

# Priorities in school choice



# Priorities in School Choice: The case of the Boston mechanism in Barcelona*

Caterina Calsamiglia†
ICREA & IPEG

Maia Güell‡
The University of Edinburgh,
CEPR, FEDEA & IZA

This version: April 2018

**Abstract**

The Boston mechanism is a school allocation procedure that is widely used around the world and has been criticized for its incentive problems. In order to resolve overdemands for a given school, most often priority is given to families living in the neighborhood of the school. Using a very rich data set on school applications for the case of the Boston mechanism in Barcelona, we exploit an unexpected change in the definition of neighborhood. This change allows us to identify that a large fraction of families systematically ranks first high priority schools, neighborhood schools in this case. Additional data on school enrollment decisions and census data shows that some seemingly unsophisticated parents are high income families that can rank hard-to-get schools because they can afford the outside option of a private school in case they do not get in. This sheds light on important inequalities beyond parents' lack of sophistication found in the literature.

**Key words**: School choice, Enrollment, Boston mechanism, Priorities, Private schools.
**JEL classification**: C78, D63, I24

*This paper was previously circulated as "The Illusion of School Choice: Empirical Evidence from Barcelona". We are very grateful to Manel Blasco, Jordi Baró, and Francesc Iñiguez for providing us with the school applications and enrollment data; to Manel Blasco, Eduard Vallory, and Frederic Udina for allowing us to merge the school application data with the census data under the CEB-BGSE-IDESCAT institutional agreement; to Miquel Delgado at IDESCAT for merging the data sets; and to Javier Asensio, Isaac Aparicio, Albert Esteve, Sabine Flamand, and Diego Puga for helping us with the street map data. Anna Muñoz provided outstanding research assistance. We also thank Guillermo Caruana, Antonio Ciccone, Doireann Fitzgerald, Guillaume Haeringer, Philipp Kircher, Laura Mayoral, Derek Neal, Mar Reguant, Sevi Rodríguez Mora, two anonymous referees, the Editor, Jonah Rockoff, and participants at numerous seminars and conferences for fruitful discussions and comments. We are grateful to Joan Gieseke for editorial assistance. Calsamiglia acknowledges financial support by the Fundación Ramón Areces, the Ramón y Cajal contract of the Spanish Ministry of Science and Technology, the Spanish Plan Nacional I+D+I (SEJ2005-01481, SEJ2005-01690 and FEDER), and the Grupo Consolidado de tipo C (ECO2008-04756), the Generalitat de Catalunya (SGR2005-00626), and the Severo Ochoa program and the ERC Starting Grant 638893. Güell acknowledges the financial support from the projects MINECO-ECO2014-59225-P and ECO2017-89240-P from the Spanish Ministerio de Economia y Competitividad, the project and the hospitality of the University of Minnesota and the Federal Reserve Bank of Minneapolis, where part of this work was written. Our online appendix is available at `http://www.maiaguell.com/SchoolsBCN/OnlineAppendixChoiceBarcelona.pdf`.

†Email: ccalsamig@gmail.com

‡Corresponding Author: Maia Güell, The University of Edinburgh, School of Economics, 30 Buccleuch Place, Edinburgh EH8 9JT, United Kingdom. Email: maia.guell@gmail.com

# 1 Introduction

During the last two decades, a large fraction of OECD countries have implemented different forms of public school choice that depart from residential-based assignment (see Musset (2012)). The aims of these school choice programs are to improve the matching between children and schools as well as to improve students' educational outcomes.[1] Yet, the concern is that disadvantaged families are less able to exercise choice, which leads to equity concerns.[2]

Under centralized school choice procedures, parents are asked to submit a list with their ranking of schools, and a set of rules determines the final allocation. A widely used procedure in school choice is the so-called Boston mechanism, henceforth the BM.[3] This mechanism assigns all applicants to the school ranked first, and if there is overdemand for a school, ties are broken according to priority points based on criteria such as having siblings in the school, living near the school or a lottery number. Those rejected from their school ranked first can only opt for the seats that remain free after the first round. The procedure in subsequent rounds follows a similar logic. Cities worldwide differ in their actual implementation of the BM but a common key feature is that the assignment in every round is final. As the literature starting with Abdulkadiroğlu and Sönmez (2003) has emphasized, the optimal strategy for parents under the Boston mechanism depends on what other parents are doing, and telling the truth is rarely optimal. Parents may avoid overdemanded schools and rank only relatively *safer* schools. The risk involved in ranking a school for a given family depends on other applicants' behaviour and on priority points. However, in general, the literature has overlooked the role of priority points in shaping parents' behaviour. Given other families' behaviour, priorities induce a discontinuity in the admission probabilities for the different schools for different families. That is, having priorities or not increases crucially the admission probabilities faced by a family.

In this paper, we use a unique and rich administrative data set containing application, enrollment and socioeconomic information to explore a variation of the Boston mechanism used in the city of Barcelona where priority is given to residence. We exploit an unexpected change in the definition of *neighborhood* in Barcelona. This provides an exogenous change in the set of schools for which families have priority and allows us to analyze the impact that priorities have on school choice, independently of housing decisions. We find that many families change their behaviour after the neighborhood change by excluding any school that is not a neighborhood school anymore and incorporating the new neighborhood schools. While the previous theoretical literature had

[1] A literature explores the impact of school choice on elementary and secondary achievement, high school graduation, and college attendance. Results are mixed, finding positive or non significant effects. Some examples are Rouse (1998), Howell and Peterson (2002), Hoxby and Rockoff (2005), Cullen, Jacob, and Levitt (2006), Hoxby (2003), Hastings, Kane, and Staiger (2008), Gibbons, Machin, and Silva (2008), Hoxby and Murarka (2009), Lavy (2010), Abdulkadiroğlu, Angrist, Dynarski, Kane, and Pathak (2011), Dobbie and Fryer (2011), Hastings, Neilson, and Zimmerman (2012), Neilson (2013), and Demin, Hastings, Kane, and Staiger (2014). Others have studied the effect of school choice on housing prices (see Machin and Salvanes (2010), Bogart and Cromwell (2000), and Ries and Somerville (2010)).

[2] Epple and Romano (2003), Brunner and Imazeki (2008) and Allen, Burgess, and McKenna (2010) explore how choice programs affect sorting in schools. Burgess, Greaves, Vignoles, and Wilson (2009) study if the type of school chosen differs across parents' characteristics.

[3] This was the mechanism used in the city of Boston until 2005. See Agarwal and Somaini (2015) for a description of the mechanisms used in different cities around the world. See Abdulkadiroğlu (2013) for a literature review.

established that parents could be strategic under the BM, this is the first paper that provides evidence that not only parents are strategic by excluding some desired school but that a large fraction of families exclude *all* schools that are not of high priority.[4]

A first contribution of this paper is to highlight the importance of priorities, overcoming large empirical challenges. Namely, (i) preferences are not observable, (ii) families choose what schools to rank as well as where to live, and (iii) some families may be able to opt for a school outside the public system.

One important concern in the debate regarding the BM is that unsophisticated parents, being unable to strategize, may be harmed by the system (see Pathak and Sönmez (2008)). In this paper, we merged our school application data with school enrollment data to provide novel evidence on the equity of the mechanism. Abdulkadiroğlu, Pathak, Roth, and Sönmez (2006) report that in Boston, 19 percent of parents seem to be unsophisticated, playing a dominated strategy. A similar fraction of parents exhibit seemingly unsophisticated behaviour in Barcelona. However, enrollment data allow us to rationalize some of this behaviour. We find that some of these families do not enroll in the school that they were assigned to initially and enroll in a private school that we refer to as outside option. Among unsophisticated families, 13 percent has an outside option. That is, if they are not allocated their first choice, they enroll in their outside option. Not surprisingly, these families have higher socioeconomics while families that are particularly harmed by the mechanism display lower socioeconomics. A second contribution of this paper is that using enrollment data allows us to quantify how many of the unsophisticated families have in fact an outside option.

The rest of the paper is organized as follows. In Section 2 we describe the school allocation mechanism in the city of Barcelona. In Section 3 we discuss the empirical challenges involved in trying to identify what drives parents' school choice and our empirical strategy. In Section 4 we describe our data. In Section 5 we analyze what drives parents' choices. In Section 6 we analyze the rationality behind parent's choices and inequalities of the system. Finally, Section 7 concludes.

# 2 School allocation mechanism in Barcelona

In Spain all children age 3 and above have universal access to a seat in the public system. This implies that even if compulsory primary education starts at the age of 6, de facto almost every child starts school at the age of 3.

In Spain families have the right to choose their children's school (see section 1 of our online appendix for more details on the education system in Spain.) Parents submit their application for the primary school for their children in March of the natural year that their children turn 3.

[4]The lack of incentives to reveal true preferences is referred to *manipulability* of the mechanism (see Pathak and Sönmez (2013)). On the other hand, Abdulkadiroğlu, Che, and Yasuda (2011) show that this *manipulability* can be useful for eliciting cardinal preferences in a way that other mechanisms cannot. Calsamiglia and Miralles (2012) theoretically show that all parents ranking high priority schools is one of the possible Nash equilibria, and it is the unique equilibrium when one of the schools is thought to be *sufficiently* bad for all families. Abdulkadiroğlu, Agarwal, and Pathak (2014) estimate preferences and provide evidence of the benefits from moving from a decentralized to a centralized choice system for secondary schools in New York City. He (2017), Calsamiglia, Fu, and Güell (2014) Agarwal and Somaini (2015) and de Haan, Gautier, Oosterbeek, and van der Klaauw (2015) evaluate the performance of different mechanisms with estimated preferences using data from Beijing (China), Barcelona (Spain), Cambridge (USA), Amsterdam (Netherlands), respectively.

In general, this school will be the one that they attend until they are at least 12 years old and possibly until they finish secondary education.[5]

Since the mid-90s, the mechanism to assign children to schools that is used in Barcelona is the so-called Boston mechanism, first described in Abdulkadiroğlu and Sönmez (2003). It can be described as follows. First, parents submit an application form ranking up to 10 schools in Barcelona (section 3 of our online appendix provides the application forms for different academic years). Once applications with families' rankings have been submitted, all applicants are allocated to the school they have ranked first. If the number of applicants is larger than the number of seats for a certain school, applicants are given points following a scale that depends mainly on the existence of a sibling in the school ranked first, whether the school is in the neighborhood that they live and other characteristics.[6] Those applicants with a higher number of points are accepted and the rest rejected. Ties in number of points are broken through a random lottery. Applicants rejected from the school ranked first opt for a seat in the next school in their submitted ranking *that has a free seat* after the previous round. If there are more applicants than free seats in the school ranked second, seats are allocated again according to the priority points of the first round. The process continues until the submitted lists have been exhausted. Those applicants that may not have been assigned a ranked school, they are assigned to a left-over school which is their closest school that has not been filled up in previous rounds. As can be immediately noticed, the chances of being admitted in any school in the second round are greatly reduced, and more so in further rounds. Waiting lists remain active until the new academic year starts. The ordering of the waiting list is based on the point system mentioned above and described next.

As mentioned above, in case of overdemand for a certain school, the allocation mechanism assigns points to different characteristics of the child, the family, and the school ranked first. Table 1 describes these aspects as well as the scale points in descending order. This point system has remained mostly stable over our period of study, except for some minor changes (Appendix B.1 explains these changes and discusses how our results are not affected by them). We can distinguish between "basic points" and "extra points". Among the basic points, there are points for siblings, neighborhood, and socioeconomic characteristics. Having a sibling in the school ranked first by the applicant is the aspect that gives most points (40). Next are criteria based on the neighborhood of the school ranked first in the application. Neighborhoods are defined by the administration. As mentioned, our empirical strategy will be based on an unexpected change in the definition of such neighborhoods. The neighborhood criteria are mutually exclusive. Living in the neighborhood of the school ranked first gives 30 points. Families have the choice to use the work address of one of the parents/guardians instead of their home. In this case, the applicant would get 20 points if the school ranked first is in the neighborhood of the workplace. After 2007, an applicant that

[5]Parents can move to another school every year, but pre-existing children in the school have priority. Given that school capacity is fixed, assignment at any time later than at age 3 is extremely difficult and requires that some other child leaves the school. Many school include both primary and secondary school, which means that changing schools is complicated before the end of secondary school. Even for schools that only include secondary school, priority is given to children form a given primary school.

[6]Chiu and Weng (2009) and Chiu and Weng (2013) consider a centralized procedure and analyze whether schools would provide preferential treatment, as the BM does, if they had a choice. In our case, as is the case in most centralized procedures, individual schools have no say on the mechanism used. and hence whether preferential treatment is warranted or not.

ranked a school outside the neighborhood but still close enough (within the *administrative district*) would obtain 15 points (more details on these districts will be given in Section 3.1). Families are given 10 points when they do not live in the neighborhood or district of the school, but do live in the city. The next set of points corresponds to the socioeconomic characteristics of the applicant: being a low-income family (10 points) and having a family member with a disability (10 points). The system also gives extra points if the applicant belongs to a large family or has a single parent (15 points) or if the applicant has a chronic digestive illness (10 points).

