

# The Illusion of School Choice: Empirical Evidence from Barcelona\*

Caterina Calsamiglia<sup>†</sup>

Universitat Autònoma de Barcelona  
& Barcelona GSE

Maia Güell<sup>‡</sup>

The University of Edinburgh,  
CEPR, FEDEA & IZA

This version: April 2015

## Abstract

The Boston mechanism is a school allocation procedure that is widely used around the world. To resolve overdemand, priority is often given to families who live in the neighborhood school. We note that such priorities define some schools as being safer. We exploit an unexpected change in the definition of neighborhood in Barcelona to show that when allowing school choice under the BM with priorities: (1) the resulting allocation is not very different from a neighborhood-based assignment, and (2) important inequalities emerge beyond parents' naivete found in the literature.

**Key words:** School choice, Boston mechanism, Priorities

**JEL classification:** C78, D63, I24

---

\*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, Derek Neal, Mar Reguant, Sevi Rodríguez Mora, 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 Ramon 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. Güell acknowledges financial support from the Spanish Ministry of Science and Innovation (ECO2011-28965) 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>.

<sup>†</sup>Email: [Caterina.Calsamiglia@uab.cat](mailto:Caterina.Calsamiglia@uab.cat)

<sup>‡</sup>Corresponding Author: Maia Güell, The University of Edinburgh, School of Economics, 30 Buccleuch Place, Edinburgh EH8 9JT, United Kingdom. Email: [maia.guell@gmail.com](mailto:maia.guell@gmail.com)

# 1 Introduction

During the last two decades, over two-thirds of OECD countries have implemented different forms of school choice that depart from residential-based assignment (see Musset (2012)). The aims of school choice programs are to improve the matching between children and schools as well as to improve students' educational outcomes.<sup>1,2</sup> Yet, the concern is that disadvantaged families are less able to exercise choice, which leads to equity concerns.<sup>3</sup>

When school choice is implemented, the presumption is that the final allocation will be largely determined by parents' preferences. In this paper, we provide empirical evidence that challenges this presumption. In a nutshell, we show that under the most widely used school choice mechanism around the world, most families avoid revealing their preferences and systematically apply for schools in their neighborhood. Only some of the advantaged families will be able to choose their preferred school.

Under centralized school choice procedures, parents are asked to submit a list with their ranking of schools, and then a set of rules determines the final allocation. One of the most widely used procedures in school choice is the so-called Boston mechanism, henceforth the B-mechanism or BM.<sup>4</sup> This mechanism assigns all applicants to the school ranked first, and if there is overdemand for a school, ties are resolved according to priorities. These priorities can be defined through a random lottery or according to criteria such as distance to the school, existence of siblings in the school, or other socioeconomic variables. Those rejected from their school ranked first can opt for the seats that remain free only after considering everyone else's school ranked first. This process drastically reduces the chances of being accepted in any particular school after being rejected from one's school ranked first. 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, thereby establishing a trade-off between *preferences* and *safety*.<sup>5</sup> On the other hand, Abdulkadiroğlu, Che, and Yasuda (2011) show that lack of truth-telling can be useful for eliciting cardinal preferences in a way that other mechanisms cannot. But

---

<sup>1</sup>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 (2012) and Calsamiglia, Fu, and Güell (2014) evaluate the performance of different mechanisms with estimated preferences using data from Beijing (China) and Barcelona (Spain), respectively.

<sup>2</sup>A large set of papers explore the impact of school choice on elementary and secondary achievement, high school graduation, and college attendance. Results are mixed, finding positive or nonsignificant 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).

<sup>3</sup>Epple and Romano (2003) and Brunner and Imazeki (2008) explore how choice programs affect sorting in schools. Allen, Burgess, and McKenna (2010) evaluate the impact on sorting that results from removing the Boston mechanism and replacing it with the so-called Gale-Shapley mechanism, and of changing priorities from strict priorities according to distance to the school to living in the catchment area of the school, which happened in Brighton and Hove in England. Burgess, Greaves, Vignoles, and Wilson (2009) study what *type* of schools parents choose in England in terms of schools' characteristics and whether the type differs across parents' characteristics.

<sup>4</sup>This was the mechanism used in the city of Boston until 2005. See Abdulkadiroğlu (2013) for a review.

<sup>5</sup>The fact that families do not have incentives to reveal their true preferences is referred to as *manipulability* of the mechanism. Pathak and Sönmez (2013) provide a measure of the degree of manipulability of mechanisms and document that local administrations have moved away from more to less manipulable mechanisms.

Abdulkadiroğlu, Che, and Yasuda (2011) assume that overdemand for a certain school is resolved randomly, whereas in most cities around the world, when a school is overdemand, ties are also broken according to priority points based on criteria such as having siblings or living near the school.<sup>6</sup> In general, the literature has overlooked the role of priority points.<sup>7</sup> These priority points may seem innocuous, but we show that they have a drastic effect on parents' behavior. Priorities induce discontinuity with regard to the safety of different schools for different families. We find that not only do families exclude some highly demanded schools from their submitted list (as the previous theoretical literature might have anticipated), but a large fraction of them exclude *any* school that is not of their highest priority, that is, any school outside of their neighborhood. To our knowledge, the previous literature had not established that strategic playing parents in the BM could result in an equilibrium where the huge majority parents ends up in their neighborhood school.

The contribution of this paper is to highlight the importance of priorities in the case of the Boston mechanism, we overcome large empirical challenges.<sup>8</sup> Namely, (i) preferences are not observable, (ii) families choose schools as well as where to live, and (iii) some families may be able to opt for a school outside of the public system. We exploit an unexpected change in the definition of *neighborhood* in the city of Barcelona (Spain), where the B-mechanism is applied.<sup>9</sup> This provides an exogenous change in the set of safer schools and allows us to analyze what drives school choice, keeping housing decisions constant. Moreover, data on final school enrollment (as explained below, the school that children enroll in can differ from their assigned school in the public system) will be key in shedding new light on equity concerns.

In Barcelona, priority points are given mainly to siblings and children living in the neighborhood of the school. Thus, the neighborhood school provides a safer choice. The unexpected change in the definition of *neighborhood* in Barcelona allows us to assess whether it is *preferences* or *priorities* that drives parents' school choice. Learning how parents choose by only observing their submitted rankings is problematic. For instance, as we will show, in the city of Barcelona in any given year, around 85 percent of the families get the school they rank first, but also 80 percent of families apply for a neighborhood school. By observing parents' choices at a point in time, however, one cannot identify whether this behavior results from families' preferences for the neighborhood schools or from families applying for a safer school. Moreover, to the extent that families can also change residence, these two reasons cannot be identified separately. Our strategy allows us to separate housing and schooling decisions. Parents who applied to schools in 2007 were not able to change their residence, because the change in the definition of *neighborhood* was very close to the school

---

<sup>6</sup>They show that if families share the ranking over schools but differ in the intensity attached to each of them, then only those families who value a given school relatively more will apply for it and therefore be assigned to it.

<sup>7</sup>Only Calsamiglia and Miralles (2012) theoretically show that all parents playing it safe 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.

<sup>8</sup>Chen and Sönmez (2006) and Calsamiglia, Haeringer, and Klijn (2010) identify the existence of this trade-off for the BM through lab experiments. But the actual relevance of this trade-off depends on the risk of applying for their preferred school and on the relative valuations for the different schools. Lab experiments cannot replicate the risk involved in actual school choice, and therefore the ability to capture the size of this trade-off is limited.

<sup>9</sup>Machin and Salvanes (2010), Bogart and Cromwell (2000), and Ries and Somerville (2010) also exploit variation in the catchment area in school choice to analyze the effect on housing price changes due to school quality.

application deadline. Moreover, crucial for our purposes is that the set of schools for which families have priority changed with this new definition of *neighborhood*, and this change is exogenous to families. Therefore, the set of safer schools that families have when they lived in any given address changed before and after 2007. On the one hand, the set of neighborhood schools was reduced, declining from an average of 24 schools to an average of 7. On the other hand, the identity of the schools included for a given family changed. Our key insight is that if parents choose according to their preferences, then a change in neighborhood definition should not change the distribution of choices. Instead, if parents play it safe, then a change in neighborhood definition can affect their choices systematically, because the incentives to apply to a given school have changed. A first thing to note is that between 2006 and 2007, the fraction of families that applied for a school in the neighborhood remained constant, despite the reduction in the set of schools and the change in the definition of neighborhood. Further, we focus on families' schools ranked first and find that after the change in neighborhood definition, families avoid applying to the *old* neighborhood schools and apply to the *new* neighborhood schools. We find that families not only change the schools they choose after the change in neighborhood definition, but they do so in a clear way that supports the idea that their preferences play a small role: they change their behavior mainly by excluding any school that is not a neighborhood school anymore and incorporating the new neighborhood schools. This implies that the gains of school choice in terms of matching of children and schools seem limited, because the equilibrium allocation is not very different from that without choice.

Studies such as Lavy (2010) and Hastings, Kane, and Staiger (2008) exploit the implementation of the B-mechanism 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 B-mechanism, from the analysis provided in this paper we learn that these results may only be a lower bound on the effects of implementing a mechanism that would better respond to parents' preferences.<sup>10</sup> Our results suggest that, in general, preferences may play a small role in determining the allocation when choice is implemented.

