preferences (over non-preferences imputed by distance), the correlation is one since these are always ranked higher. Thus, the only possibility for the TTC algorithm to improve over the SIC algorithm is to find trades involving the static sibling priority. However, as shown in Table 15 (and explained in the footnote), that is not feasible.

To illustrate this, imagine a system with two schools (A and B), both with only one vacancy, and three applicants(i, j and k). i has priority in A but prefers B over A.j has priority in school B, but prefers A over B. k has priority in both schools, prefers A over B, and has the worst lottery number of the system. The result of the DA and SIC assignment would be i assigned to A and j assigned to B. The TTC algorithm would allow them to trade their priorities and switch their assignments. With that assignment switch, applicant k is now unassigned but has a higher priority in both schools that rejected him (higher priority pre-trade, of course). Such a situation can only arise when the correlation between preference and priority is not one, thus leaving room to trade the priority and get a better assignment.