Table 1: Priority points in case of overdemand for a school, 2005-2010

| | Number of Points |
|---|---|
| **Basic Points** | |
| Applicant has a sibling in school of ranked first | 40 |
| *Neighborhood*([a]) | |
| Residence in neighborhood of school ranked first | 30 |
| One of the parents/guardians works in neighborhood of school ranked first | 20 |
| Residence in the *administrative district* (not neighborhood) of school ranked first([b]) | 15 |
| Residence in the city (not neighborhood) of school ranked first | 10 |
| *Socioeconomic characteristics* | |
| Low-income family([c]) | 10 |
| Applicant, one of the parents/guardians, or a sibling is disabled | 10 |
| **Extra Points**([d]) | |
| Large family (three or more siblings) or single parents([e]) | 15 |
| Applicant has a chronic digestive illness([f]) | 10 |

Notes: ([a]) *Neighborhood* criteria are mutually exclusive: The work address of one of the parents/guardians can be considered *instead* of the residence address of the applicant. *Neighborhoods* of the city of Barcelona are explained in Section 3.1. ([b]) *Administrative district* criteria apply only from 2007 onward. *Administrative districts* of the city of Barcelona are explained in Section 3.1 and Figure 1. ([c]) *Low-income* indicates a family in which one of the parents/guardians receives welfare income or the total family income per capita is less than one-third of it (around 12,000 Euros per year). ([d]) Before 2007, the extra points were automatically added to the basic points. From 2007 on, the extra points were taken into account only to break ties between people with the same number of basic points. ([e]) Single parents have received the same points as large families since 2010. ([f]) Before 2007, points for chronic digestive illness were given only if school ranked first was in the neighborhood.
Source: Resolutions from the Catalan Ministry of Education EDC/712/2005, EDC/449/2006, EDU/904/2007, EDU/349/2008, EDU/553/2009, and EDU/107/2010 (see Section 2 of our online appendix).

The case of Barcelona is particularly useful for our purposes. From Table 1, we note that there are two types of families: those that have socioeconomic characteristics for which they get priority and those that do not. These points are exogenous to the school choice decision, whereas the neighborhood points are endogenous and depend on the location of the school ranked first. In the case of Barcelona, the qualifying socioeconomic characteristics are narrow, and few applicants obtain points from these characteristics. In fact, 90 percent of families do not receive any points for their socioeconomic characteristics. This is particularly useful for our goal because the huge majority of families, at least with respect to the priority points, will be equally constrained in their choice problem. However, some families have the outside option of a private school, and we will be able to highlight important inequalities that these options generate within this mechanism.

Cities worldwide use different versions of the Boston mechanism. These differ in terms of the criteria used to warrant priorities for overdemanded schools, or number of seats that are prioritized in a given school, or in terms of the number of schools children can rank, or even in terms of the

number of seats that are subject to the mechanism. For instance, in Cambridge and Beijing some of the seats are assigned outside the mechanism. In Boston itself, different priorities were defined for seats in a given school (50% were prioritized and the remaining were not). In almost all of them, the number of options that can be submitted is smaller than the total number of schools and the number of options differs across cities. Despite the heterogeneity in the school systems and variations of the BM in different cities, there is an empirical regularity that persists across cities: more than 80% of the assignments are resolved in the first round. As we will discuss Barcelona is no exception (see Table 2).[7] The fact that in the BM the assignment is final in each round makes the choice of the school ranked first of utmost importance. The particularity of the mechanism in Barcelona is that the points for the school in the first round are the points that are used to break ties in any round. However, the additional advantage for having high priority in future rounds is small given the equilibrium played in Barcelona and in cities worldwide.[8]

# 3 Empirical strategy

Families choose their submitted school list according to their preferences and admission probabilities (which are determined by both priority points and other families' behaviour). As explained in Section 2, priority points are mainly given by residence in the neighborhood of the school ranked as well as socioeconomic characteristics of the families.

We start documenting a stylized fact for the city of Barcelona. Table 2 displays for years before and after the change of neighborhood definition the percentage of children that were assigned the school that they ranked first in their school application, those that were assigned the school that they ranked second, and so on, for families without siblings which will be the main focus of our analysis. This table shows that on average, as many as 86 percent of children were assigned to the school they ranked first.[9] Also worth noting is that 3.3 percent of children, on average, were assigned to a non-ranked school. This shows again that under the BM, the school ranked first is crucial.

One might think that the fact that so many children get allocated to the school they rank first indicates that the school system accommodates parents' preferences nicely. However, Abdulkadiroğlu and Sönmez (2003) have shown that under the BM, parents may not have incentives to provide their true preferences. Further experimental evidence, such as Chen and Sönmez (2006) and Calsamiglia, Haeringer, and Klijn (2010), suggests that priority points for schools affect the ranking of schools.[10] Priorities impose a discontinuity in the admission probabilities of the different schools for a given family. Hence, parents may exclude from their submitted list any school that

[7]For instance, in Charlotte, 83% of assignments for grade K2, 94% of assignments for grade 6, and 89% of assignments for grade 9 were resolved in the first round (see Hastings, Kane, and Staiger (2008)). In Tel Aviv, the percentage is 90% (see Lavy (2010)). In Boston, it's 80%, see Abdulkadiroğlu, Pathak, Roth, and Sönmez (2006). In Amsterdam, it's 95% (see de Haan, Gautier, Oosterbeek, and van der Klaauw (2015)).

[8]In Barcelona, 44% of schools get filled up in the first round and 40% are never filled up. Hence, points in the subsequent rounds are only relevant for a very small fraction of assignments.

[9]This number goes up to 91 percent if those with a sibling in the school are included in the baseline population. Those with a sibling are always assigned to their first choice (the sibling school).

[10]These studies observe that subjects in the lab manipulate the submitted rankings by reducing the position of the neighborhood schools with respect to the true position of the school in the preference ranking.

Table 2: Parental behaviour and school allocation in Barcelona

| Percentages | 2005-2006 | 2007-2010 |
|---|---|---|
| Assigned in school ranked 1st | 86.11 | 86.36 |
| Assigned in school ranked 2nd | 5.50 | 6.05 |
| Assigned in school ranked 3rd-10th | 4.30 | 4.70 |
| Assigned to a non-ranked school | 4.10 | 2.89 |
| School ranked 1st is in residence neighborhood | 79.85 | 82.93 |
| Observations | 12,963 | 26,575 |

Source: Barcelona primary school applications data set matched with school enrollment and population census data, *Consorci d'Educació de Barcelona* and IDESCAT.

is preferred to a school for which they get priority points.[11] But learning how parents choose by only observing their submitted rankings is problematic. The following piece of evidence further illustrates this empirical challenge.

The last line of Table 2 displays the percentage of children who rank first a school that is in their residence neighborhood. On average across the years, this number is as much as 82 percent. Our main objective in this paper is to understand to what extent families are ranking the neighborhood school because of their implied priority points. When answering this question we also need to take into account that families can move their residence to the neighborhood of their most preferred school.

We circumvent these identification problems by exploiting an unexpected change in the definition of *neighborhood* that occurred in the city of Barcelona in 2007. This change had two consequences. The first implication is that for families that applied for schools in 2007, it was almost impossible to move residence, thus essentially allowing us to separate housing and schooling decisions. The change was announced on March 27, 2007, and families ranking a school that year were told on March 30 that they had to submit their applications between April 10 and 20, 2007.[12] Moreover, in order to get the corresponding residential points, applicants need to prove to have resided in the claimed address for at least six months prior to the application time. In other words, in 2007, the cases in which the choice of residence and school coincides should be minimal. Since moving residence can take time, most likely this was also the case (even if to a lesser extent) for some years after 2007.

The second implication of the change in the definition of *neighborhood* is that for each family, the set of schools with the highest priority also changed. Comparing school choices before and after 2007 we identify the impact that priorities have on choice. We concentrate our analysis on those families that live in Barcelona (for whom the neighborhood changed) and that do not have a sibling in the school that they have ranked first. The latter approximates the families that are submitting an application for the first sibling because the choice for a second sibling involves a

[11]Note that this is different from saying that parents exclude their top choices, say, the best schools in the city, where everyone wants to go.

[12]The change of neighborhoods was made public through the regional edict 4850/2007, and the application deadline was made public through the regional resolution EDU/904/2007 (see Section 2 of our online appendix).

different problem.[13]

Note that the school application for the first sibling at age 3 is submitted only *once* by each family. Thus, we necessarily need to compare different families over time. The change in neighborhood definition implies that some schools become priority schools and others are not priority schools any more. Identification would be harmed if all families changed their preferences towards the new priority schools against the old priority schools. Nothing changed in the schools in that short period of time, but the peers could have changed. We checked in our data that the new priority schools did not improve their average peer socioeconomic characteristics and that the old priority schools did not worsen their average peer socioeconomic characteristics. Hence we can safely state that peer composition will not affect our identification.

Another consideration relates to whether families coordinate among each other. The school choice literature has modeled individual choice abstracting from interdependent preferences that may arise, for example, from neighbors wanting to coordinate where they send their children to school to facilitate logistics and building additional social capital. Our empirical strategy does not take a stand on the nature of the decision making problem, whether it is individual or collective. The reason is that the change in priorities has a similar impact on every family living nearby. Hence, we are not able to exclude coordination among families, but cannot prove it either. If families decide independently, priorities affect their behaviour by affecting admission probabilities. Similarly, if the decision making is collective, priorities affect probability of admissions for each of them in the same way since priorities are mostly defined by residence. Priorities can equally affect the school that families individually rank first or the school that families coordinate to. Hence, when we observe individuals change their behaviour we do not know whether this results from the priorities affecting each of the individuals independently or as part of a group. In sum, we establish that priorities shape the ranking of schools submitted by families.

## 3.1 Change in the definition of *neighborhood*

As explained in Section 2, in the city of Barcelona, to resolve overdemands for schools, families are given priority points for all schools within their residence neighborhood. The definition of what constitutes such a neighborhood changed in March 2007. Before 2007, the city was divided into fixed neighborhoods. The neighborhoods varied in size for semi-public and public schools (the two types of schools whose seats are allocated through the analyzed centralized system), but were conceptually the same. For semi-public schools, the neighborhood coincided with the administrative district.[14] Figure 1 displays the 10 administrative districts in the city of Barcelona. For public schools, the neighborhoods were smaller areas within the administrative district (see Appendix 4 of our online appendix for more details).

In March 2007, new neighborhoods were defined, moving away from the old fixed administrative

[13]The reasons are that additional priority points are given for children with older siblings in a school (see Table 1), and it is reasonable to assume that families prefer that siblings attend the same school.

[14]There was the possibility for semi-public schools to request that their neighborhood would consist of two adjacent administrative districts. Unfortunately, however, the body that granted these petitions no longer exists, and there is no record of these requests. Please note that this implies that the share of applicants ranking a school that we identify as their neighborhood is a lower bound for the years 2005 and 2006.

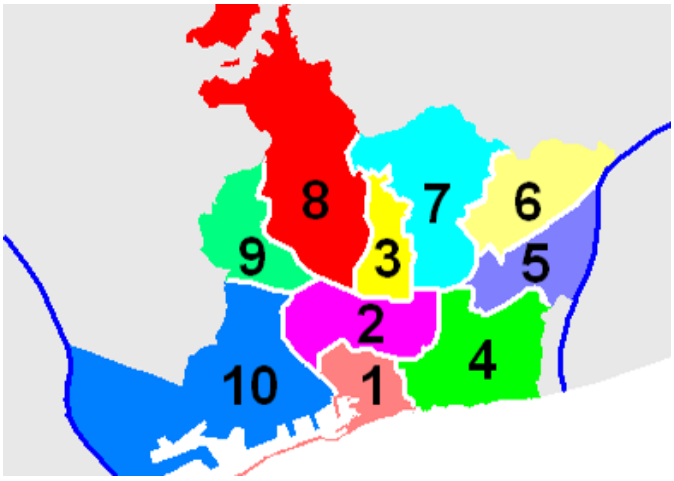


Figure 1: Administrative districts in the city of Barcelona

districts. The change in the definition of *neighborhood* occurred because the previous definition (i.e., a *fixed* area) had two issues from the administration's point of view. First, the administration wanted to unify the definition of *neighborhood* for public and semi-public schools. Second, the old neighborhoods did not capture well the distance to schools (e.g., families living close to the border of two neighborhoods did not have priority points for nearby schools while having priority points for schools farther away within the neighborhood). Finally, the old neighborhoods implied that the size of the set of schools for which a family could have priority could be very different because it depended on the density of schools in the neighborhoods.

One could interpret that the change in neighborhood definition tried to capture parents' preferences to go to closer schools. Even if this were the case, this would not be a problem for our empirical strategy. The reason is that distance to any school is constant before and after the change in neighborhood definition. The only thing that changes are the priority points given for different schools, which allows to identify if families react to this.