One important concern in the debate regarding the B-mechanism is that naive parents, being unable to strategize, may be harmed by the system (see Pathak and Sönmez (2008)). In this paper, we merged our school application and assignment data with school enrollment data to provide novel evidence of this important issue. Abdulkadiroğlu, Pathak, Roth, and Sönmez (2006) report that in Boston, 19 percent of parents seem to be naive, playing a dominated strategy. We also find a similar fraction of parents that exhibit seemingly naive behavior. Enrollment data allow us to rationalize some of this behavior. We find that of those who are unlucky and do not get their school ranked first, 14 percent of them go to private school (only 4 percent of schools are private in Barcelona), and around 30 percent do not get any of the schools that they ranked (enrolling in some other public school). Calsamiglia, Miralles, and Martinez-Mora (2013) show theoretically that parents with the option of private school can take higher risks and apply for the best schools in the public system, appearing to be naive if the private option is ignored. Our empirical evidence

---

<sup>10</sup>Hastings, Kane, and Staiger (2008) show, in that case, policy makers strongly advised parents to state their true choices and were not clear about the details of the mechanism, which lead to limited strategic behavior.

suggests a new and important source of inequality that the B-mechanism induces.

Our application data set was merged with the population census data in order to yield information about parents' socioeconomic characteristics. Having this information allows us to analyze the level of education for these families. We find that those who are particularly harmed by the system, that is, by not getting any of the schools ranked in their application, have systematically lower levels of education than those applicants who do not take any risk. Similarly, those with higher levels of education can take higher risk and therefore have greater access to the best schools in the city.

This paper shows that with the BM, priorities overtake the role of preferences for most applicants. The risk involved in stating preferences is not worth taking, leading most of the applicants to apply for one of the neighborhood schools, independently of the identity of those schools. Only a few families who have the option of private school if they are unhappy with their allocation can take the risk of stating their preferences. Those who dare to take risks without having an outside option are particularly harmed by the system, which rationalizes why most families do not take risks.

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 describe our data. In Section 4 we discuss the empirical challenges involved in trying to identify what drives parents' school choice. In Section 5 we explain our empirical strategy. In Section 6 we analyze what drives parents' choices. In Section 7 we analyze the rationality behind parent's choices and inequalities of the system. Finally, Section 8 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.<sup>11</sup> 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.<sup>12</sup>

Educational policies in Spain are determined at three levels: national, regional, and municipal.<sup>13</sup> In Spain families have the right to choose their children's schools.<sup>14</sup> Parents choose the primary

---

<sup>11</sup>This right started in 1990 with the recognition of preschool education in the national law Ley Orgánica de Ordenación General del Sistema Educativo (LOGSE). This has been developed further since then. Preschool education is divided into two stages: the first one (0-3) and the second one (3-6), which is appended to primary schools.

<sup>12</sup>Most schools include both preschool (3-6 years old) and primary school (from 6 years old on). The LOGSE in 1990 established that in order to universally incorporate 3-year-old children into the public system, primary schools in the public system should expand and include preschool. In fact, public childcare centers that offer care for children under 3 do not offer care for older children, and the supply of care is limited.

<sup>13</sup>In the case of the city of Barcelona, whose region is Catalonia, the regional government is called the Generalitat de Catalunya, and the municipal government is called the Ajuntament de Barcelona. Together they form the Consorci d'Educació de Barcelona, which deals with the management of public education in Barcelona. All education legislation cited in this section can be found in Section 1 of our online appendix.

<sup>14</sup>This was first established in the 1985 national law called Ley Orgánica reguladora del Derecho a la Educación (LODE). Further developments on the application process, the allocation algorithm, and priorities for the region of Catalonia are developed in decrees; the relevant decrees for our period of study are 252/2004 (for the years 2005 and 2006) and 75/2007 (from 2007 until 2010). Finally, further details on the exact weights of the established priorities are provided every year (even if there are no changes) under resolutions from the regional educational ministry.

school for their children in March of the natural year that their children turn 3. In general, this school will be the one that they will attend until they are at least 12 years old and possibly until they finish secondary education.<sup>15</sup>

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.<sup>16</sup> Once applications with families' rankings have been submitted, all applicants are allocated to the school they have ranked first.<sup>17</sup> 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 (more details to follow). 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, spots are allocated again according to the priority points of the first round.<sup>18</sup> The process continues until all children have an assigned seat in some school. 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 A 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 the 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

---

The relevant ones for our period of study are EDC/712/2005, EDC/449/2006, EDU/904/2007, EDU/349/2008, EDU/553/2009 and EDU/107/2010.

<sup>15</sup>Schools either offer primary and secondary education or, if they do not offer the latter, they still give priority to a specific secondary school when applying.

<sup>16</sup>Section 2 of our online appendix provides the application forms for the academic years 2006-2007 and 2007-2008, which in the paper we refer as years 2006 and 2007, respectively. One can notice that in 2006, families could rank up to 7 schools, and in 2007 families could rank up to 6 schools. Although the application form included only 6 or 7 schools by default, up to 10 schools could be requested. In practice, very few applicants used the additional slots.

<sup>17</sup>Applicants with special needs are allocated first before everyone else. These include important disabilities, extreme socioeconomic circumstances, or being a recent immigrant to the country (having arrived in the last two years).

<sup>18</sup>Other cities around the world resolve ties by assigning priorities according to the school each applicant applies to in each round. The way ties are broken in Barcelona gives even more weight to the school ranked first.

of the parents/tutors 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 applied to a school living outside the neighborhood but still close enough (with the *administrative district*) would obtain 15 points (more details on these districts will be given in Section 5.1). Families can apply to a school in another city from their residence. But priority (10 points) is given to families that apply to a school in a city where they live if it is not in their neighborhood. 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
<b>Basic Points</b>	
<i>Siblings</i>	
Applicant has a sibling in school of ranked first	40
<i>Neighborhood<sup>(a)</sup></i>	
Residence in neighborhood of school ranked first	30
One of the parents/tutors works in neighborhood of school ranked first	20
Residence in the <i>administrative district</i> (not neighborhood) of school ranked first <sup>(b)</sup>	15
Residence in the city (not neighborhood) of school ranked first	10
<i>Socioeconomic characteristics</i>	
Low-income family <sup>(c)</sup>	10
Applicant, one of the parents/tutors, or a sibling is disabled	10
<b>Extra Points<sup>(d)</sup></b>	
<i>Socioeconomic characteristics</i>	
Large family (three or more siblings) or single parents <sup>(e)</sup>	15
Applicant has a chronic digestive illness <sup>(f)</sup>	10

Notes: <sup>(a)</sup> *Neighborhood* criteria are mutually exclusive: The work address of one of the parents/tutors can be considered *instead* of the residence address of the applicant. *Neighborhoods* of the city of Barcelona are explained in Section 5.1. <sup>(b)</sup> *Administrative district* criteria apply only from 2007 onward. *Administrative districts* of the city of Barcelona are explained in Section 5.1 and Figure 1. <sup>(c)</sup> *Low-income* indicates a family in which one of the parents/tutors receives welfare income or the total family income per capita is less than one-third of it (around 12,000 euros per year). <sup>(d)</sup> 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. <sup>(e)</sup> Single parents have received the same points as large families since 2010. <sup>(f)</sup> 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 1 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, trading off safety and preferences. However, we will find that some families

have an outside option, and we will be able to understand how they choose schools as we highlight important inequalities that the mechanism generates.

### 3 Data

We exploit four data sets that provide us with detailed information on: (i) the demand of schools of each family in Barcelona, (ii) the supply of schools in Barcelona, and (iii) the final allocation for each child and school in Barcelona for the years 2005-2010. Additionally, the family microdata have been merged with the population census data to obtain socioeconomic characteristics. We describe these in the following subsections.

#### 3.1 Barcelona primary school applications data set

The demand for each school in Barcelona comes from a unique and rich administrative data set that consists of the universe of applications for primary schools in Barcelona of children that turned 3 in the years 2005-2010. This population consists of 77,825 children.

For all children, we have a complete application form that consists of the following variables: their ranking of schools (up to 10 schools) and all 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).<sup>19</sup> 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.<sup>20</sup> Table 2 describes this data set year by year.

Our first analysis concentrates on understanding the choice of the school ranked first by applicants. As we explain in more detail in Section 6, for this exercise our population of interest consists of those applicants that: live in the city of Barcelona and have no siblings in the school ranked first. That is, on the one hand, we want to analyze those families for whom the definition of *neighborhood* changed in 2007 (see Section 5.1 for details).<sup>21</sup> On the other hand, for families with older siblings, 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. We therefore exclude them from our analysis.<sup>22</sup> Crucial for our first analysis is to identify the address

---

<sup>19</sup>For the year 2005, we only have the *aggregate* number of priority points obtained and not the detailed characteristics. For most of our analysis, such details will not be necessary. However, whenever we need to use the detailed points, the results for 2005 will be less precise because of this fact. Still, all the results of the paper can be derived simply by using the years 2006 and 2007. We include the other years for completeness.

<sup>20</sup>In our data, 1.25 percent of the observations for the school that they assigned is missing because they end up in a school outside Barcelona.

<sup>21</sup>Our data set only includes the residential address of every applicant and not the work addresses of the applicants' parents. This means that we cannot tell how the neighborhood changed for those that live outside Barcelona and apply to 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 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.

<sup>22</sup>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

of the applicants in order to identify how the change in definition in neighborhood affected them. 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 (see Table 2). In Appendix C 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. We will refer to our population of interest (i.e., applicant lives in the city of Barcelona and has no siblings in the school ranked first) for whom we can identify their address as our baseline population (highlighted in bold in Table 2).

In our second analysis, we analyze how risky the behavior of families is. For this purpose, we calculate the aggregate demand for each school using the total population of applicants.