The new neighborhoods are defined by *distance* between families' residence and schools. An area (specifically, a minimum convex polygon) around every block of houses in the city was established to include at least the closest six schools (three public and three semi-public). An algorithm was constructed to define the new neighborhoods citywide, which amounted to 5,300 for the whole city. Given this, it is not useful to draw the map of the city with all the new neighborhoods, but rather to illustrate how it changes for a given family as in Figure 2.[15]

We can map every block of houses in the city to its neighborhood and thus to the schools for which that address would have priority points if ranked first. Next, we explain how the change in the definition of *neighborhood* changed this set.

[15]Before 2007, families knew which schools were in their neighborhood by looking at maps like the one in Online Appendix Figure 1, which were available in all schools. Since 2007, since every block in the city has a different neighborhood, the administration would send a letter to each family specifying which schools were in their neighborhood. One might think that this would increase the number of families ranking schools in the neighborhood first, but as Table 2 shows, this was not the case. The number of families that ranked a neighborhood school first is constant over time.

## 3.2 Changes in the set of neighborhood schools

The change in the definition of *neighborhood* implied two changes in terms of the set of schools for which each family had priority points. These changes are the basis of our empirical strategy. For simplicity let us refer to the neighborhoods before 2007 as the *old* neighborhoods and the neighborhoods defined in 2007 as the *new* neighborhoods.

The two changes in the set of priority schools are as follows. First, the new neighborhoods were defined to include the closest 6 schools (3 public and 3 semi-public). This implied that the *size* of the set of neighborhood schools for which families have priority was reduced. Before 2007, families had priority points on average in 22 schools, whereas, after 2007, the number of priority schools was reduced to 8 on average. Also, the standard deviation went from 7 to 2 (see Appendix B.2 for details). So, the size of the set of priority schools indeed became smaller as well as more equal across families in the city. Second, for some families (those living near the border of an old neighborhood), new schools were added in their priority set because the new neighborhood included nearby schools that are on the other side of the border of the old neighborhood. On average, around 83 percent of the families had at least a new school added in their priority set. For these families, this change involved the inclusion of an average of 31 percent of new schools in their set (see Appendix B.2 for details). As will become apparent shortly, it will be relevant to distinguish between: (i) families that live in the *centre* of the old neighborhood and (ii) families that live in the *corner* of the old neighborhood.

In order to illustrate these changes, we classify schools for each family, given their address of residence, as follows:

- Yes-Yes Schools (YY hereafter): in the old neighborhood and in the new neighborhood.
- Yes-No Schools (YN hereafter): in the old neighborhood but not in the new neighborhood.
- No-Yes Schools (NY hereafter): not in the old neighborhood but in the new neighborhood.
- No-No Schools (NN hereafter): not the old neighborhood and not in the new neighborhood.

Figure 2 illustrates the change in the set of priority schools for different families (or different addresses). Diamonds, pentagons, triangles, and hexagons in these graphs represent schools. The (orange) building is a particular family/address. The (green) square represents the old neighborhood, and the (purple) circle refers to the new neighborhood for a family that lives in the *centre* of the old neighborhood (left panel of Figure 2) and for a family that lives in the *corner* of the old neighborhood (right panel of Figure 2).

For families living in the center of the old district, schools inside the circle are YY schools; that is, these schools were also in the old neighborhood (square). Schools outside the circle but inside the square are YN schools; they used to be in the old neighborhood but are not in the new neighborhood. For families living in the corner of the old neighborhood, there are also NY schools, which were not inside the old neighborhood but are inside the new neighborhood. Finally, schools outside the square are NN schools; they were not in the old neighborhood and are not in the new

neighborhood either.[16]

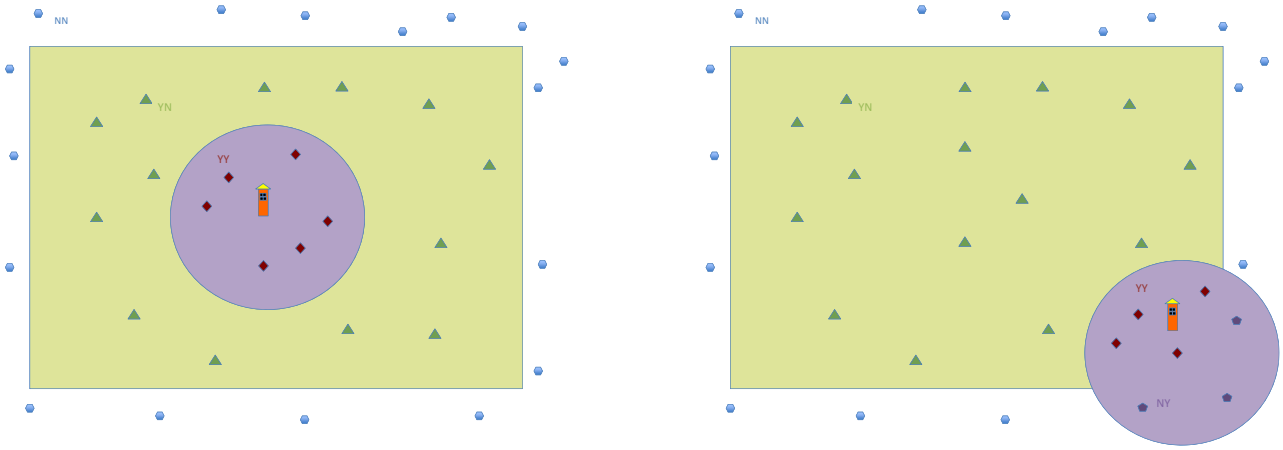


Figure 2: Change in the school priority set for an address
in the center (left) and in the corner (right) of the old neighborhood

Note: Diamonds are YY schools, triangles are YN schools, hexagons are NN schools and pentagons are NY schools for the (orange) building.

The reduction in the size of the set of schools for which families have priority allows us to identify YN schools as well as YY schools for any address in the city (see Figure 2). The inclusion of new schools and the exclusion of old schools in the set of schools for which families have priority allows us to identify NY schools for those addresses that are sufficiently far away from the center of the old neighborhood (see right panel in Figure 2). Overall, these changes of the priority set allow us to identify the different types of schools for each block of houses.

Our key insight is that we can learn the role of priorities in driving families' choices by looking at the changes in the rankings for these different types of schools before and after the change of the definition of *neighborhood*. We explain our strategy in detail in the next subsection.

## 3.3 Testable hypotheses

As seen in Table 2, a large majority of families rank a school in their neighborhood. But from this fact alone we cannot infer what drives parents' choices. The objective of our analysis is to examine the impact of the change in the neighborhood definition in 2007 on parents' behaviour. As already argued, for parents who submit a school application in 2007, it is very unlikely that they were able to change residence.

As mentioned before, it could be the case that the new (old) set of high priority schools are systematically better (worse) schools in terms of their peers. We compare socioeconomic characteristics of the set of individuals that have high priority for a given school before and after the neighborhood change and find that average peers did not improve in NY schools and did not worsen in YN schools. Hence, we can isolate the effect of priorities on parents' school choice.

[16]In terms of priority points, families ranking their YY schools received 30 priority points both before and after 2007; families ranking their YN schools received 30 priority points before 2007 and 15 points after 2007; families ranking their NY schools received 10 priority points before 2007 and 30 points after 2007; and families ranking their NN schools received 10 priority points before 2007 and 10 points after 2007 (except for public schools in the district for which the change would be from 10 to 15; see Appendix B.1 for a further discussion).

Our analysis focuses on the school ranked first and whether it is a NY, YN, YY or NN. If priorities play no role in parents' behaviour, then we should observe no change in choices after the change in the neighborhood definition. Instead, if on the other extreme, families rank high priority schools, then we should observe that their choices change. In particular, we should observe the following changes:

**Changes over time**: *If families mostly rank high priority schools*, the proportion of families ranking the different types of schools should only change between the years 2006 and 2007, when the neighborhoods were changed. And we should observe no changes between the years 2005 and 2006, nor between the years 2007, 2008, 2009, and 2010. But *if priorities do not matter for school choice*, the proportion of families ranking the different types of schools should not change at any point in time.

**Changes in YN and NY schools between 2006 and 2007**: *If families rank mostly high priority schools, the proportion of families that rank first*:

1. YN schools should decrease, given that these schools were high priority before but are not anymore after the neighborhood change;

2. NY schools should increase, given that these schools were not high priority before but are so after the neighborhood change;

But *if priorities do not matter for school choice*, we should observe no change for neither YN, nor NY schools.

**Changes in YY and NN schools between 2006 and 2007**: Regarding YY schools, since the set of schools with high priority has been reduced, if families' choices are driven by priorities, then ranking of any given neighborhood school should increase. But increased ranking of YY schools depends on whether families live in the center or the corner of the old neighborhood. In particular, *if families rank mostly high priority schools*, for families living in the center of the old neighborhood, the proportion of families that rank first YY schools should increase, given that there is a smaller set of schools with highest priority after 2007. In other words, families that were mostly ranking high priority schools before 2007 were ranking YY or YN schools but only for YY schools after 2007. Therefore, if they continue to rank high priority schools, the number of applicants for YN schools would be shifted to YY schools. But *if priorities do not matter for school choice*, we should observe no change.

Note that no prediction can be made for YY schools for families living in the corner of the old neighborhoods. Two forces are at play here. On the one hand, as above, the proportion of families ranking YY schools should increase as the number of applicants for YN schools is being shifted to YY schools. On the other hand, for families living in the corner of the old district, within their set of high priority schools, they will be comparing whether the new schools in their set (NY) are preferred to the remaining schools (YY), which would tend to decrease the proportion of families ranking YY schools. The final effect is ex-ante unpredictable.

Similarly, there is no prediction for the change in the proportion of families ranking NN schools because it is unclear how the aggregate number of applicants for those schools should change. Before

2007 the set of families that had high priority for that school was larger (old neighborhoods were larger), but it was also true that the set of schools for which each applicant had highest priority was larger. Therefore, it is not clear whether the number of applicants from the neighborhood would be larger or smaller, and therefore unclear what the chances of being admitted would be for an individual outside the neighborhood.

# 4 Data

We use a unique and rich administrative data set that consists of the union of the following four data sets containing detailed information on: the primary school application of each family in Barcelona, the supply of schools in Barcelona, and the enrollment for each child and school in Barcelona for the years 2005-2010. School application and allocation data sets were obtained from the *Consorci d'Educació de Barcelona* (CEB), which is the institution that deals with management of public education in Barcelona. Additionally, the school application microdata were merged with the population census data to obtain socioeconomic characteristics of the parents. We describe these data sets below.

**Barcelona primary school applications data set** This data set consists on the universe of applications for primary schools in Barcelona of children that turned 3 in the years 2005-2010. For all children, we have a complete application form that consists of the following variables: their ranking of schools (up to 10 schools) and the priority points obtained for their different characteristics (as described in Table 1: whether they are a low-family income, presence of a sibling in the school ranked first, whether they suffer any chronic illness or disability, etc.). We also know where they live, which allows us to assign them the set of schools for which they would have priority points if they were to rank them first. Finally, we also know the school to which each applicant was assigned by the administration if this is a school in Barcelona.

**Barcelona supply of schools data set** Spain has three types of schools: public, semi-public, and private. The system described in Section 2 is for all public and all semi-public schools. That is, the application and allocation processes include both public and semi-public schools. Semi-public schools are publicly funded but privately managed. By law, no fees are involved in semi-public schools.[17]

Our second database includes all existing schools in the region of Catalonia, their address, and their characteristics (whether they are public, semi-public, or private, and the grades they offer: preschool only, preschool and primary school, etc.). Among the schools in the city of Barcelona,

[17]Semi-public schools are the so-called *concertadas*. Although fees are not compulsory, we know that in reality, families are asked to pay for other services that most families pay for. Families' preferences for a school may be shaped by this fact. However, this is not an issue in this paper because (i) our aim is to understand the role of priorities in school choice (see Section 3) and (ii) the structure of these payments did not change before and after 2007, which is key because our identification relies on the change in the neighborhood definition that occurred in 2007 (see Section 3). One may think that semi-public schools could be a safer option for those families that can afford the voluntary fees. But as we show in Table B3 in Appendix B.3, in any given year, a larger number of public schools are underdemanded than for semi-public schools.

around 48 percent are public, 48 percent are semi-public, and 4 percent are private. Hence, our paper studies the allocation of children in 96 percent of the schools in the city. We refer to these schools as those in the Barcelona public system (i.e., public and semi-public primary schools in Barcelona).

We merge this school data set with the application data set to incorporate school's characteristics into the analysis. In any given year, all available public and semi-public schools are ranked by families. For our baseline population, for the whole period of study, 368 schools are ranked, 337 of these schools are in the city of Barcelona.

**Barcelona school enrollment data set** The third data set that we exploit is the enrollment data set for all schools in the public system in Barcelona. We merge these data with the application data set in order to understand enrollment beyond assignment. This allows us to verify whether families enroll in their assigned school or not as well as to identify families who, failing to be assigned to their desired school, enroll their children in some other school.

**Population census: Parents' socioeconomic characteristics** Under the institutional agreement between the *Consorci d'Educació de Barcelona*, the Barcelona Graduate School of Economics, and the Catalan Institute of Statistics, the school application data set was merged with the Catalan population census data (more information on the census can be found at `http://www.idescat.cat/en/poblacio/censos2011/`) and the municipal population register (more information on the register can be found at `http://www.idescat.cat/en/poblacio/padro.html`). More specifically, each of the applicants was looked for in the population census or the municipal register in order to identify his or her father and mother and obtain their socioeconomic characteristics (see Appendix A.3 for details). The resulting data allow us to study how socioeconomic characteristics can relate to behaviour in the context of school choice (see Section 6).

**Baseline Population** Demand for public primary schools in Barcelona is determined by all children that apply to a school in Barcelona. The application data will be used to compute the demand for each school.[18] Our analysis aims to understand how families' behaviour changes with the change of neighborhood definition. We concentrate on the school ranked first. For this purpose, our population of interest consists on those children that live in Barcelona and apply to schools in Barcelona and that have no sibling in the school ranked first.[19] That is, we want to analyze those families for whom the definition of *neighborhood* changed in 2007. We exclude from the behaviour analysis those children with older siblings as for these families both the preferences for the school and the incentives provided by the mechanism are very different and completely independent of the policy change that we are exploiting.

[18]The school application submitted by families needs to be a regular and valid application and thus we exclude late applications and multiple applications as these do not constitute demand at the allocation time (their school assignment is decided after the regular assignment). We also exclude children with special needs as the system assigns before all the other applicants and offer additional seats if needed.

[19]Dur (2011) theoretically considers the school choice problem of the second sibling for families with more than one child.

The administrative data used are very high quality but there are some minor issues that do not allow us to have complete information for all observations in our baseline population. For instance, crucial for our first analysis is to identify the address of the applicants in order to identify how the change in definition in neighborhood affected them.[20] The school applications forms are submitted to the school or the central administration and then manually entered into the centralized system. This process implies that typos may be introduced, which can make our identification difficult. Overall, we are able to identify the address for 91 percent of the observations in our population of interest. In Appendix A.1 we show that the sub-population for which we cannot identify their address is not statically different from that for which we can in terms of the other observable variables. The fact that this sub-population is rather small and that there does not seem to be any non-random selection implies that our analysis based on the sub-population for which we can identify the address is not biased. Another issue is with respect to enrollment information, which contains information on all schools except preschools in Barcelona (i.e., those schools that offer education up to age 6 only).[21] We match the Barcelona primary school applications data set with the school enrollment data set on common variables (see Appendix A.2 for details). We are able to match 99 percent of the observations in our population of interest. Finally, in terms of matching the application data with the population census, it was possible to find at least one of the parents for 98 percent of the children in our population of interest. Table 3 describes these data year by year. We will refer to our population of interest for whom we have complete information as our baseline population (highlighted in bold in Table 3).

Table 3: Barcelona Primary School Applications Data Set, 2005-2010

| Number of applicants | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 |
|---|---|---|---|---|---|---|
| Live and apply in Barcelona & no sibling | | | | | | |
| Complete information | **6,669** | **6,294** | **6,548** | **6,743** | **6,700** | **6,584** |
| Incomplete information | 884 | 1,031 | 951 | 805 | 715 | 737 |
| Live and apply in Barcelona & sibling | 3,730 | 4,340 | 4,634 | 4,786 | 4,794 | 4,807 |
| Others | 1,094 | 839 | 596 | 798 | 1,012 | 1,173 |
| Total Population | 12,377 | 12,504 | 12,931 | 13,366 | 13,346 | 13,301 |

Notes: *Complete information* includes address, enrollment, and socioeconomic information. For 2005, applicants with a sibling in the school ranked first are approximated as described in Appendix A.4. Source: Barcelona primary school applications data set matched with enrollment and population census data, *Consorci d'Educació de Barcelona* and IDESCAT.

[20] Our data does not include the work addresses of the applicants' parents. This means that we cannot tell how the neighborhood changed for those that live outside Barcelona and rank a school in Barcelona for work reasons. Similarly, for those that live in Barcelona we can only tell how their neighborhood changed in terms of their residential address. In Appendix B.4 we show that the fraction that uses their work address is small and fairly constant over time; thus our results should not depend on being able to identify the work addresses of the parents.

[21] This amounts to six centers in Barcelona. For the years 2005-2010, this amounts to not having enrollment information for 504 children in our baseline population of applicants.

# 5 Results: priority points and choice

In this section, we analyze the school that families ranked first in their application form for our baseline population. In particular, we track whether the school ranked first in their application was a school in their neighborhood (i.e., had highest priority) both before and after 2007 (keeping track of the change in neighborhood) and test each of the hypotheses outlined above.

Figure 3 displays the percentage of families ranking YN schools over the years, that is, schools that were located in their old neighborhood but not in their new neighborhood. We observe the following facts.

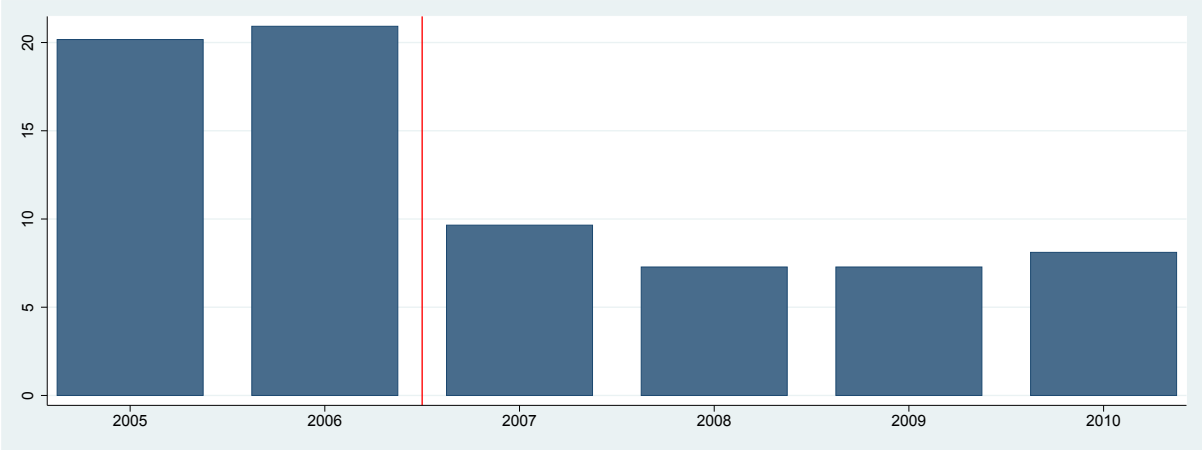


Figure 3: **Share of families over time ranking a YN school first**
Notes: the t-statistic for the difference in the share of families between 2006 and 2007 is 18.25.

First, it can be seen that the only substantial change occurs between 2006 and 2007, coinciding with the change in the neighborhood definition. This provides evidence that the change in the neighborhood definition did change families' school choice, giving a first indication that priorities matter, as outlined by our hypothesis that if this is the case the only change over time should be between 2006 and 2007 and not between any other two years. If priorities did not matter, we should not observe any change. Moreover, as described in Section 3.3, if families mostly rank high priority schools, the change in behaviour should be such that they stop ranking first schools that are no longer in their neighborhood. Consequently, the proportion of families that ask for YN schools should decrease between 2006 and 2007, which is exactly what we observe Figure 3. The proportion of families that ask for YN schools was reduced from 21 percent to 10 percent, which is a decrease in number of applicants of 52 percent (Table B5 in the Appendix displays the yearly figures). This large decrease indeed suggests that priorities play a crucial role in school choice.

We now turn to the analysis of families that rank NY schools, that is, schools that were not located in their old neighborhood but are in their new neighborhood. Figure 4 displays the percentage of families ranking NY schools over the years. In this case, we concentrate on those families in our baseline population for whom the neighborhood change implied not only a change in the size of the set of priority schools but also the inclusion of new school in their priority set. That is, families are close enough to the corner of the old neighborhood so that their new set of priority schools includes at least one school that was not in their set before (see right panel in Figure 2). Recall that families living in the center of the old district only experience a change

in the size of the set of priority schools, and therefore there are no NY schools in their set. We observe the following facts.

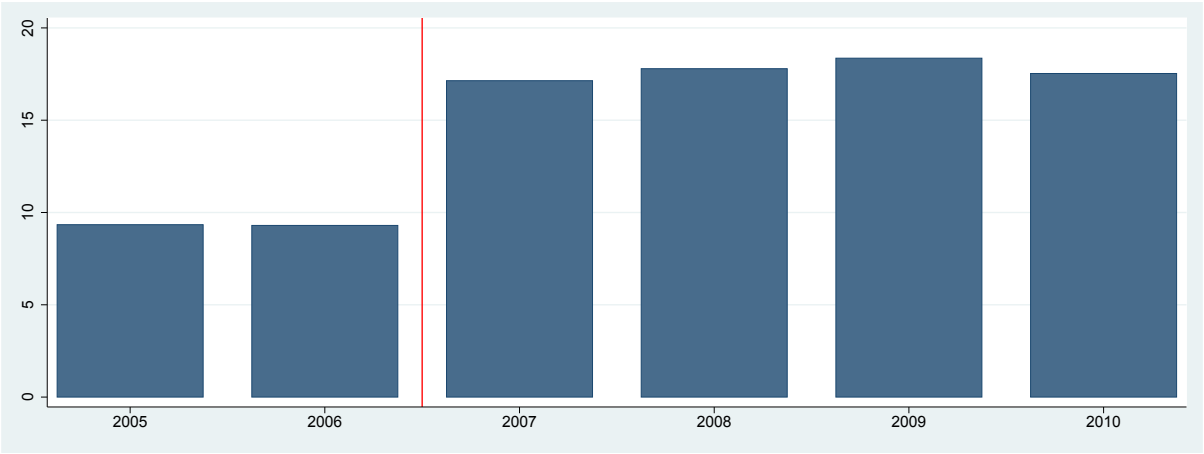


Figure 4: **Share of families over time ranking a NY school first**
Notes: the t-statistic for the difference in the share of families between 2006 and 2007 is -11.89.

Figure 4 shows again that the only substantial change occurs between 2006 and 2007, coinciding with the change in the neighborhood definition, thereby providing further evidence that the change in the neighborhood definition did change families' school choice, giving another indication that priorities play a big role in school choice. Moreover, as described in Section 3.3, if families mostly rank high priority schools, the change in behaviour should be such that they start ranking first schools that are in their new neighborhood, because these schools now give them priority points. Consequently, the proportion of families that ask for NY schools should increase between 2006 and 2007, which is exactly what we observe in Figure 4. The proportion of families that ask for NY schools increased from 9 percent to 17 percent, which is an increase in the number of applications of 89 percent (Table B6 in the Appendix displays the yearly figures). This very large increase indeed suggests that priorities play a crucial role in school choice.

To sum up, we find evidence that a large fraction of families change systematically the school they rank first because of the change in the definition of *neighborhood*. Moreover, it is worth noting that the way the new neighborhoods were designed implies that YN schools are schools that are relatively far away from families' residences (see left panel in Figure 2), whereas NY schools are schools that are relatively close to families' residences (see right panel in Figure 2). This means that in the old neighborhoods, families were going to schools that were quite far away (YN schools) and were not going to schools that were nearby (NY schools) because of lack of priority points for them. Therefore, we find that preference for a shorter distance to schools is not the sole determinant for parents' actual preferences or choice. These results are not incompatible with the fact that within the set of neighborhood schools, families may then rank schools according to distance. However, note that the analysis on the NY schools does not allow us to establish whether families coordinate or not. The figures in Table B6 suggest that either families were not coordinating before 2007 or that they were coordinating but that they have changed where they coordinate to after 2007.

We now turn to the analysis of YY schools, that is, schools that for any given family were located in their old neighborhood and are still located in their new neighborhood. We first analyze

those families that live in the *center* of the old neighborhood for whom the set of priority schools from 2007 is composed only of YY schools. Figure 5 displays the percentage of families ranking YY schools over the years.

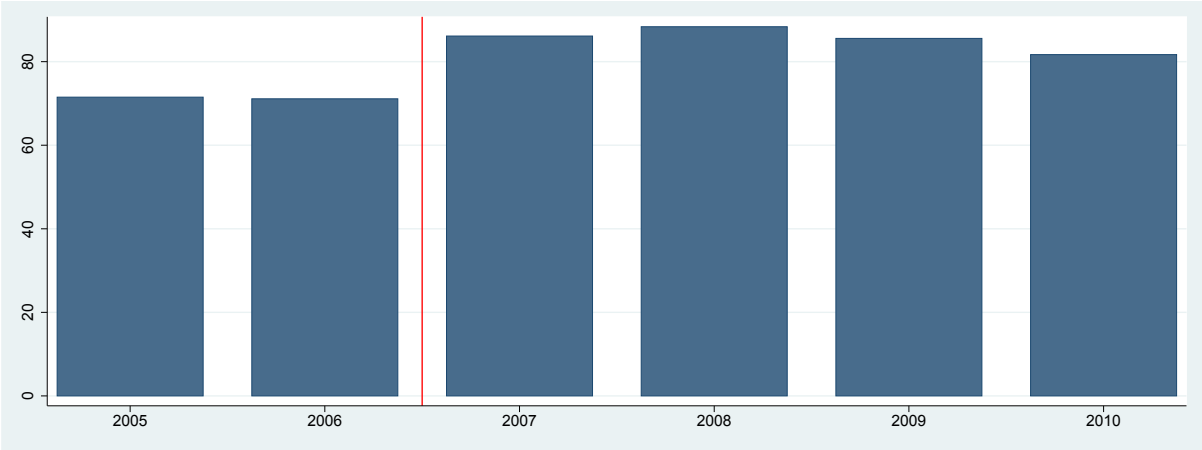


Figure 5: **Share of families over time ranking a YY school first and live in the center of the old neighborhood**

Notes: the t-statistic for the difference in the share of families between 2006 and 2007 is -9.23.