Table 2: Barcelona Primary School Applications Data Set, 2005-2010

Number of applicants	2005	2006	2007	2008	2009	2010
Live in Barcelona & no sibling in school ranked first:						
Address identified	<b>7,775</b>	<b>6,649</b>	<b>6,893</b>	<b>7,198</b>	<b>7,134</b>	<b>7,085</b>
Address not identified	718	880	800	606	576	542
Live in Barcelona & have sibling in school ranked first	3,087	4,342	4,642	4,794	4,798	4,812
Do not live in Barcelona	797	633	596	768	838	862
<b>Total Population</b>	<b>12,377</b>	<b>12,504</b>	<b>12,931</b>	<b>13,366</b>	<b>13,346</b>	<b>13,301</b>

Note: For 2005, applicants with a sibling in the school ranked first are approximated as described in footnote 22.  
Source: Barcelona primary school applications data set, *Consorti d'Educació de Barcelona.*, 2005-2010.

### 3.2 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 process and the allocation process are *identical* for 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.<sup>23</sup>

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.

<sup>23</sup>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 whether *priorities* or *preferences* determine the choice for schools (see Section 4) and (ii) the structure of these payments did not change before and after 2007, which is key because we identify *priorities* versus *preferences* from a change in the neighborhood definition that occurred in 2007 (see Section 5). 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 A3 in Appendix D,

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, around 48 percent are public, 48 percent are semi-public, and 4 percent are private. In this sense, our paper studies the allocation of children in 96 percent of the schools in the city.

We can merge the application data set with the school data set to learn about the demand for schools. In any given year, there is demand for all available public and semi-public schools. For our baseline population, for the whole period of study, there is demand for 368 schools, 337 of these schools are in the city of Barcelona.

### 3.3 Barcelona school enrollment data set

The third data set that we will exploit is the school enrollment data set of the residents in Barcelona. The school where applicants end up enrolling is not necessarily the same school to which they were allocated by the administration (and could include a private school). In the second part of our analysis (see Section 7), we merge these data with our application and assignment data in order to first verify whether families enroll in their assigned school or not, but also to identify families who, failing to be assigned to their desired school, enroll their children in some other school. We refer to this as an outside option, which can be either a private or a public school outside of the city of Barcelona. This approach allows us to provide some novel evidence on so-called naive parents.

The school enrollment data contain information on enrolled children from two types of schools. First, it contains *all* of the children that enrolled in any given school in the Barcelona public system (i.e., public and semi-public primary schools in Barcelona).<sup>24</sup> Second, it also includes school enrollment information for *some* schools outside the Barcelona public system (i.e., private schools in Barcelona and schools outside Barcelona) in which Barcelona residents have enrolled. However, this enrollment is not complete.<sup>25</sup>

We match the Barcelona primary school applications data set with the school enrollment data set on common variables (see Appendix E for details). We are able to match 96 percent of the observations in the applications data set. The resulting data set allows us to further analyze the potential inequalities generated by the system (see Section 7).

---

in any given year, a larger number of public schools are underdemanded (i.e., demand is lower than the capacity of the school).

<sup>24</sup>Preschools in Barcelona (i.e., those schools that offer education up to age 6 only) are not in this data set. 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.

<sup>25</sup>Children that submit an application to a school in the Barcelona public system may end up enrolling in a school outside the public system in Barcelona (i.e., private schools in Barcelona or schools outside the city of Barcelona). The institution that deals with the school applications in Barcelona (the *Consorci d'Educació de Barcelona*) tries to keep track of the enrollment school for all the children resident in Barcelona. It therefore asks for the enrollment information to all potential schools outside the Barcelona public system. However, schools are not obliged to provide this information, and this is why the school enrollment data set does not include the enrollment information for all of those that end up outside the Barcelona public system.

### 3.4 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 searched 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 F for details). It was possible to find one of the parents in 95 percent of the children in our application data set. The resulting data allow us to study how socioeconomic characteristics can relate to behavior in the context of school choice (see Section 7).

## 4 Empirical challenges

Table 3 displays for each year the percentage of children that enter the school that they ranked first in their school application, those that enter the school that they ranked second, and so on. We concentrate on our baseline population (those that live in Barcelona and have no siblings in the school ranked first), which were allocated to a school in Barcelona.<sup>26</sup> This table shows that on average, as many as 86 percent of children enter their school ranked first.<sup>27</sup> Also worth noting is that 3.5 percent of children, on average, does not enter *any* of the schools they ranked. This shows again that under the BM, the school ranked first is crucial.

Table 3: Parental behavior and school allocation in Barcelona

Percentages	2005	2006	2007	2008	2009	2010
Enter in school ranked 1st	88.10	85.05	83.19	85.44	86.83	88.64
Enter in school ranked 2nd	4.82	5.69	5.98	6.11	6.21	5.64
Enter in school ranked 3rd-10th	3.75	4.38	5.56	5.00	4.71	4.21
Does not enter in any ranked school	3.33	4.89	5.28	3.45	2.79	1.51
School ranked 1st is in residence neighborhood	72.85	79.89	82.02	83.34	82.32	82.16
Observations	7,713	6,649	6,893	7,183	7,122	7,075

Note: For 2005, applicants with a sibling in the school ranked first are approximated described in footnote 22.

Population: Baseline population that has been allocated to a school in Barcelona.

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

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 a trade-off between *preferences* and *safety*

<sup>26</sup>For those that were not allocated a school in Barcelona, 1.25 percent of the total population, we do not know the exact school to which they were allocated. We only know that they were allocated a public school out of the city.

<sup>27</sup>This number is almost 100 percent for those that have a sibling in the school ranked first.

when ranking schools (i.e., schools for which they obtain priority points, like the neighborhood school).<sup>28</sup> Priorities impose a discontinuity on the safety of the different schools for a given family. Hence, this trade-off may imply that parents exclude from their submitted list any school that is preferred to a school for which they are given priority points.<sup>29</sup> 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 3 displays the percentage of children who rank first a school that is in their residence neighborhood. On average across the years for our baseline population, this number is as much as 80 percent.

Therefore, the question is: is it *priorities* or *preferences* that explains these facts? Or could both reasons explain it, to the extent that families can move their residence to the neighborhood of their most preferred school? The schooling and housing decisions can be summarized as follows.

**Parents’ choice of neighborhood school: Three reasons** Considering housing and schooling decisions, parents may choose their neighborhood school for one of the following reasons:

- For families that *cannot* change residence:
  - (1) preference for the neighborhood school<sup>30</sup>
  - (2) opting for a safer school rather than another more preferred school.
- For families that *can* change residence:
  - (3) the neighborhood school is *both* their most preferred school and a safer option.

We will refer to reason (1) as “preferences” and to reason (2) as “safety.” Therefore, by observing parents’ choice of a neighborhood school at a point in time, we cannot identify whether their behavior results from preferences for that school or from the increased safety provided by having *priority* for that school. Ultimately, we cannot identify whether the school ranked first in the list reflects *preferences* or *priorities* for schools. Reason (3) is the union of the first two. Thus, to the extent that families can also choose their residence makes it even harder to distinguish the reason behind their choice. In this paper we exploit a change in the definition of *neighborhood*, which will allow us to separately identify these reasons. In the next section, we explain our empirical strategy.

## 5 Empirical strategy

We circumvent the identification problem explained in the previous section by exploiting an unexpected change in the definition of *neighborhood* that occurred in the city of Barcelona in 2007. As

---

<sup>28</sup>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.

<sup>29</sup>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.

<sup>30</sup>This could be the quality of the school and/or shorter commuting time. The way neighborhoods changed in Barcelona will allow us to provide some insights.

explained below, this change implied that while distance to schools and average quality of schools remained constant, priority points for different schools and families changed.

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 minimizing channel (3) described above. The change was announced on March 27, 2007, and families applying for a school that year were told on March 30 that they had to submit their applications between April 10 and 20, 2007.<sup>31</sup> 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 that were *safer* (i.e., those with the highest priority) also changed. In turn, this allow us to distinguish between reasons (1) and (2) described above (priorities versus preferences). In other words, it allows us to identify whether parents change their choices following the change of neighborhood definition. Our key insight is that if parents choose according to their preferences, then the change of neighborhood definition should *not* change the distribution of choices.<sup>32</sup> Instead, if parents play it safe, then the change of neighborhood definition can affect their choices by excluding the *old* safer school and including the *new* safer school.

Therefore, in our empirical analysis we analyze the changes in school choices before and after this change and, in particular, between 2006 and 2007. Our assumption is that distribution of preferences will not change between these two years. As mentioned before, we will concentrate 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 choosing for the first sibling because the choice for a second sibling is a different problem.<sup>33</sup> Note that the school for the first sibling at age 3 is chosen only *once* by each family. Thus, we necessarily need to compare different families over time. As mentioned above, our empirical strategy relies on the assumption that the distribution of families' preferences does not change between 2006 and 2007. More precisely, we need to assume that preferences of families in the same neighborhood do not change in a systematic manner that coincides with the direction in which the incentives provided by the mechanism have changed.

---

<sup>31</sup>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 1 of our online appendix).

<sup>32</sup>One concern is that the change in neighborhoods may induce changes in peer composition of schools, thereby affecting parents preferences. We compare peer socioeconomic characteristics before and after the neighborhood change and find no statistical differences in the average quality of YY, YN or NY schools. That is, peer composition in schools does change, but it is not the case that increased safety is systematically correlated with improvement in peers.

<sup>33</sup>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.

## 5.1 Change in the definition of *neighborhood*

As explained in Section 2, in the city of Barcelona, to resolve overdemand 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, but were conceptually the same. For semi-public schools, the neighborhood coincided with the administrative district.<sup>34</sup> 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 G for more details).