Figure 5 shows again that the main substantial change occurs between 2006 and 2007, coinciding with the change in the neighborhood definition, thereby providing further evidence that the change in the neighborhood definition did change families' school choice, giving another indication that priorities play a big role in school choice. Moreover, as described in Section 3.3, if families mostly rank high priority schools, there should be an increase in the share of families that ask for YY schools between 2006 and 2007. We find that these families are moving away from schools that were in their neighborhood (YN schools) and concentrating on the schools around their residence. (Table B7 in the Appendix displays the yearly figures).

We next turn to the analysis of YY schools for families that live in the *corner* of the old neighborhood. As explained in Section 3.3, there is no unambiguous prediction for this case. Figure 6 displays the percentage of families ranking schools that are YY that live in the corner of the old neighborhood over the years. We find that the number of applicants for YY schools has moderately increased (Table B8 in the Appendix displays the yearly figures). This increase is consistent with a shift in applicants from YN schools to YY schools as well as some substitution between YY schools and NY schools.

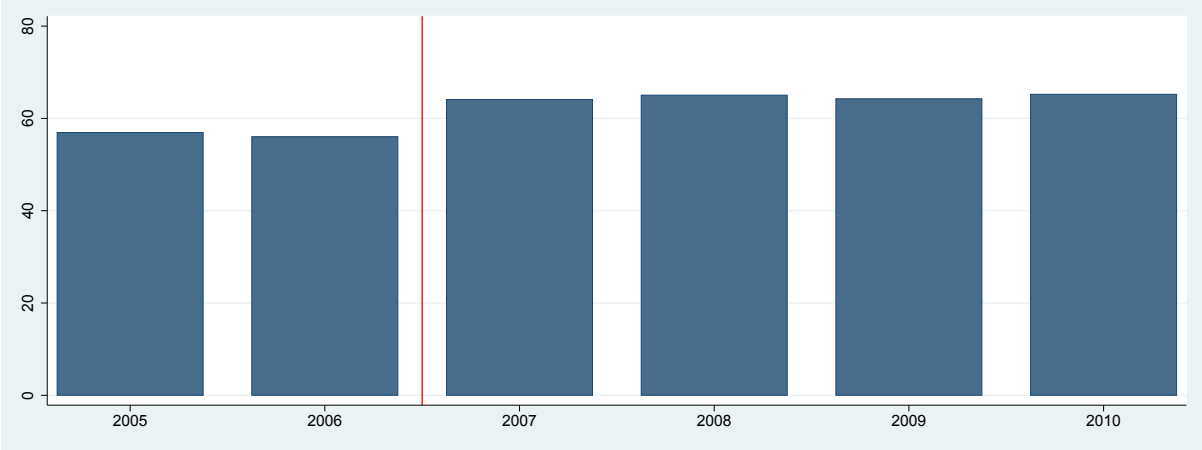


Figure 6: **Share of families over time ranking a YY school first and live in the corner of the old neighborhood**
Notes: the t-statistic for the difference in the share of families between 2006 and 2007 is -8.41.

Finally, we turn to the analysis of NN schools. Again, as explained in Section 3.3, there is no unambiguous prediction for this case. Figure 7 displays the percentage of families ranking schools that are NN over the years. We find that the number of applicants for NN schools has decreased after 2006 (Table B9 in the Appendix displays the yearly figures).

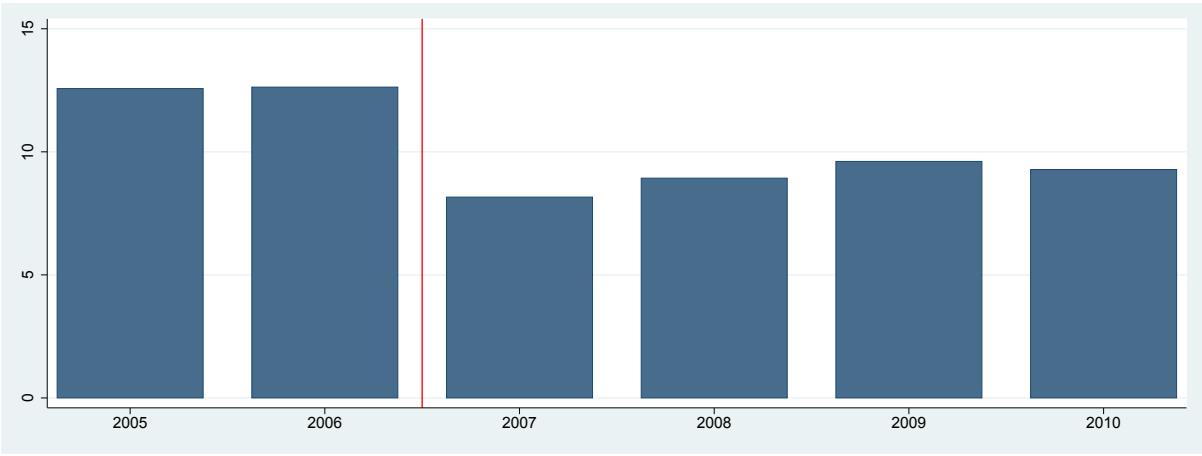


Figure 7: **Share of families over time ranking a NN school first**
Notes: the t-statistic for the difference in the share of families between 2006 and 2007 is 8.30.

# 6 Results: further understanding behaviour

In this section we pool the 2005-2010 data together to better understand parents' behaviour through analyzing the consequences of their actions, that is, their final school enrollment.[22] In Subsection 6.1, we first identify those applicants whose behaviour could be considered unsophisticated based on the definition in Abdulkadiroğlu, Pathak, Roth, and Sönmez (2006), which requires that the probability of being assigned to the school ranked first is less than one, and the probability of being assigned to the school ranked second is zero. That is families that rank first a school that

[22]The change in neighborhoods does not have implications for the willingness for families to take risk. Accordingly, we did not find in our data any significant differences in risk-taking behaviour before and after 2007.

is overdemanded (i.e. there is some risk of not getting in) and second a school that was already filled up in the first round. In Barcelona, applications are submitted in schools and information about previous and current number of applicants is available.[23] We also analyze those families that rank only one (overdemanded) school, leaving the rest of the list blank. In Subsection 6.2, we incorporate information about the school where children enroll (which may be different from the assigned school) to understand further the consequences of such behaviour and to learn about the equity properties of this mechanism. Finally, in Subsection 6.3, we analyze the socioeconomic characteristics of these families in order to reevaluate whether, as the literature has emphasized, disadvantaged families may be less able to exercise choice.

## 6.1 Unsophisticated families and families that rank only one school

We start by identifying families that rank first a school that their probability of assignment is less than one. We compute the chances of being assigned the ranked school given the priority points that families have for the ranked school. This will depend on how many other families (and with how many priority points) rank each school.

Let $MinPoints_{st}$ be the minimum amount of priority points necessary to be assigned to a school $s$ in year $t$ in the first round. The probability of assignment of applicant $i$ that has ranked first school $s$ in year $t$ can be defined as follows:

$$ProbAssignement_{is1t} = \begin{cases} 1 \text{ if } points_{it} > MinPoints_{st}, \\ \dfrac{Capacity_{st} - Applicants_{s1tp+}}{Applicants_{s1tp}} \text{ if } points_{it} = MinPoints_{st}, \\ 0 \text{ if } points_{it} < MinPoints_{st}. \end{cases}$$

where $Capacity_{st}$ is the number of seats available in school $s$ in year $t$, $Applicants_{s1tp}$ is the number of applicants ranking first school $s$ in year $t$ with priority points $p$, $Applicants_{s1tp+}$ is the number of applicants ranking first school $s$ in year $t$ with *more* priority points than $p$, and $points_{it}$ is the number of priority points that applicant $i$ in year $t$ has for the school ranked first.[24] (see Appendix B.3 for details on how we calculate these with our data).[25]

That is, the variable $ProbAssignement_{is1t}$ takes value 1 when an applicant has more priority points than the minimum necessary to be assigned to the school ranked first. It takes value 0 when an applicant has fewer priority points than the minimum necessary to be assigned to a school. For applicants with the same number of priority points than the minimum necessary to be assigned to a school, then $ProbAssignement_{is1t}$ is the total capacity left after considering applicants that

[23] This definition tries to identify parents who are clearly not playing the optimal strategy and can be checked with school application data. More recent structural estimations (such as He (2017), Calsamiglia, Fu, and Güell (2014) and Agarwal and Somaini (2015)) have refined the definition of unsophisticated because the structural estimates of the model allows to identify how true preferences deviate from the reported preferences.

[24] Recall that for the case of Barcelona, all ranked schools have the priority points of the school ranked first.

[25] An example will be useful: school $s$ in year $t$ will have $MinPoints_{st} = 25$ if: (i) the number of children that have more than 25 points that rank the school is smaller than the capacity of the school, that is $Applicants_{s1t25+} < Capacity_{st}$ and (ii) the number of children that have exactly 25 points that rank the school is greater than or equal to the capacity of the school, that is, $Applicants_{s1t25} > Capacity_{st}$.

have more points than the minimum necessary ($Capacity_{st} - Applicants_{s1tp+}$) divided by the total number of competitors (i.e., who rank the same school with the same number of priority points, $Applicants_{s1tp}$).

Table 4 reports the probability of assignment for the schools ranked first calculated for our baseline population.[26] We observe the following. First, for 57 percent of the families, their school choice entails no risk. Second, 41 percent of the families entail some risk with the school they rank first. This means that they are ranking schools that are overdemanded by other families with the same number of priority points. This number may seem high at first. But among these families, 95 percent have 30 points or more (and 96% have obtained these points by ranking a neighborhood school). So, interestingly, most of the risk entailed by families comes from ranking their neighborhood schools, in an equilibrium in which most families rank high priority schools. Third, 2 percent of families are ranking a school in which they have no chance of being assigned to. Their choice *without further information* cannot be easily rationalized.

Table 4: Probability of being assigned to the school ranked first

| $ProbAssignement_{is1t}$ | Proportion of Families (%) |
|---|---|
| 1 | 56.75 |
| $(0, 1)$ | 41.30 |
| 0 | 1.96 |
| Observations | 39,538 |

Source: Barcelona primary school applications data set matched with enrollment and population census data, 2005-2010.

Among families that rank a school such that $ProbAssignement_{is1t} < 1$, we further distinguish among three groups: (i) those whose behaviour could be considered unsophisticated (because the probability of being assigned to the school they have ranked second is zero); (ii) those families that only rank one school, leaving the rest of the list blank and (iii) the remaining families are those that take some risk in the school ranked first but include other schools in the list that may be alternative options if they are not allocated to their first.

The first group, the unsophisticated families, are those who rank second a school that was already full in the first round and rank first a school that does not guarantee assignment. That is, unsophisticated families clearly play suboptimally by ranking second a school with zero probability of assignment, which implies that they will miss an opportunity into some school if they are not assigned to the school ranked first. We find that 21 percent of the families seem to be unsophisticated, at least *without further information*.[27] This magnitude is in line with the 19 percent found in Boston by Abdulkadiroğlu, Pathak, Roth, and Sönmez (2006). In the next section, we reevaluate the behaviour of these unsophisticated families using school enrollment information.

The second group includes the families that rank only one (overdemanded) school and have left the rest of the list blank. We find that 3 percent of families behave this way. As we will explore in the next section, further analysis of their behaviour is worth pursuing to understand the role of

[26] If we included in the sample those applicants with a sibling in the school, these proportions would be 73.00, 25.78 and 1.22, respectively.

[27] If we included in the sample those applicants with a sibling in the school, this magnitude would be 13 percent.

outside options.

In the next section we concentrate on the families that rank first a school with some risk. Among this population, the families that are unsophisticated represent the 48 percent and the families that rank only one (overdemanded) school represent the 6 percent.

## 6.2 Rationalizing risky behaviour using enrollment data

The aim of this section is to understand what fraction of the families taking risks have an outside option. Calsamiglia, Martinez-Mora, and Miralles (2017) show theoretically that parents with the option of private school can take higher risks and rank the best schools in the public system, appearing to be unsophisticated if the private option is ignored. Key for an outside option to have an effect is that equal access is *not* guaranteed for all families. Private schools requiring tuition fees make clear candidates.[28]

Access to any school in the public system in Barcelona or in any city in Catalonia goes through the same centralized mechanism. Our analysis focuses on families who have only ranked schools in Barcelona and hence, have not ranked schools outside of Barcelona. This implies that if they opt out from the school they are assigned to by the centralized system, their only alternative is a private school, as it is the only type of school where the assignment is not centralized. Since we have enrollment data for all schools in the public system in Barcelona, if a child in our population of interest is not enrolled in any of the public schools, he/she would have enrolled in a private school. Therefore, in our setting, outside option identifies private schools.