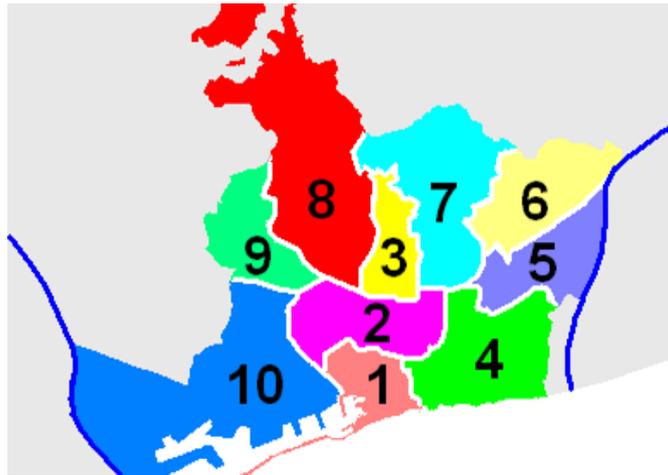


Figure 1: Administrative districts in the city of Barcelona

In March 2007, new neighborhoods were defined, moving away from the old fixed administrative districts. The change in the definition of *neighborhood* is unrelated to our concerns about the properties of the B-mechanism. It 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.

<sup>34</sup>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 demanding a school that we identify as their neighborhood is a lower bound for the years 2005 and 2006.

The new neighborhoods are based on *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.<sup>35</sup>

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. We next explain how the change in the definition of *neighborhood* changed this set.

## 5.2 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 H 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), there was also a change in the *composition* of schools in the priority set because the new neighborhood includes nearby schools that are on the other side of the border of the old neighborhood. On average, around 83 percent of the families had a change in the composition of the schools in their priority set. For these families, the change in composition involved an average of 31 percent of schools in their set (see Appendix H for details). As will become apparent shortly, it will be relevant to distinguish between: (i) families that live in the *center* 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.

---

<sup>35</sup>Before 2007, families knew which schools were in their neighborhood by looking at maps like the one in 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 demand for schools in the neighborhood, but as Table 3 shows, this was not the case. The number of families that ranked a neighborhood school first is constant over time.

Figures 2 and 3 illustrate 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 *center* of the old neighborhood (Figure 2) and for a family that lives in the *corner* of the old neighborhood (Figure 3). 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.<sup>36</sup>

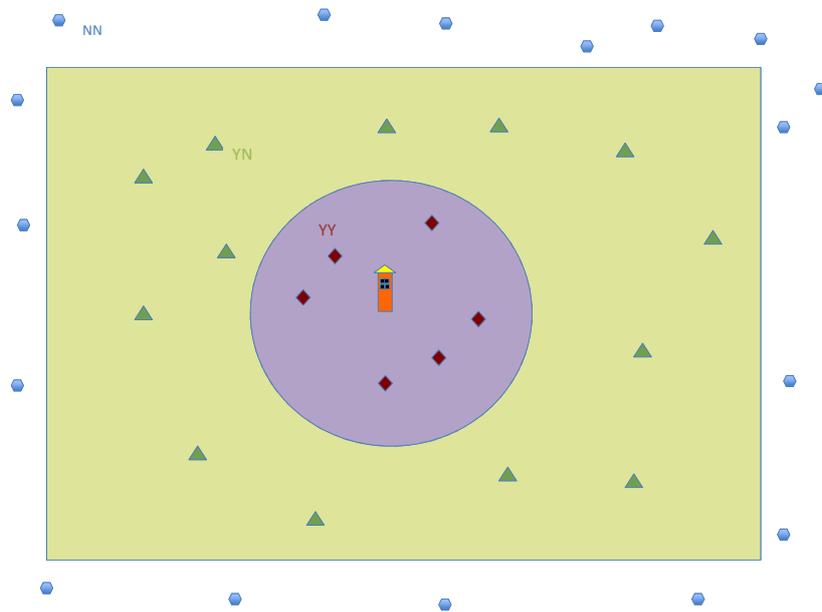


Figure 2: Change in the school priority set for an address in the center of the old neighborhood

Note: Diamonds are YY schools, triangles are YN schools, and hexagons are NN 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 Figures 2 and 3). The change in composition of 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 Figure 3). On average for the period of study, we have 83 percent of families for whom the set of priority schools changed both in size and in composition, whereas for 17 percent it only changed in size. The two changes (size and composition) of the priority set allow us to identify the

<sup>36</sup>In terms of priority points, families applying to their YY schools received 30 priority points both before and after 2007; families applying to their YN schools received 30 priority points before 2007 and 15 points after 2007; families applying to their NY schools received 10 priority points before 2007 and 30 points after 2007; and families applying to 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 A for a further discussion).

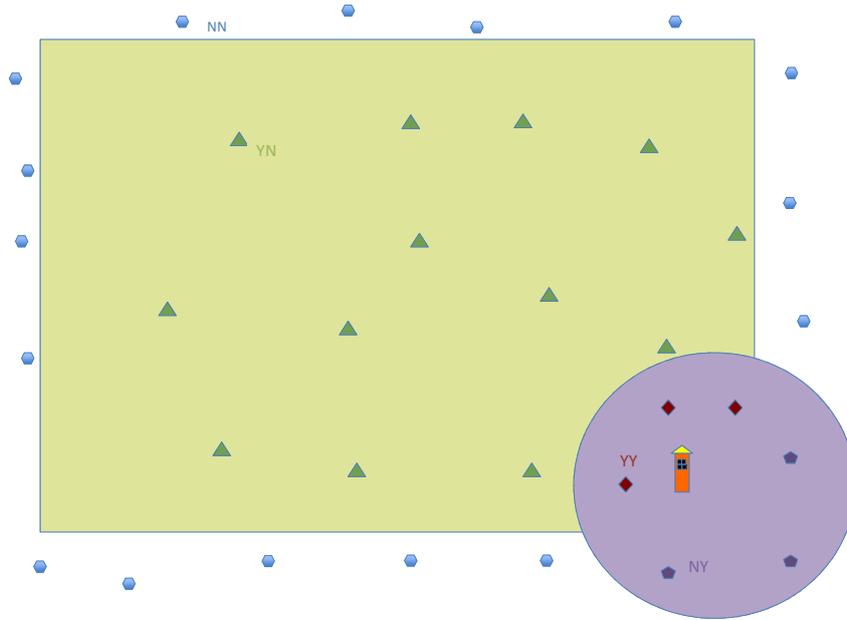


Figure 3: Change in the school priority set for an address in the corner of the old neighborhood

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

different types of schools for each block of houses. It is worth noting that by construction of new neighborhoods, the distance from families' residences to any school in the neighborhood is reduced after 2007. However, this fact carries no mileage for our identification strategy. If preferences drive choices (and families care about a shorter distance to school), then the definition of *neighborhood* is irrelevant. And if priorities are driving choices, then distance is irrelevant.

Our key insight is that we can learn what drives families' choices (preferences or priorities) by looking at the changes in demand 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.

### 5.3 Testable hypothesis

As seen in Table 3, a huge majority of families apply for a school in their neighborhood. But from this fact alone we cannot infer what drives parents' choices. Key for our analysis is to examine the response to incentives by families after the change in the neighborhood definition in 2007. As already argued, for parents who chose a school in 2007, it is very unlikely that they were able to change residence. Therefore, as discussed in Section 4, we are left with two mutually exclusive reasons for choice: *preferences* versus *priorities*.

The key for our identification is the following. If families choose according to their preferences, then we should observe no change in choices after the change in the neighborhood definition (assuming that preferences have not changed over time). Instead, if families play it safe, then we should observe that their choices change. In particular, if safety is driving choice, we should observe the following changes:

**Hypothesis 0:** *If priorities are driving school choice*, the proportion of families asking for 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 preferences are driving school choice*, the proportion of families asking for the different types of schools should not change over time.

**Hypothesis 1:** *If priorities are driving school choice*, the proportion of families that ask for YN schools should decrease between 2006 and 2007, given that these schools become less safe after the neighborhood change. But *if preferences are driving school choice*, we should observe no change.

**Hypothesis 2:** *If priorities are driving school choice*, the proportion of families asking for NY schools should increase between 2006 and 2007, given that these schools become safer to apply for after the neighborhood change. But *if preferences are driving school choice*, we should observe no change.

Regarding YY schools, since the set of schools with high priority has been reduced, if families' choices are driven by priorities, then demand for any given neighborhood school should increase. But rising demand for YY schools depends on whether families live in the center or the corner of the old neighborhood. In particular,

**Hypothesis 3:** *If priorities are driving school choice*, for families living in the center of the old neighborhood, the proportion of families asking for YY schools should increase between 2006 and 2007, given that there is a smaller set of schools with highest priority after 2007. In other words, families that were playing it safe before 2007 were asking for YY or YN schools but only for YY schools after 2007. Therefore, if they continue to play it safe, the demand for YN schools would be shifted to YY schools. But *if preferences are driving 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 in Hypothesis 3, the proportion of families asking for YY schools between 2006 and 2007 should increase as demand from 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 safer 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 asking for YY schools between 2006 and 2007. The final effect is *ex ante* unpredictable.

Similarly, there is no prediction for the change in the proportion of families asking for NN schools between 2006 and 2007 because it is unclear whether these schools are safer. The safety of applying to an NN school depends on the number of applicants that have priority points for that school. 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 demand from the neighborhood would be larger or smaller, and therefore unclear what the chances of being admitted would be for an individual outside of the neighborhood.

In the next section, we test these hypotheses in order to learn whether preferences or safety drives parents' school choice.

## 6 Results: Preferences versus priorities

In this section, we analyze the school that families ranked first in their application form. 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.

Table 4 displays the percentage of families applying to YN schools, that is, schools that were located in their old neighborhood but not in their new neighborhood. We observe the following facts.

Table 4: Testing Hypothesis 1: Families that ranked a YN school first

	2005	2006	2007	2008	2009	2010
Proportion of families	18.24	20.71	9.72	7.29	7.64	8.26
Observations	7,775	6,649	6,893	7,198	7,134	7,085

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, 2005-2010.

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 families play it safe (as outlined in Hypothesis 0). If it was families' preferences that drove choice, we should not observe any change. Moreover, as described in Hypothesis 1, if families play it safe, the change in behavior 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 in Table 4. The proportion of families that ask for YN schools was reduced from 21 percent to 10 percent, which is a decrease in demand of 52 percent. This large decrease indeed suggests that safety, rather than parents' preferences, plays a crucial role in choosing a school.

It is worth noting that in Table 4, starting in 2008, the proportion of families asking for YN schools is decreasing over time. This partly reflects the fact that families can over time move their residence to another neighborhood with their preferred school (which stops being a YN school).

We now turn to the analysis of families that choose NY schools, that is, schools that were not located in their old neighborhood but are in their new neighborhood. Table 5 displays our results.

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 a change in its composition. 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 Figure 3). 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. Again Table 5 shows 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 safety

Table 5: Testing Hypothesis 2: Families that ranked an NY school first

	2005	2006	2007	2008	2009	2010
Proportion of families	10.15	9.35	17.23	17.62	18.33	17.38
Observations	6,352	5,380	5,566	5,955	5,947	6,121

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, 2005-2010.

plays a big role in school choice (as outlined in Hypothesis 0). Moreover, as described in Hypothesis 2, if families play it safe, the change in behavior 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 Table 5. The proportion of families that ask for NY schools increased from 9 percent to 17 percent, which is an increase in demand of 89 percent. This very large increase indeed suggests that safety, rather than parents' preferences, plays a crucial role in choosing a school.

To sum up, we find evidence that families' preferences play a limited role in school choice because a change in the definition of *neighborhood* makes families change their choices. 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 Figure 2), whereas YN schools are schools that are relatively close to families' residences (see Figure 3). 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 safety. 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 chose according to distance.

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 district for whom the set of priority schools from 2007 is composed only of YY schools. Table 6 displays our results.

Table 6: Testing Hypothesis 3: Families that ranked a YY school first and live in the center of old neighborhood

	2005	2006	2007	2008	2009	2010
Proportion of families	67.04	71.08	85.91	87.45	84.67	81.33
Observations	1,423	1,269	1,327	1,243	1,187	964

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, 2005-2010.

Again, as Table 6 shows, 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 safety plays a big role in school choice (as outlined in Hypothesis 0). Moreover, as described in Hypothesis 3, if families play it safe, there should be an increase in the share of families that ask for YY schools between 2006 and 2007. This is exactly what we find. For these families, living in the center of the old district, since they play it safe, they are moving away from schools that were in their neighborhood (YN schools) and concentrating on the schools around their residence.

We next turn to the analysis of YY schools for families that live in the *corner* of the old neighborhood. As explained in Section 5.3, there is no unambiguous prediction for this case. Table 7 displays the percentage of families applying schools that are YY and live in the corner of the old neighborhood. We find that demand for YY schools has moderately increased, which is consistent with a shift in demand from YN schools to YY schools as well as some substitution between YY schools and NY schools.

Table 7: Families that ranked a YY school first and live in the corner of old neighborhood

	2005	2006	2007	2008	2009	2010
Proportion of families	51.83	56.38	63.87	64.87	63.53	64.91
Observations	6,352	5,380	5,566	5,955	5,947	6,121

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, 2005-2010.

Finally, we turn to the analysis of NN schools. Again, as explained in Section 5.3, there is no unambiguous prediction for this case. Table 8 displays the results. We see a reduction in the demand for NN schools. This suggests that families having a small set of schools and playing it safe leads to a more homogeneous demand across schools. Hence, finding an NN school with available seats may be more difficult.

Table 8: Families that ranked an NN school first

	2005	2006	2007	2008	2009	2010
Proportion of families	18.86	12.54	8.25	9.36	10.04	9.58
Observations	7,775	6,649	6,891	7,198	7,132	7,085

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, 2005-2010.

## 7 Results: Risk analysis

In this section we pool the 2005-2010 data together to study the rationality behind parents' behavior when choosing a school and the consequences of their behavior on their final allocation.<sup>37</sup> In Subsection 7.1, we first identify those applicants whose behavior could be considered naive based

<sup>37</sup>The change in neighborhoods should not have any implications for the willingness for families to take risk. Accordingly, we did not find in our data any significant differences in risk-taking behavior before and after 2007.

on their exante chances to enter the school for which they apply. In Subsection 7.2, we incorporate information about the school where children enroll (which may be different from the assigned school) to understand further the possible naivete in parents' behavior and to learn about the equity properties of this mechanism. Finally, in Subsection 7.3, we analyze the socioeconomic characteristics of families who may seem naive in order to reevaluate whether, as the literature has emphasized (see Abdulkadiroğlu, Pathak, Roth, and Sönmez (2006)), disadvantaged families may be less able to exercise choice.

## 7.1 Possible naive behavior

In this section we identify those applicants whose behavior could be considered naive, that is, that the probability of entering in the school they have chosen is zero, thereby reducing their chances of being assigned to other schools. We first concentrate on the risk that families take when deciding on the school ranked first and then do the analogous computation for the schools ranked second.<sup>38</sup>

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) are demanding each school. In order to do so, we define the following variables (see Appendix D for details on how we calculate these with our data):

- $Capacity_{s1t}$ : number of seats available in school  $s$  in year  $t$  in the first round.
- $Demand_{s1tp}$ : number of applicants ranking first school  $s$  in year  $t$  with priority points  $p$ .
- $Demand_{s1tp+}$ : number of applicants ranking first school  $s$  in year  $t$  with *more* priority points than  $p$ .
- $MinPoints_{s1t}$ : minimum amount of priority points necessary to enter a school  $s$  in year  $t$  in the first round.
- $Points_i$ : the number of priority points that applicant  $i$  has for the school ranked first.<sup>39</sup>

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 demand the school is smaller than the capacity of the school, that is  $Demand_{s1t25+} < Capacity_{s1t}$  and (ii) the number of children that have exactly 25 points that demand the school is greater than or equal to the capacity of the school, that is,  $Demand_{s1t25} > Capacity_{s1t}$ .

We define the risk of applicant  $i$  that has ranked first school  $s$  in year  $t$  as follows:

$$Risk_{i1t} = \begin{cases} 0 & \text{if } points_i > MinPoints_{st}, \\ 1 - \frac{Capacity_{s1t} - Demand_{s1tp+}}{Demand_{s1tp}} & \text{if } points_i = MinPoints_{st}, \\ 1 & \text{if } points_i < MinPoints_{st}. \end{cases}$$

<sup>38</sup>We do not go further than the second school ranked because it seems too demanding for anyone to be able to calculate the chances of entering a school ranked third and below.

<sup>39</sup>Recall that for the case of Barcelona, the number of priority points is constant for all schools ranked.

That is, the variable  $Risk_{is1t}$  takes value 0 when an applicant has more priority points than the minimum necessary to enter the school ranked first. It takes value 1 when an applicant has fewer priority points than the minimum necessary to enter a school. For applicants with the same number of priority points than the minimum necessary to enter a school, then  $Risk_{is1t}$  is 1 minus the probability of entering. The latter is the total capacity left after considering applicants that have more points than the minimum necessary ( $Capacity_{s1tp} - Demand_{s1tp+}$ ) divided by the total number of competitors (i.e., who demand the same school with the same number of priority points,  $Demand_{s1tp}$ ).

Once we have calculated  $Risk_{is1t}$ , we will turn to the calculation of the risk around the school ranked second. In order to calculate the probability of entering in schools ranked second, we compute the expected second round demand for each school. Demand in the second round results from applicants not getting their school ranked first, which depends on the realization of the lottery, which is unknown to applicants at the time they submit their application. We therefore run 500 simulations on what the demand for the second round would have been under different realizations of the lottery. Then,  $Prob_{is2t}$  is the average probability of being accepted in the school ranked second for individual  $i$  in year  $t$ , and  $Risk_{is2t} = 1 - Prob_{is2t}$  is the corresponding risk.

In order to calculate  $Risk_{is1t}$  and  $Risk_{is2t}$  in our data, we focus on those children in our baseline population for whom we can potentially have information about the school in which they enrolled.<sup>40</sup> We also exclude from the analysis those that either applied after the deadline or were found to have submitted multiple applications, since these cases are assigned only *after* those that have applied on time have a secured seat. This amounts to 41,266 applicants and 331 schools.<sup>41</sup>

Table 9 reports the value of the variable risk for the schools ranked first. We observe the following. First, for 58 percent of the families, their school choice entails no risk.<sup>42</sup> Second, 40 percent of the families entail some risk with the school they rank first. This means that they are demanding 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, 91 percent have 30 points or more and demand a school in their neighborhood; 4 percent have 30 points and demand a school outside their neighborhood; and 5 percent have fewer than 30 points and demand a school outside their neighborhood. So, interestingly, most of the risk entailed by families comes from applying to their neighborhood schools, in an equilibrium in which most families play it safe. Third, 2 percent of families are choosing a school in which they have no chance of entering. Their choice *without further information* seems naive.

As mentioned in Section 2, once the assignment is complete, overdemanded schools open waiting

---

<sup>40</sup> Again, we exclude those with a sibling in the school ranked first because in our data, we find that their behavior involves no risk, i.e.,  $Risk_{is1t}=0$ .

<sup>41</sup> This means 1,468 fewer observations than in our original baseline population: 504 children applied to a preschool in Barcelona for which we do not have enrollment information (see footnote 24). Other children for whom we do not have enrollment information are: 58 children that did not rank a school first in Barcelona (although they did rank a school in another position in Barcelona), and 197 children that although they ranked a school first in Barcelona, they ranked a school second outside Barcelona. Finally, 516 children applied late and 194 children had multiple applications.