The challenge of identifying who is using an outside option as a back up plan is that outside options can only observed in the data for those families that end up having to opt for it when they are not assigned to their first choice. We exploit the fact that getting the first choice when this entails risk, that is when the $ProbAssignement_{is1t} \in (0,1)$, depends only of a random lottery draw. Hence among families ranking a first choice that entails risk, those not getting their first choice is a random representation of all families taking such risk. Therefore we infer what families' backup plans are by looking at what those not getting their first option do, which should be a good representation of what any of the families taking risks would have done had they not gotten their first choice. We find that 31 percent of families that their first choice involves risk do not get it. For the rest of this section we focus on the subset of families for whom their school ranked first involves risk and are not assigned to it.

We now proceed to analyze the outcome for those families by considering *enrollment data*. Again, school enrollment information is particularly useful to distinguish among the four and very heterogeneous outcomes: (1) At one extreme, families that do not enroll in *any* of the schools they have ranked in their application (labeled as "Non ranked school"). In between, (2) there are the families that enroll in a school ranked in positions third through tenth in their application (labeled as "A school ranked 3rd-10th") (3) families that enroll in their school ranked second (labeled as "School ranked 2nd"). At the other extreme, (4) there are the families that have an outside option

[28]In the structural models in Calsamiglia, Fu, and Güell (2014) and Agarwal and Somaini (2015), an outside option is included, however, it is not analyzed as the goals of these papers is to analyze the performance of different mechanisms.

and are able to improve their allocation in the public system in Barcelona by going to a private school (labeled as "Outside option").

Table 5 reports the share of families obtaining each of these four possible outcomes. We report these outcomes for the three different groups explained in the previous section. We find that 13 percent of these families enroll in an outside option (5.7 percent considering the overall population). This number goes up to 32 percent for those families who only rank one school.

Table 5: Final enrollment of unlucky risk takers

| | Proportion of Families (%) | | | |
|---|---|---|---|---|
| *Final enrollment* : | Unsophisticated | Rank only one school | Other risky behaviour | All |
| Non-ranked school | 44.91 | 67.69 | 13.71 | 30.49 |
| A school ranked 3rd-10th | 39.82 | - | 11.46 | 23.78 |
| School ranked 2nd | 1.77 | - | 63.93 | 32.61 |
| Outside option | 13.51 | 32.31 | 10.90 | 13.12 |
| Observations | 2,436 | 260 | 2,670 | 5,366 |

Population: Baseline population with $ProbAssignement_{is1t} < 1$ and were not assigned to the school ranked first. Source: Barcelona primary school applications data set matched with enrollment and population census data, 2005-2010.

These results have two important implications. On the one hand it shows that, a substantial fraction of parents use an outside option as as a backup option to their public system application, allowing them to take higher risks and hence to get increased access to their preferred schools. This is quite large in a context where there are few private schools (only 4 percent of the schools in Barcelona are private).

On the other hand, these results allow us to rationalize some of the initially diagnosed as unsophisticated behaviour when only using assignment data. The initial 21 percent of unsophisticated parents gets reduced to 18 percent. Our findings shed new light on the equity properties behind this mechanism.

## 6.3 Risky behaviour and socioeconomic characteristics

The literature has emphasized the fact that disadvantaged families may be less able to exercise choice, which could explain the presence of less rational behaviour (e.g., Hastings, Kane, and Staiger (2008)). We are able to analyze this issue with our data. Our application data set was merged with the census data and municipal register in order to give us information about the education of the parents of the applicants. In particular, we determined the share of highly educated parents in the household (that is, those that completed at least secondary education; see Appendix A.3 for details).

The shares of highly educated parents among those who do not take a risk is 0.73, while among those that do take a risk it is 0.78. The standard error of the difference is 0.004 so the difference is statistically different from zero. That is, families that are unsophisticated are *not* associated with lower education levels than those that are not. This may seem puzzling but as we analyze below, once we consider the fact that families may have an outside option, we will be able to understand this fact better.

We next explore the relationship between families' education and final outcomes. Figure 8 displays the average share of highly educated parents in the household of the applicant by the different outcomes. The average shares are displayed separately for those in the groups defined in Table 5. The horizontal line represents the average share of highly educated parents in the household for those whose choice entailed no risk, i.e., $Risk_{is1t} = 0$. The first striking fact is that there is substantial heterogeneity in terms of education across the different outcomes in each of the three groups. This highlights the importance of looking at final outcomes when analyzing the sophistication behind school choice. Looking at Figure 8, we notice the following. First, those with an outside option have higher education levels than the non-risk takers.[29] These are families that decide to rank their first best in the public system and opt for private school if they are not lucky (thereby stating that they prefer the private school to the schools that have a free seat in the second round). Second, those that end up in a school in their ranking (but not the school ranked first) seem to have similar levels of education as the non-risk takers.[30] Third, among those who are not assigned to any ranked school, there is only one group that has lower levels of education than the non-risk takers. These are those who rank only one school and do not have an outside option.[31] Two rationales are possible. One is that these families did not understand the mechanism. The alternative is that, having come from a worse neighborhood with schools that are underdemanded at any round, they risk ranking a better school and if unlucky are placed in a "leftover" school, which is their neighborhood school.

It is remarkable although not surprising that those with an outside option have the highest levels of education. These facts suggest that the role of the outside option may be a relevant aspect in determining parents' willingness to take risk and thereby increase their chances of getting the best schools in the system. Consistent with our findings, He (2017) shows that families with such outside options have higher than average socioeconomics, and play less cautiously.[32]

Therefore, the B-mechanism raises an equity concern, but of a different type from those concerns suggested by previous literature.

# 7 Conclusions

Abdulkadiroğlu and Sönmez (2003) show theoretically that the Boston mechanism does not provide incentives to families to report their preferences truthfully. Choosing the ranking of schools to submit under the Boston mechanism involves considering many factors such as school quality, distance to school or cardinal preferences. The goal of this paper is to highlight the importance

[29]The difference in means t-statistics for each the three groups and the non-risk takers (in the same order as in Figure 4) are 7.4, 3.7 and 4.0, respectively.

[30]The difference in means t-statistics for each the two groups that rank more than one school and the non-risk takers (in the same order as in Figure 4) that end up in a school ranked 3rd-10th are 2.6 and 1.4, respectively. The equivalent t-statistics for those being assigned to the school ranked second are 5.1 and 1.6, respectively. Note that the latter is significant for the unsophisticated families, who in principle should have zero probability of being assigned to that option. This is a small group of families among the unsophisticated (1.77 percent) and they are exceptional in terms of their socioeconomics and their outcomes.

[31]The difference in means t-statistic for each the three groups and the non-risk takers (in the same order as in Figure 4) are 0.8, -2.1, and -1.1, respectively.

[32]Outside option in this case captures a mix of public schools or seats outside one's district that can be accessed through alternative procedures than the centralized mechanism, such as side payments or entrance exams.

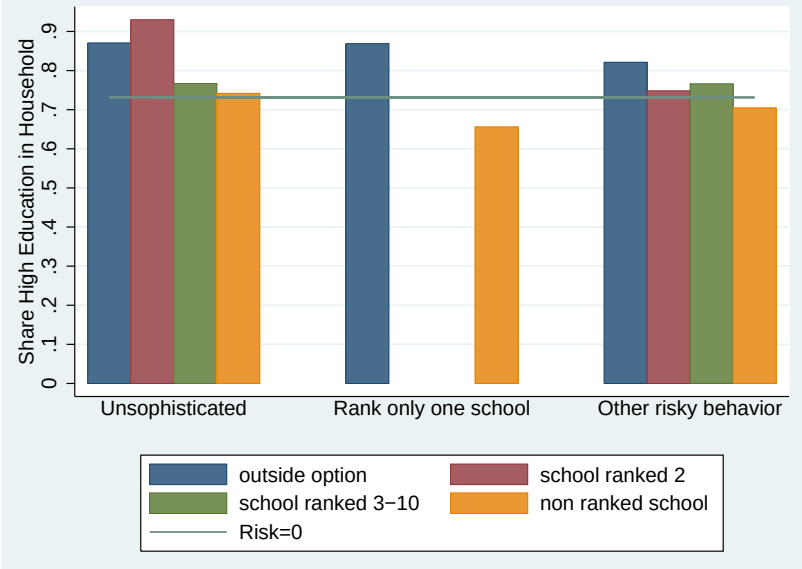


Figure 8: Household Education and Final Outcomes

Notes: Final outcomes as defined in Table 5. High education is defined as completed secondary education or more. Population: Baseline population with $ProbAssignement_{is1t} < 1$ and were not assigned to the school ranked first. Source: Barcelona primary school applications data set matched with enrollment and population census data, 2005-2010.

of another factor in determining the solution to this problem that has been mostly overlooked by the literature. To resolve overdemands for a given school, mechanisms use priority points such as neighborhood, siblings or a lottery number. These priority points may seem innocuous, but we show that they have a drastic effect on parents' behaviour under the Boston Mechanism used in Barcelona.

In a nutshell, we provide clean empirical evidence that a large fraction of families systematically rank schools in their neighborhood. We exploit an unexpected change in the definition of neighborhood in Barcelona. This provides an exogenous change in the set of schools for which families have priority and allows us to analyze the impact that priorities have on school choice, independently of housing decisions. We find that many families change their behaviour after the neighborhood change by excluding any school that is not a neighborhood school anymore (such applications are reduced by 52 percent) and incorporating the new neighborhood schools (such applications increase by 89 percent).

On the other hand, full access to data on final school enrollment allows us to further understand the consequences of unsophisticated behaviour, and clarifies that a substantial fraction of the unsophisticated can be rationalized by their access to an outside option of a private school. We find that, in a school system in which the private sector is rather small (4 percent of schools), 13 percent of applicants to the public school system that rank first an overdemanded school end up using an outside option if they do not get admitted to it. This implies that the initial 21 percent of applicants diagnosed as unsophisticated goes down to 18 percent when using enrollment information.

It is often presumed that implementing school choice will shift the determination of a child's school away from the administration and closer to parents' preferences. This paper shows that

under the Boston Mechanism used in Barcelona, priorities used to break ties have a very large impact on the list submitted by parents unless one has an outside option. Hence, even when school choice is implemented, the school that children are assigned to is greatly determined by the administration. Our findings suggest that policy makers should think further about the role that priorities and private schools ought to play in the final assignment. Similarly, researchers should incorporate the existence of priorities and private schools in their analysis. For instance, Abdulkadiroğlu and Che (2010) suggest that the BM will provide efficiency gains by having parents who value relatively more a given school rank it. In their model, there are neither priorities nor private schools. Our paper shows that both priorities and private schools have a direct impact on behaviour and suggest that the potential efficiency gains of the BM may be reduced.

On the other hand, studies such as Lavy (2010) and Hastings, Kane, and Staiger (2008) exploit the implementation of the BM with neighborhood priorities to provide empirical evidence of the positive impact of choice on outcomes. The underlying assumption in their analysis is that when choice is implemented, families can access a different school or a school they like better. Since the mechanism implemented in these studies is the BM, from the analysis provided in this paper we learn that these results may only be a lower bound on the effects of implementing school choice.

# References


Abdulkadiroğlu, A. (2013). School choice. In *The Handbook of Market Design*, ed. N. Vulkan, A. E. Roth and Z. Neeman. Oxford: Oxford University Press.

Abdulkadiroğlu, A., N. Agarwal, and P. A. Pathak (2014). The welfare effects of congestion in uncoordinated assignment: Evidence from the nyc hs match. Mimeo.

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. *Quarterly Journal of Economics 126*(2), 699–748.

Abdulkadiroğlu, A. and Y.-K. Che (2010). The role of priorities in assigning indivisible objects: A characterization of top trading cycles. Mimeo, Duke University.

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, A. E. Roth, and T. Sönmez (2006). Changing the Boston school choice mechanism: strategy-proofness as equal access. Mimeo, MIT.

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

Agarwal, N. and P. Somaini (2015). Demand analysis using strategic reports: An application to a school choice mechanism. Mimeo.

Allen, R., S. Burgess, and L. McKenna (2010). The early impact of Brighton and Hove's school admission reforms. CMPO Working Paper No. 10/244, University of Bristol.

Bogart, W. T. and B. A. Cromwell (2000). How much is a neighborhood school worth? *Journal of Urban Economics 47*(2), 280–305.

Brunner, E. J. and J. Imazeki (2008). Tiebout choice and universal school vouchers. *Journal of Urban Economics 63*(1), 253–279.

Burgess, S., E. Greaves, A. Vignoles, and D. Wilson (2009). Parental choice of primary school in england: what type of schools do parents choose? CMPO Working Paper No. 09/224, University of Bristol.

Calsamiglia, C., C. Fu, and M. Güell (2014). Structural estimation of a model of school choices: the boston mechanism vs. its alternative. Mimeo.

Calsamiglia, C. and M. Güell (2014). The illusion of school choice: Empirical evidence from barcelona. CEPR working paper No 10011.

Calsamiglia, C., G. Haeringer, and F. Klijn (2010). Constrained school choice: An experimental study. *American Economic Review 100*(4), 1860–1874.

Calsamiglia, C., F. Martinez-Mora, and A. Miralles (2017). Sorting in public school districts under the boston mechanims. Mimeo, Universitat Autònoma de Barcelona.

Calsamiglia, C. and A. Miralles (2012). All about priorities: No school choice under the presence of bad schools. Barcelona GSE Working Paper No 631.

Chen, Y. and T. Sönmez (2006). School choice: An experimental study. *Journal of Economic Theory 127*(1), 202–231.

Chiu, Y. S. and W. Weng (2009). Endogenous preferential treatment in centralized admissions. *The RAND Journal of Economics 40*(2), 258–82.

Chiu, Y. S. and W. Weng (2013). Student enthusiasm and preferential treatment in centralized admissions. Mimeo.

Cullen, J. B., B. A. Jacob, and S. Levitt (2006). The effect of school choice on participants: Evidence from randomized lotteries. *Econometrica 74*(5), 1191–1230.

de Haan, M., P. A. Gautier, H. Oosterbeek, and B. van der Klaauw (2015). The performance of school assignment mechanisms in practice. Mimeo, Vrije Universiteit.

Demin, D. J., J. S. Hastings, T. J. Kane, and D. O. Staiger (2014). School choice, school quality and postsecondary attainment. *American Economic Review 104*(3), 991–1013.

Dobbie, W. and R. G. Fryer (2011). Are high-quality schools enough to increase achievement among the poor? evidence from the Harlem children's zone. *American Economic Journal: Applied Economics 3*(3), 158–187.

Dur, U. (2011). Dynamic school choice problem. Mimeo, North Carolina State University.

Epple, D. N. and R. Romano (2003). Neighborhood schools, choice, and the distribution of educational benefits. In *The Economics of School Choice*, ed. C. M. Hoxby. Chicago: University of Chicago Press.

Gibbons, S., S. Machin, and O. Silva (2008). Choice, competition, and pupil achievement. *Journal of the European Economic Association 6*(4), 912–947.

Hastings, J. S., T. J. Kane, and D. O. Staiger (2008). Heterogeneous preferences and the efficacy of school choice. Combines and replaces NBER Working Papers No. 12145 and No. 11805. Available at `http://www.justinehastings.com/images/downloads/HKS_Combined_201006.pdf`.

Hastings, J. S., C. A. Neilson, and S. D. Zimmerman (2012). The effect of school choice on intrinstic motivation and academic outcomes. NBER Working Paper No. 18324.

He, Y. (2017). Gaming the boston school choice mechanism in Beijing. Working Paper, Rice University.

Howell, W. G. and P. Peterson (2002). *The Education Gap: Vouchers and Urban Schools*. Washington, DC: Brookings Institution Press.

Hoxby, C. M. (2003). School choice and school productivity. Could school choice be a rising tide that lifts all boats? In *The Economics of School Choice*, ed. C. M. Hoxby. Chicago: University of Chicago Press.

Hoxby, C. M. and S. Murarka (2009). Charter schools in New York City: Who enrolls and how they affect their students' achievement. NBER Working Paper No. 14852.

Hoxby, C. M. and J. Rockoff (2005). The impact of charter schools on student achievement. Mimeo, Columbia University.

Lavy, V. (2010). Effects of free choice among public schools. *The Review of Economic Studies 77*, 1164–1191.

Machin, S. and K. G. Salvanes (2010). Valuing school quality via a school choice reform. IZA Discussion Paper No. 4719.

Musset, P. (2012). School choice and equity: current policies in OECD countries and a literature review. OECD Education Working Papers, No. 66, OECD Publishing.

Neilson, C. (2013). Targeted vouchers, competition among schools, and the academic achievement of poor students. Mimeo, Yale Univesity.

Pathak, P. A. and T. Sönmez (2008). Leveling the playing field: sincere and sophisticated players in the Boston mechanism. *American Economic Review 98*(4), 1636–1652.

Pathak, P. A. and T. Sönmez (2013). School admissions reform in Chicago and England: comparing mechanisms by their vulnerability to manipulation. *American Economic Review 103*(1), 80–106.

Ries, J. and T. Somerville (2010). School quality and residential property values: evidence from Vancouver rezoning. *Review of Economics and Statistics 92*(4), 928–944.

Rouse, C. E. (1998). Private school vouchers and student achievement: an evaluation of the Milwaukee parental choice program. *Quarterly Journal of Economics 103*(2), 553–602.

# A Appendix: Further data details

## A.1 Address identification in applications data set

As explained in Section 4, in order to assign to each applicant his/her neighborhood in terms of priority schools, we need to identify their addresses from their application forms. Specifically, we need to match the applicant's address to a database that contains all addresses in Barcelona with their block. In turn, there is a database that relates the schools in the neighborhood for every block in the city.[33] However, addresses in the school applications can have typos. We were able to identify around 89 percent of the addresses in our population of interest. In this appendix we analyze whether there is any selection issue around the address identification.

Table A1 reports the differences in all available observable variables between the sub-population for which we have been able to identify their address as well as for the sub-population for which we have not been able to identify their address. As can be seen, in general, the differences are not significant at conventional levels. The exception is the share of low-income families, which is somehow larger for the sub-population for which we have not been able to identify their address. However, it is worth noting that in general, the share of low-income families is extremely small (due to the extreme definition of low income; see Table 1). The share of families that are *not* low income is 99.72 percent and 99.47 percent, respectively, for these two sub-populations. Therefore, in general we are not concerned about any selection issues related to the identification of the families' address.

Table A1: Differences among population with and without identified address

| | All | Identified Address | Not Identified Address | P-value |
|---|---|---|---|---|
| % Low income family | 0.30 | 0.28 | 0.52 | 0.00 |
| | (0.05) | (0.05) | (0.07) | |
| % Disability($^a$) | 2.10 | 2.10 | 2.15 | 0.77 |
| | (0.14) | (0.14) | (0.14) | |
| % Large family | 4.05 | 4.02 | 4.21 | 0.52 |
| | (0.20) | (0.20) | (0.20) | |
| % Chronic illness($^b$) | 2.07 | 2.05 | 2.22 | 0.41 |
| | (0.14) | (0.14) | (0.15) | |
| Number priority points | 28.01 | 27.99 | 28.11 | 0.28 |
| | (7.31) | (7.16) | (8.37) | |
| Number of observations | 44,661 | 39,538 | 5,133 | |

Population: Population of interest with identified address (Identified Address) or without identified address (Not Identified Address). Notes: ($^a$) Disability refers to "Applicant, one of the parents/guardians or a sibling is disabled." ($^b$) Chronic illness refers to "Applicant has a chronic digestive illness." Standard deviations are in parentheses. Reported p-values are for a two-tailed test. Source: Barcelona primary school applications data set matched with enrollment and population census data, 2005-2010.

[33]For the years 2005 and 2006, such a database was not readily available. However, a map for the whole city like the one shown in Online Appendix Figure 1, was available in DGN (design) format, which allowed us to construct such a database. We are grateful to the staff at Geodata for their help.

## A.2 Matching applications and school enrollment data sets

The matching of the applications data set and the school enrollment data set is done through the following steps. First, the two data sets are matched on five common variables: (1) the academic year of the application, (2) the first name of the child, (3) the first surname of the child, (4) the second surname of the child,[34] and (5) the town where the child resides. Among those observations that did not match in the first round, on a second round they are matched on the five variables above plus (6) the name of the street of their residence.

Among those observations that did not match in the second round, we then tried a series of matching rounds that replicated the first round, but we used part of the name and/or parts of the two surnames rather than the full name and full surnames in order to allow for typos in these variables (e.g., we matched on (1) the academic year of the application, (2) the first four letters of the first name of the child, (3) the first surname of the child, (4) the second surname of the child, and (5) the town where the child resides). We undertook 30 different rounds.

Among the remaining observations that did not match in any of the previous 32 steps, we then used the Stata module to probabilistically match records (for more details see `http://ideas.repec.org/c/boc/bocode/s456876.html`) on (1) the academic year of the application, (2) the first name of the child, (3) the first surname of the child and, (4) the second surname of the child.

For each of the rounds described above, we checked that the matches produced were correct. We considered that matches were not correct if the address in the two data sets was different[35] or if in the case that the address was different, the child had a very common name.[36] For those that we considered incorrect matches, we used the Stata module to probabilistically match records as described above.

In all our steps, it can be the case that a child from the application merges with two different entries in the school enrollment data set. This is mainly due to children enrolling at two different schools. This can happen for different reasons (e.g., trying to improve their allocation or a school canceling some seats at the last minute). We deal with these double matches in the following way: First, we create a variable that tells us whether the two schools are from the application list. There are three cases. (i) In the first case, one school is from the application list and one is not. In this case, we keep the school *not* from the application list if this school is not from the Barcelona public system (i.e., private schools in Barcelona and schools outside Barcelona), and we keep the school from the application list if the one that is *not* from the application list is from the Barcelona public system.[37] (ii) In the second case, both schools are from the application list. In this case, we keep the one that has a higher ranking in the application list. (iii) In the third case, both schools are *not* from the application list. Within this case, there are several situations. If both schools are from the Barcelona public system, we then keep the one from the residence neighborhood of the child (in case one is from the neighborhood and the other is not) or the one that was assigned by the system (in case the two schools are from the child's residence neighborhood). In the case that both schools are from outside the child's residence neighborhood, we assigned one of the schools

[34] All Spanish people have two surnames. The first surname is the first surname of their father and the second surname is the first surname of their mother. Non-Spanish people often include a second surname in Spanish applications, but they could also leave it blank.

[35] Matching the addresses in the two different data sets would have been very complicated, since even in the case where the addresses were the same, they were very often written in very different ways in the two data sets.

[36] For instance, García, Martínez, López, Sánchez, Rodríguez, Fernández. A complete list can be found at `http://www.idescat.cat/pub/?id=aec&n=948&lang=en`.

[37] The idea behind this is to try to respect preferences. That is, schools listed in the application are the ones kept unless the enrollment school is from outside the system (that by definition was impossible to include in the application).

randomly.[38]

There are some observations in the application data set that we were not able to match with the school enrollment data set after our very thorough matching process. In a previous version, see Calsamiglia and Güell (2014) we explored any possible selection of the unmatched observations. We excluded families that we did not manage to match in the school enrollment data set (but we did match to the census data) and obtained the same qualitative results as in Figure 8 for the whole population of interest.

## A.3 Matching applications and census data sets

The Statistics Institute of Catalonia matched the applications data set with the 2001 Catalan population census (the last census available for public use). In order to increase the number of applicants found, the applications data set was then also matched to the 2005-2010 Barcelona municipal register. This is the administrative register that accounts for the residents of a municipality. It is updated more frequently than the census, so it allowed us to extend the number of applicants found. Fortunately for our purposes, the municipal register also asks for the educational attainment of each resident. The results of this matching process are as follows. Among our population, 95 percent of the applicants' mothers were found and 84 percent of the applicants' fathers were found. For the parents found (for both mothers and fathers), for around 78 percent of them we have information from both the census and the municipal register. For the remaining 22 percent of the parents, information was found on the municipal register.

When constructing our socioeconomic variables of interest, we generate the average at the household level, which can involve the two parents or just one of them (if the other was not found). This means that for variables in the municipal register, we have information for 98 percent of the children (83 percent of both parents, 14 percent of the mother only and 3 percent of the father only). For variables in the census, we have information for 81 percent of the children (53 percent of both parents, 18 percent of the mother only, and 10 percent of the father only). Given these figures, in general, we do not expect much of a selection issue with the socioeconomic variables. It can be checked the that share of non-missing observations for the different values of the variable $Risk_{is1t}$ (as defined in Section 6) is quite similar. For the variables from the Barcelona municipal register, the share of non-missing observations is 98 percent for all the values of $Risk_{is1t}$. For variables from the census, there are more missing observations, but their distribution is not too different across risk groups: the share of non-missing observations is 84 percent for those with $Risk_{is1t} = 0$ and those with $Risk_{is1t} \in (0, 1)$, and 82 percent for those with $Risk_{is1t} = 1$.

For our analysis, we are particularly interested in the education of the parents of our applicants whose decision entails some risk. We can use education information from both the census and the municipal register. In the census the variable education takes 10 values that range from (1) unable to read to (10) PhD. In the municipal register, the variable education is organized around four categories: (1) does not read or write, (2) less than primary school, (3) completed primary education and some secondary education, and (4) completed secondary education and beyond. Within these categories there are finer subcategories, but respondents do not always use these finer categories.

We define the share of high education from the two sources as those who have at least completed secondary education. We define for both the mother and the father and then generate the average for the household. In order to maximize the number of observations we obtain information for either source, and in cases in which we have information from both sources we take the maximum value. In a previous version, we used information only from the census and then only from the municipal register and results

[38] There is only one child under this situation.

were qualitatively the same as results here.