<sup>42</sup> This includes, by definition, those applicants with special needs. In our data, these children were recorded for the years 2005 and 2008-2010 and are around 1 percent of our baseline population. Since we do not have a record of them for the years 2006-2007, our estimates of  $Risk_{is1t} > 0$  are an upper bound, but given the small fraction that these applicants represent, we expect any bias to be minor.

Table 9: Risky behavior, schools ranked first

$Risk_{is1t}$	Proportion of Families (%)
0	57.98
(0, 1)	40.11
1	1.92
Observations	41,266

Population: Baseline population with enrollment information and did not apply late or have multiple applications (see footnote 41). Source: Barcelona primary school applications data set matched with school enrollment data set, 2005-2010.

lists for those that ranked a school first and were not initially allocated. If a child does *not* enroll in the assigned school (because the child ends up attending a private school or another school outside of the city), then a seat is freed up in that school. This seat can be filled by children in the waiting list for that school. In turn, these children free up a space in the school to which they were originally assigned, which again implies some readjustment will be necessary with respect to the original assignment of children to schools. This means that upon enrollment, there may be changes to the assignment, after some children leave the public system in Barcelona, thereby generating a reallocation process that changes the final allocation.

School enrollment information is particularly useful in understanding the behavior of those that may seem naive without further information. In particular, it allows us to detect successful cases of applicants that were on a waiting list. Thus, enrollment information clarifies that those that could be classified as naive may actually not be naive at all. In particular, for those who apply to a school with  $Risk_{is1t} = 1$ , we find that as much as 18 percent still end up in their school ranked first. Therefore, with the school enrollment information, we can no longer claim that those with  $Risk_{is1t} = 1$  are as naive as we initially thought: they are risk takers, indeed, but they have anticipated that the waiting list gives them a positive probability of entering the school they ranked first.<sup>43</sup> Therefore, in order to assess the rationality of applicants, we need to analyze the risk associated with the school ranked second.

For those applicants in the first round with  $Risk_{is1t} \in (0, 1]$ , we calculate the risk in the school ranked second, i.e.,  $Risk_{is2t}$ . We find that among these, 53 percent have  $Risk_{is2t} = 1$ .<sup>44</sup>

Based on the risk analysis in the schools ranked first and second, we classify as naive those where  $Risk_{is1t} \in (0, 1]$  and  $Risk_{is2t} = 1$ . Table 10 displays the share of families that seem to have a naive behavior and those that do not. We find that 23 percent of the families seem to be naive, at least *without further information*. This magnitude is in line with the 19 percent found in Boston by Abdulkadiroğlu, Pathak, Roth, and Sönmez (2006).<sup>45</sup> In the next section, we reevaluate the behavior of these naive families using school enrollment information.

<sup>43</sup>Among these, 10 percent are children with special needs for whom a seat is systematically created, increasing class size accordingly. For them, then, there is no such risk. We therefore argue that we cannot claim that these individuals are naive, but just risk takers, and include them in our analysis with the rest of the risk takers.

<sup>44</sup>Again, some of these may not be naive because of the presence of waiting lists. However, we find that among these, only 3 percent manages to get into the school ranked second.

<sup>45</sup>If we impose not only that the two first schools be overdemanded in the first round, but also that the individual has applied for some other school in the remainder of his submitted list where the probability of getting in was positive had he listed it second, then share of naive becomes 15 percent.

Table 10: Rational and Naive Behavior

	Proportion of Families (%)
Rational	77.35
Naive	22.65
Observations	41,266

Population: Baseline population with enrollment information and did not apply late or have multiple applications (see footnote 41). Source: Barcelona primary school applications data set matched with school enrollment data set, 2005-2010.

Please note that we include as *rational* those that leave the second school in the ranking empty. Their behavior is rational if they are indifferent between any other school in the city other than the one ranked first. In this case, if they were not to enter the school ranked first, they would be assigned to the nearest school to their home that has a space. Since this is an extreme type of behavior, in the next section we will separate these families from the rest.

## 7.2 Consequences of risky behavior

We now proceed to analyze the consequences for those whose choices are risky (and in particular for those that seem naive) by considering *both* the final allocation by the administration (i.e., the school where the applicants are assigned) as well as the final outcome (i.e., the school where the applicants enroll).

We analyze the consequences of any risky behavior, that is, for all applicants for whom  $Risk_{i1t} \in (0, 1]$ . Among these families, 69 percent end up entering that school. This reflects that although a large fraction of families take a risk, most of them do so when competing for their neighborhood school, and still, chances of getting in are relatively high (see discussion of Table 9).

The next question is, what happens to the rest, the other 31 percent of families, that do not make it to the school they ranked first? Again, school enrollment information is particularly useful to distinguish among the three possible outcomes that we describe below.

1. *Do not enter/enroll in any school of their ranking:* Families that are not allocated to any school of their ranking and enroll in a school outside their ranking.
2. *Enter and enroll in a school ranked in the second through tenth positions:* Families that are allocated to a school of their ranking (second through tenth positions) and enroll in a school of their ranking (second through tenth positions).
3. *Outside option:* Families that end up enrolling in a school outside the Barcelona public system (i.e., a private school in Barcelona or a school outside the city of Barcelona).<sup>46</sup> As mentioned in Section 3.3, the school enrollment data are not complete for schools outside the Barcelona public system. Therefore, we identify those with an outside option as either (i) those who enroll in schools outside the Barcelona public system for which we *do* happen to have their

<sup>46</sup>Indeed it can be the case that families apply to both a school in the Barcelona public system as well as to another school in the public system of another town nearby Barcelona. Another form of outside option could be making the application but not enrolling at any school at all (school is not compulsory until the age of 6).

enrollment information or (ii) families for which we have their school application information but not their enrollment information.<sup>47</sup>

Table 11 reports the share of families in each of these three possible outcomes. We distinguish among those are rational risk takers (i.e.,  $Risk_{is2t} < 1$ ) and naive (i.e.,  $Risk_{is2t} = 1$ ). Within the rational risk takers, we further distinguish among those who have only ranked one school in their application (i.e., leaving the second school blank) and the rest.

Table 11: Final outcomes of risky behavior

	Proportion of Families (%)			
	Rational risk-takers		Naive	All
	Rank only one school	Rank more than one school		
<i>Final outcome:</i>				
Do not enter/enroll in any school of ranking	67.79	10.43	43.74	30.47
Enter and enroll in a school ranked 2nd-10th	-	79.44	42.22	56.14
Outside option	32.21	10.13	14.04	13.38
Observations	267	2,340	2,835	5,542

Population: Baseline population with enrollment information and did not apply late or have multiple applications (see footnote 41), with  $Risk_{is1t} \in (0, 1]$ , and who do not enter school ranked first. Source: Barcelona primary school applications data set matched with school enrollment data set, 2005-2010.

As can be seen in Table 11, among those families with  $Risk_{is1t} \in (0, 1]$  that do not enter the school ranked first, there is a large heterogeneity in outcomes: at one extreme are the families with the worst final outcomes: those that do not enter (nor enroll) in *any* of the schools they have ranked in their application. In between, there are the families that enter some school in their application. At the other extreme, there are the families that have an outside option and are able to improve their allocation in the public system in Barcelona by going to another school, either private or outside Barcelona. The literature has typically emphasized the former case, but the role of the outside option has not been explored before. We find that this is important: it represents on average as much as 14 percent of those families with risk that do not enter the school they rank first. That is, these are families that have an outside option but still apply for a school in the public system in Barcelona in which their chance to enter is not one. This strategy entails some risk that involves using the outside option when failing to enter the school ranked first, and that is when we observe them in the data. If it works, then the outside option is not used and is unobservable. This is why unless one has school enrollment data, this phenomenon cannot be detected. This is important because it implies that the share of naive families is approximately 14 percent lower.<sup>48</sup> As shown in Calsamiglia, Miralles, and Martinez-Mora (2013), the presence of an

<sup>47</sup>These are 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 (see Appendix E for details). In Appendix I we explore any possible selection of the unmatched observations.

<sup>48</sup>We obtain this number assuming that the share of families with an outside option that *do* enter in the school ranked first is the same as for families that do not enter in school ranked first, since the outcome depends on the realization of a fair lottery.

outside option can help rationalize why a risky behavior is not necessarily associated to extreme preferences for the particular school (as suggested by Abdulkadiroğlu, Che, and Yasuda (2011)).

We find that the outside option is particularly relevant in understanding those who only rank one school. We find that 32 percent of these families only apply for one school because if they are unlucky and do not get in their first choice, they will opt for their outside option.

To conclude this section, school enrollment data allow us to detect those that successfully end up in the school they ranked first because of the waiting list, as well as those who end up in a school outside the public system (outside option) because they were not successful at getting a set in the school they rank first. This allows us to understand better those that at first, without further information, seem naive. This evidence suggests that the fraction of applicants that have an outside option can take higher risks and get increased access to the best schools in the system, as suggested in Calsamiglia, Miralles, and Martinez-Mora (2013). Our findings shed new light on understanding perceived naive behavior in school choice.

### 7.3 Risky behavior 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 behaviors (e.g., Hastings, Kane, and Staiger (2008)). We are able to analyze this issue with our data. As explained in Section 3.4, our application data set was merged with the census data 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 F.1 for details).

Figure 4 displays the average household education for rationals and naive as defined in Table 10. Families that are naive are not associated with lower education levels than those that behave rationally. Admittedly, the differences in education between the two sub-populations are not very large, but, if anything, naivete is associated with families with higher levels of education. 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.

Remember that  $Risk_{ist} \in (0, 1]$  implies that whether the applicant is accepted or not depends on the realization of the lottery. Of those taking risk, some are successful and get the school they wanted, but a random sample of them are unlucky. Precisely from observing the outcomes of a random representation of risk takers, the unlucky, we can infer what the outcomes would have been for all risk takers. We next explore the relationship between families' education and final outcomes. Figure 5 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 11, namely, rational risk takers and naive, and among the rational risk takers those that only apply to one school versus those that apply to more than one. The horizontal line represents the average share of highly educated parents in the household for those whose choice entailed no risk, i.e.,  $Risk_{ist} = 0$ . 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 rationality behind school choice. Looking at Figure

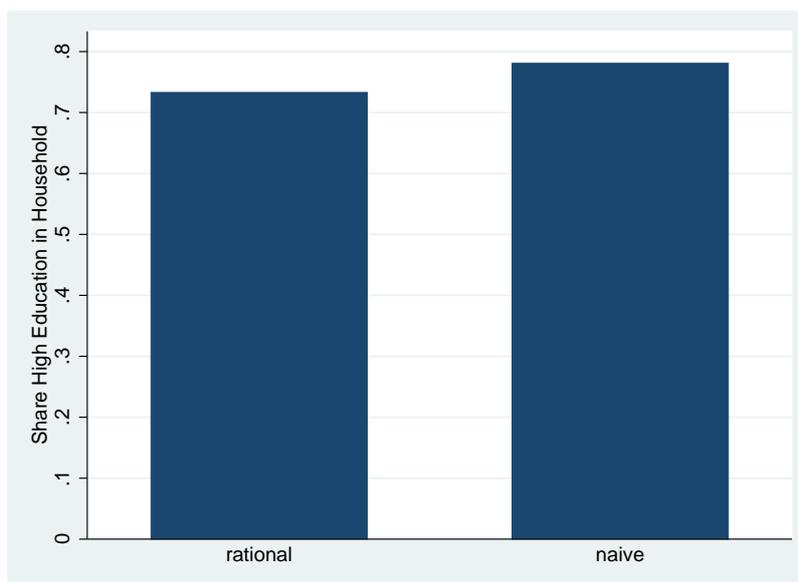


Figure 4: Household Education and Rationality

Note: High education is defined as completed secondary education or more using the population census. Population: Baseline population with enrollment information and did not apply late or have multiple applications (see footnote 41). Source: Barcelona primary school applications data set matched with school enrollment data set and Catalan population census, 2005-2010.

5, we notice the following. First, those with an outside option have higher education levels than the non-risk takers. These are families that decide to apply for 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. Third, those that do not enter any school of their ranking have lower levels of education than the non-risk takers. 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 applying for a better school and if unlucky are placed in a “leftover” school, their neighborhood school.

For rational risk-takers and the naive, the pattern of education and final outcomes are similar to those described above. It is remarkable that for both the rational risk takers that only rank one school as well as the naive, 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.

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

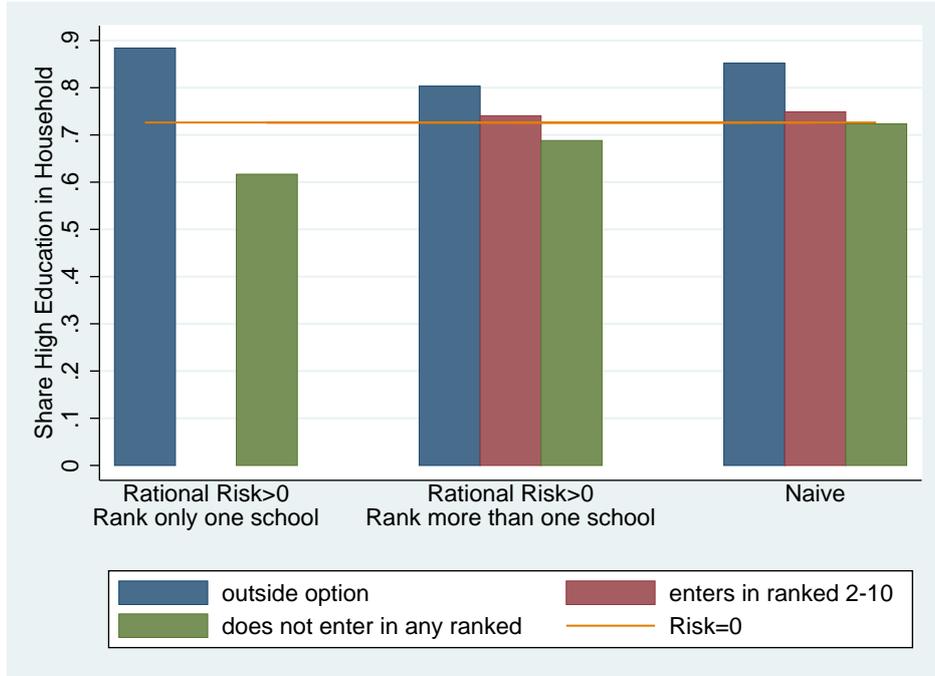


Figure 5: Household Education and Final Outcomes

Notes: Final outcomes as defined in Table 11. High education is defined as completed secondary education or more using the census data. Population: Baseline population who rank first a school in Barcelona, with  $Risk_{i_s1t} > 0$ , and who do not enter in school ranked first. Source: Barcelona primary school applications data set matched with school enrollment data set and Catalan population census, 2005-2010.

## 8 Conclusions

School choice is one of the most hotly debated policies in education. This paper provides empirical evidence that challenges the underlying presumption that implementation of school choice implies school choice *in equilibrium*. The evidence also sheds new light on the two main issues of this debate: the matching between children and schools as well as students' educational outcomes, and equity concerns.

We exploit an unexpected change in the definition of neighborhood in the city of Barcelona, which provides an exogenous change in the set of safer schools and allows us to analyze the extent to which preferences affect the submitted list, keeping housing decisions constant.

This paper finds that the final allocation that results from the implementation of choice under the B-mechanism with priorities for the neighborhood to resolve overdemands is, to a large extent, determined by priorities, not by preferences. In other words, implementation of school choice does not imply school choice *in equilibrium*. Priorities typically include a criterion based on residence. Safety induces families to respond to priorities, since they fear being assigned to a worse school if they apply for their most preferred school. This implies that moving from a purely residential-based assignment to the implementation of the BM with priority for neighborhood seems to leave the majority of families unaffected. We find two exceptions: those who have an outside option, e.g., a private school, that allows them to take risks without suffering the consequences. These families come from substantially higher levels of education. The other exception includes families

with lower levels of education who either do not anticipate the consequences of their risk taking or have little to lose. A seat in their neighborhood school is guaranteed because their neighborhood school is never oversubscribed.

An important literature has evaluated the impact of school choice on school and individual performance (see footnote 2). More recently, the literature has exploited the implementation and expansion of choice within districts implemented through the B-mechanism. Our paper suggests that, depending how the choice was implemented, their results may underestimate the potential effect of preferences that determine the final allocation. More generally, it also suggests that any large-scale choice program with limited school capacity should carefully consider how overdemand is resolved to ensure that parents' preferences determine the final allocation.

The debate on school choice has generated concern about possible equity problems that stem from the fact that sophisticated parents can benefit from the existence of naive ones – see Pathak and Sönmez (2013). We identify new dimensions along which the B-mechanism generates inequalities by using school enrollment and education data. The group formerly identified as naive now includes two differentiated groups: 1) parents with low levels of education that are harmed by the system by being overly unassigned to any school of their ranking and 2) parents with high levels of education who can take high risk by applying for overdemanded schools, opting for a private school if they do not get their preferred school. The second group can now be identified as advantaged by the system: its initially diagnosed naivete can now be rationalized as a result of the presence of an outside option. Abdulkadiroğlu and Che (2010) suggest that the BM will provide efficiency gains by having parents who value relatively more a given school apply to it. In their model, there are neither priorities nor private schools. With the existence of private schools, those who can afford them incur a smaller risk when applying for the best schools and therefore will apply to them and get assigned to them more often, as suggested by Calsamiglia, Miralles, and Martínez-Mora (2013). Thus, for those who can afford a private school, choice implies having increased access to the best schools. The potential efficiency gains from BM are reduced by the presence of priorities and private schools in the city, despite the fact that only 4 percent of the schools are private.

Our findings are important because theoretical analysis has not been conclusive and the empirical challenges are large. By exploiting the change in neighborhood priorities that occurred in Barcelona, we can provide novel evidence of the adequacy of the B-mechanism and urge policy-makers to question the role that priorities ought to play in the final assignment.

Having understood how parents choose in the B-mechanism with priorities, the next question is to compare alternative allocation mechanisms in terms of welfare. We undertake such analysis in Calsamiglia, Fu, and Güell (2014).

## 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.
- 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., G. Haeringer, and F. Klijn (2010). Constrained school choice: An experimental study. *American Economic Review* 100(4), 1860–1874.
- Calsamiglia, C. and A. Miralles (2012). All about priorities: No school choice under the presence of bad schools. Barcelona GSE Working Paper No 631.
- Calsamiglia, C., A. Miralles, and F. Martinez-Mora (2013). Designing school choice in a Tiebout model. Mimeo, Universitat Autònoma de Barcelona.
- Chen, Y. and T. Sönmez (2006). School choice: An experimental study. *Journal of Economic Theory* 127(1), 202–231.
- 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.
- 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.
- 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](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 intrinsic motivation and academic outcomes. NBER Working Paper No. 18324.
- He, Y. (2012). Gaming the boston school choice mechanism in beijin. Working Paper, Toulouse School of Economics.
- 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 University.
- 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: 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 apply to 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 4).

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 applying to these obtain 15 points rather than 10 points). Still, we find an important decline in the demand when considering all NN schools (see Table 8).