### A.4 Siblings identification in 2005

For the year 2005, we only have the *aggregate* number of priority points obtained and not the more detailed by characteristics. However, given the structure of points (see Table 1), we can approximate a lower bound on this population. More precisely, anyone with more than 75 points, the maximum that one can obtain with the other characteristics, must have points for a sibling. Also, anyone that has *exactly* 70 points must have a sibling because there is no other way to obtain exactly these points. With this approximation, those that have siblings and some other points are being counted as having no siblings. This makes our analysis for 2005 more conservative because these families are less constrained in their choices than those who do not have siblings.

## B Appendix: Robustness checks

### B.1 Changes in priority points

Section 2 and Table 1 describe the system of priority points in Barcelona. This point system has remained mostly stable over our period of study, except for the following three changes, which occurred from 2007 onward:

- Change # 1: If the school ranked first was in the *administrative district* (but not the neighborhood), 15 points were given.
- Change #2: Before 2007, the extra points were added to the basic points automatically. From 2007, the extra points (for large family and chronic digestive illness) were only taken into account to break ties between people with the same number of basic points.
- Change #3: Before 2007, points for chronic digestive illness were given only if school ranked first was in the neighborhood.

These changes occurred at the same time as the change in the neighborhood definition that we exploit in this paper. However, we argue below that it is reasonable to assume that our results are driven by the change in the neighborhood definition and not by the minor changes to the point system.

- Regarding Change #1: This change affects two types of schools. First, it affects YN schools, that is, schools that for a given family are not in the new neighborhood but were in the old neighborhood *and* in the administrative district (Figure 1 displays the administrative districts). This change in points implies that families living in Barcelona who rank these schools obtain 15 points instead of 10 points (i.e., the default for schools outside the neighborhood for Barcelona residents; see Table 1). The difference in points (5 points extra) is small, especially in the context that a neighborhood school provides 30 points (and in the context where it is hard to get other points for socioeconomic characteristics, as explained in Section 2). In any case, if anything, this change in points goes in the opposite direction of the change in the neighborhood definition that we exploit in this paper because this makes YN schools a little bit safer in 2007 than we are implicitly assuming. Yet, we still find an important decline in their demand (see Table B5).

Second, this change in points affects public NN schools that for a given family are located outside the old neighborhood as well as outside the new neighborhood but still inside the administrative district. As above, this makes these particular types of NN schools a little bit safer in 2007 than we are implicitly assuming (families ranking these obtain 15 points rather than 10 points). Still, we find an important decline in the demand when considering all NN schools (see Table B9).

- Regarding Change #2: In general, it should not make a difference whether these points are counted automatically or in order to break ties. After 2007, families know that if necessary the points will be counted.

- Regarding Change #3: In general, we do not expect that this can affect our results because the share of applicants with chronic digestive illness is small (see Table 1) and the points given for this condition are also low (see the first column of Table A1). Again, if anything, this would imply that after 2007 it was safer to rank a school outside the neighborhood, but we do not find evidence of this.

- Please note that, additionally, since 2010, single parents receive the same priority points as large families (three or more siblings). However, this does not affect our main years of comparison, namely 2006 and 2007. Moreover, the fraction of single parents in any given year is small.[39]

## B.2 Changes in the set of priority schools

As explained in Section 3.2, the new definition of *neighborhood* had two implications. First, the size of the set of neighborhood schools for which families have priority was reduced. Table B1 displays the number of schools in the priority set for the period 2005-2010. Second, new schools were added in the priority set for some families because the new neighborhood includes nearby schools that are on the other side of the border of the old neighborhood. Table B2 displays the percentage of new schools for families that had at least one new school added to their priority set for the period 2005-2010.

Table B1: Change in the size of the set of priority schools

| Number of schools | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 |
|---|---|---|---|---|---|---|
| Average | 21.86 | 22.10 | 7.03 | 7.62 | 7.73 | 7.94 |
| Standard deviation | 7.35 | 7.51 | 1.41 | 1.94 | 1.99 | 2.00 |
| Number of neighborhoods | 52 | 52 | 5300 | 5300 | 5300 | 5300 |

Source: Barcelona primary school applications data set matched with enrollment and population census data, 2005-2010.

[39] In the data, only one variable captures both large families and single parents. Between 2009 and 2010, it increased from 4.5 percent to 7.4 percent, which gives an indication of the incidence of single parent families in 2010.

Table B2: New schools in the priority schools

| | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 |
|---|---|---|---|---|---|---|
| Families with at least one new school (%) | 81.21 | 80.81 | 80.73 | 82.68 | 83.34 | 86.48 |
| Among these families: | | | | | | |
| Average of new schools in the set (%) | 31.48 | 31.29 | 31.60 | 31.15 | 31.12 | 31.17 |
| Standard deviation | (0.16) | (0.16) | (0.16) | (0.16) | (0.16) | (0.16) |
| Observations | 6,669 | 6,294 | 6,548 | 6,743 | 6,700 | 6,584 |

Source: Barcelona primary school applications data set matched with enrollment and population census data, 2005-2010.

## B.3 Calculation of demand and capacity of schools

In this appendix, we explain how we calculate for each school and year its demand and its capacity as defined in Section 6. Please note that in order to calculate demand and capacity, we use the *entire* population (except those who applied late or have multiple applications, since they are allocated *after* all the others).

The variable $Capacity_{st}$ (i.e., the number of seats available in school $s$ in year $t$) is calculated as follows. It can be calculated using the assigned schools in the applications database as well as in the enrollment data set. In Spain, by law, class size is 25 students.[40]

For each school and year, we define $CA_{st}$ as the capacity of school $s$ in year $t$ according to the assigned schools in the application data set (that is, the number of children assigned to a school $s$ in year $t$). We also define $CM_{st}$ as the capacity of school $s$ in year $t$ according to the school enrollment data set (that is, the number of children enrolled in school $s$ in year $t$). It can happen that due to budget constraints, the government allows class sizes to be larger than 25. When calculating capacities, we therefore allow schools with $CA_{st}$ or $CM_{st}$ equal to 25 to have up to 30 children, and schools with $CA_{st}$ or $CM_{st}$ equal to 50 to have up to 60 children and so on. This is simply not to overestimate school capacities (see more below). But, 25 is the relevant number for parents when deciding what school to rank. The allocation mechanism never assigns more than 25 children in a class. Given $CA_{st}$ and $CM_{st}$, for each school and year, we can have four cases: (i) First, if $CA_{st} > CM_{st}$, then $Capacity_{st} = CA_{st}$. (ii) Second, if $CA_{st} < CM_{st}$ and $CA_{st}$ is a multiple of 25 (i.e., class size), then $Capacity_{st} = CA_{st}$. (iii) Third, if $CA_{st} < CM_{st}$ and $CA_{st}$ is not a multiple of 25 and $CA_{st} >$(number of children assigned$_{st}$), then $Capacity_{st}$ could be either $CA_{st}$ or $CM_{st}$. But note that this is a case that will not affect our risk analysis in Section 6 because fewer children are assigned to the school than its capacity. (iv) Fourth, if $CA_{st} < CM_{st}$ and $CA_{st}$ is not a multiple of 25 and $CA_{st} <$(number of children assigned$_{st}$) $< CM_{st}$, then $Capacity_{st} = CM_{st}$.

The demand variables (i.e., $Demand_{s1tp}$, $Demand_{s1tp+}$, etc.) can be calculated directly from the application data set by counting the number of applicants ranking first school $s$ in year $t$ with priority points $p$. We can also calculate $Demand_{s1t}$, which is the total demand for school $s$ in year $t$ (i.e., $Demand_{s1t} = \sum_p Demand_{s1tp}$).

As explained in Section 4, there are two types of public schools: those strictly public and those semi-public in which there are voluntary fees. For the 337 schools demanded among our baseline population, Table B3 shows the share of semi-public schools that are underdemanded ($Demand_{s1t} < Capacity_{st}$),

[40]The relevant laws are Ley Orgánica de Ordenación General del Sistema Educativo (1990), which have remained unchanged regarding this aspect during the period of study (see Section 2 of our online appendix).

the share overdemanded ($Demand_{s1t} > Capacity_{st}$), and those for which demand is equal to capacity ($Demand_{s1t} = Capacity_{st}$). As can be seen, there are more underdemanded public schools than underdemanded semi-public schools. This means that, on average, semi-public schools do not represent a safer option for those families that can afford it.

Table B3: Capacity and demand of schools

| | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 |
|---|---|---|---|---|---|---|
| Public Schools: | | | | | | |
| Number | 161 | 160 | 162 | 164 | 167 | 168 |
| Share underdemanded | 0.64 | 0.66 | 0.61 | 0.62 | 0.62 | 0.56 |
| Share overdemanded | 0.32 | 0.30 | 0.33 | 0.33 | 0.36 | 0.39 |
| Share demand = supply | 0.04 | 0.04 | 0.06 | 0.05 | 0.02 | 0.05 |
| Semi-Public Schools: | | | | | | |
| Number | 160 | 161 | 161 | 164 | 165 | 165 |
| Share underdemanded | 0.51 | 0.46 | 0.45 | 0.43 | 0.50 | 0.55 |
| Share overdemanded | 0.45 | 0.48 | 0.46 | 0.48 | 0.38 | 0.38 |
| Share demand = supply | 0.04 | 0.06 | 0.09 | 0.09 | 0.12 | 0.07 |

Source: Barcelona primary school applications data set matched with enrollment and population census data, 2005-2010.

## B.4 Applicants to work-neighborhood-schools

As explained in Section 4, the school applications data set only includes the residential address of every applicant and not the work addresses of the applicants' parents. However, we do know if applicants received 20 points because they applied to a school in the neighborhood where one of the parents works (see Table 1). Table B4 shows that after excluding those that have a sibling in the school ranked first, the share of applicants to a school in the work neighborhood is small and fairly constant over time.

Table B4: Applicants to a school in the work neighborhood

| | 2006 | 2007 | 2008 | 2009 | 2010 |
|---|---|---|---|---|---|
| Fraction applicants | 3.40 | 3.15 | 2.85 | 2.76 | 2.83 |

Source: Barcelona primary school applications data set matched with enrollment and population census data, 2005-2010.

## B.5 Detailed results section 5

Table B5: Families ranking a YN school first over time

| | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 |
|---|---|---|---|---|---|---|
| Proportion of families | 20.17 | 20.92 | 9.65 | 7.28 | 7.28 | 8.11 |
| Observations | 6,669 | 6,294 | 6,548 | 6,743 | 6,700 | 6,584 |

Note: YN schools are schools that, for any given family, are located in their old neighborhood but not in their new neighborhood. Population: Baseline population. Source: Barcelona primary school applications data set matched with enrollment and population census data.

Table B6: Families ranking a NY school first over time

| | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 |
|---|---|---|---|---|---|---|
| Proportion of families | 9.34 | 9.30 | 17.14 | 17.79 | 18.36 | 17.53 |
| Observations | 5,416 | 5,086 | 5,286 | 5,575 | 5,584 | 5,694 |

Note: NY schools are schools that, for any given family, are not located in their old neighborhood but are in their new neighborhood. Population: Baseline population living close enough to the corner of the old neighborhood such that their new set of priority schools includes at least one school that was not in their old set. Source: Barcelona primary school applications data set matched with enrollment and population census data.

Table B7: Families ranking a YY school first and live in the center of old neighborhood over time

| | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 |
|---|---|---|---|---|---|---|
| Proportion of families | 71.51 | 71.11 | 86.13 | 88.36 | 85.57 | 81.69 |
| Observations | 1,253 | 1,208 | 1,262 | 1,168 | 1,116 | 890 |

Note: YY schools are schools that, for any given family, are located both in their old neighborhood and in their new neighborhood. Population: Baseline population living in the center of the old neighborhood such that their new set of priority schools includes only schools that were in their old set. Source: Barcelona primary school applications data set matched with enrollment and population census data.

Table B8: Families ranking a YY school first and live in the corner of old neighborhood over time

| | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 |
|---|---|---|---|---|---|---|
| Proportion of families | 56.94 | 56.04 | 64.11 | 65.04 | 64.26 | 65.23 |
| Observations | 5,416 | 5,086 | 5,286 | 5,575 | 5,584 | 5,694 |

Notes: YY schools are schools that, for any given family, are located both in their old neighborhood and in their new neighborhood. Population: Baseline population living close enough to the corner of the old neighborhood such that their new set of priority schools includes at least one school that was not in their old set. Source: Barcelona primary school applications data set matched with enrollment and population census data.

Table B9: Families ranking an NN school first over time

| | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 |
|---|---|---|---|---|---|---|
| Proportion of families | 12.57 | 12.63 | 8.16 | 8.93 | 9.61 | 9.28 |
| Observations | 6,669 | 6,294 | 6,548 | 6,743 | 6,700 | 6,584 |

Note: NN schools are schools that, for any given family, were not located in their old neighborhood and are not in their new neighborhood. Population: Baseline population. Source: Barcelona primary school applications data set matched with enrollment and population census data.