- 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 A2). Again, if anything, this would imply that after 2007 it was safer to apply 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.<sup>49</sup>

## B Appendix: Applicants to work-neighborhood-schools

As explained in Section 3.1, 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 A1 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 A1: Applicants to a school in the work neighborhood

	2006	2007	2008	2009	2010
Fraction applicants	5.41	4.38	3.84	3.90	4.08

Population: Applicants without a sibling in school ranked first that obtained 20 points because of neighborhood. For 2005, only information on aggregate points is available. Source: Barcelona primary school applications data set, 2005-2010.

## C Appendix: Address identification

As explained in Section 3.1, 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.<sup>50</sup> However, addresses in the school applications can have typos. We were able to identify around 91.2 percent of the addresses in our population of interest (i.e., live in Barcelona and have no sibling in the school ranked first). In this appendix we analyze whether there is any selection issue around the address identification.

Table A2 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 99.70 percent and 99.27 percent, respectively, for these two sub-populations. Also, when considering all the population that lives in Barcelona (with and without siblings in school ranked first) the relevant p-value becomes 0.0007. Therefore, in general we are not concerned about any selection issues related to the identification of the families' address.

<sup>49</sup>In the data, only one variable captures both large families and single parents. Between 2009 and 2010, it increased from 4.6 percent to 7.6 percent, which gives an indication of the incidence of single parent families in 2010.

<sup>50</sup>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 Figure A1, 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.

Table A2: Differences among population with and without identified address

	All	Identified Address	Not Identified Address	P-value
% Low income family	0.33 (0.058)	0.30 (0.054)	0.73 (0.085)	0.000
% Disability <sup>(a)</sup>	2.30 (0.150)	2.30 (0.150)	2.33 (0.151)	0.914
% Large family	4.10 (0.198)	4.06 (0.197)	4.51 (0.208)	0.164
% Chronic illness <sup>(b)</sup>	2.04 (0.141)	2.04 (0.141)	2.09 (0.143)	0.844
Number priority points	28.320 (8.130)	28.343 (8.062)	28.072 (8.799)	0.041
Number of observations	46,856	42,734	4,122	

Population: Applicants that live in Barcelona without a sibling in school ranked first (All) with identified address (Identified Address) or without identified address (Not Identified Address). Notes: <sup>(a)</sup> Disability refers to “Applicant, one of the parents/tutors or a sibling is disabled.” <sup>(b)</sup> 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, 2005-2010.

## D Appendix: 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 7. Please note that in order to calculate demand and capacity, we use the *entire* population (except those who apply late or submit 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.<sup>51</sup>

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 choose. 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  $CA_{st} > (\text{number of children assigned}_{st})$ , 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 7 because fewer children are assigned to the school than its capacity. (iv) Fourth, if  $CA_{st} < CM_{st}$  and  $CA_{st}$  is not

<sup>51</sup>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 1 of our online appendix).

a multiple of 25 and  $CA_{st} < (\text{number of children assigned}_{st}) < CM_{st}$ , 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 3.2, 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 A3 shows the share of semi-public schools that are underdemanded ( $Demand_{s1t} < Capacity_{st}$ ), 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 A3: Capacity and demand of schools

	2005	2006	2007	2008	2009	2010
<u>Public Schools:</u>						
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
<u>Semi-Public Schools:</u>						
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

Population: Baseline population that applies to schools in Barcelona. Source: Barcelona primary school applications data set, 2005-2010.

## E Appendix: 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,<sup>52</sup> 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

<sup>52</sup>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.

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<sup>53</sup> or if in the case that the address was different, the child had a very common name.<sup>54</sup> 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.<sup>55</sup> (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 randomly.<sup>56</sup>

## F Appendix: 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

---

<sup>53</sup>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.

<sup>54</sup>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>.

<sup>55</sup>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).

<sup>56</sup>There is only one child under this situation.

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 that the share of non-missing observations for the different values of the variable  $Risk_{is1t}$  (as defined in Section 7) 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  $Risk_{is1t} = 1$ .

## F.1 Education variables from census and municipal register data sets

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. All of our analysis in Section 7 is done using the education variable from the census. Results using the education variable from the municipal register are qualitatively the same and available in Section 3 of our online appendix.

## G Appendix: Neighborhoods before 2007

As explained in Section 5.1, before 2007, the city was divided into *fixed* neighborhoods. For semi-public schools, the neighborhood coincided with the administrative district (see Figure 1). For public schools, the neighborhood was a smaller area within the administrative district.

An example will be useful. For instance, consider the administrative district 1 in Figure 1. Figure A1 plots the neighborhoods within administrative district 1 (as well as some parts of the adjacent districts). As can be seen within administrative district 1, there are 4 neighborhoods marked with numbers 1 to 4 (and colored yellow, green, light blue, and red). The schools within each neighborhood are marked as dots (in the same color as the neighborhood).

Since the amount of schools within a neighborhood could vary substantially for different neighborhoods, some neighborhoods were defined so they could overlap to compensate for the variation in the density of

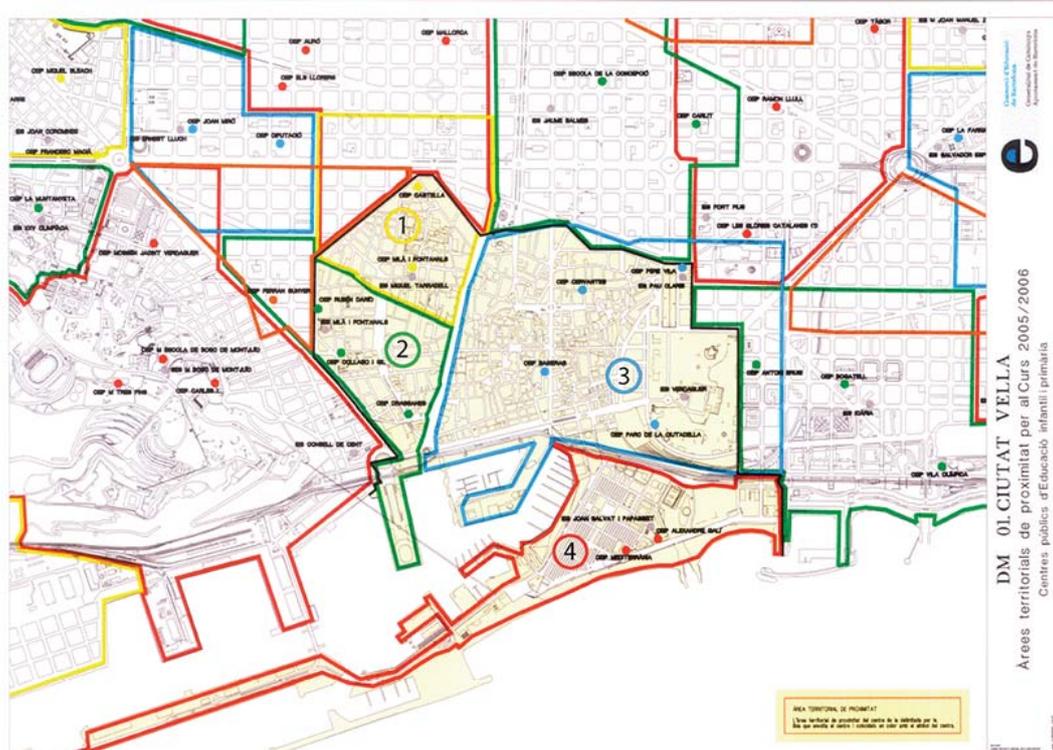


Figure A1: Neighborhoods within administrative district 1

Notes: Primary schools of each neighborhood are the colored dots (colored with the same color as the neighborhood). Schools colored gray are secondary schools. Source: Consorci d'Educació de Barcelona.

schools. For instance, neighborhood number 3 in Figure A1 also includes some families near the border of the adjacent district. Families living in the overlapped areas would have priority in the schools of both areas. For the overall city, there were 52 different neighborhoods within the administrative districts. The maps for these neighborhoods are available upon request from the authors.

## H Appendix: Changes in the set of priority schools

As explained in Section 5.2, 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 A4 displays the number of schools in the priority set for the period 2005-2010. Second, the composition of schools in the priority set changed for some families because the new neighborhood includes nearby schools that are on the other side of the border of the old neighborhood. Table A5 displays change in the composition of the priority set for the period 2005-2010.

Table A4: Change in the size of the set of priority schools

Number of schools	2005	2006	2007	2008	2009	2010
Average	21.82	22.07	7.03	7.62	7.73	7.95
Standard deviation	7.26	7.51	1.41	1.94	1.96	1.99
Number of neighborhoods	52	52	5300	5300	5300	5300

Population: Baseline population. Source: Barcelona primary school applications data set, 2005-2010.

Table A5: Change in the composition of the set of priority schools

	2005	2006	2007	2008	2009	2010
Families with a composition change (%)	81.70	80.91	80.75	82.73	83.36	86.39
For families with a composition change:						
Average of different schools in the set (%)	31.86	31.39	31.68	31.19	31.12	31.17
Standard deviation	(0.16)	(0.16)	(0.16)	(0.16)	(0.16)	(0.16)
Observations	7,772	6,649	6,891	7,198	7,132	7,085

Population: baseline population. Source: Barcelona primary school applications data set, 2005-2010.

## I Appendix: Final outcomes and outside option

As explained in Section 7, one of the possible outcomes for families with  $Risk_{is1t} > 0$  that do not enter the school they rank first is that they enroll in a school outside the Barcelona public system. In other words, these are families with an *outside option*. We identify the families by either (i) directly observing that they enroll in schools outside the Barcelona public system (for which we *do* happen to have their enrollment information) or (ii) indirectly by noticing that we have their school application information but not their enrollment information. One might be concerned that even if we have undertaken a very thorough matching process (as described in Appendix E), it is still not perfect. However, Figure A2 shows that this is unlikely to be the case. When we exclude families that we did not manage to match in the school enrollment data set, we obtain the same qualitative results as in Figure 5 for the whole population of interest.

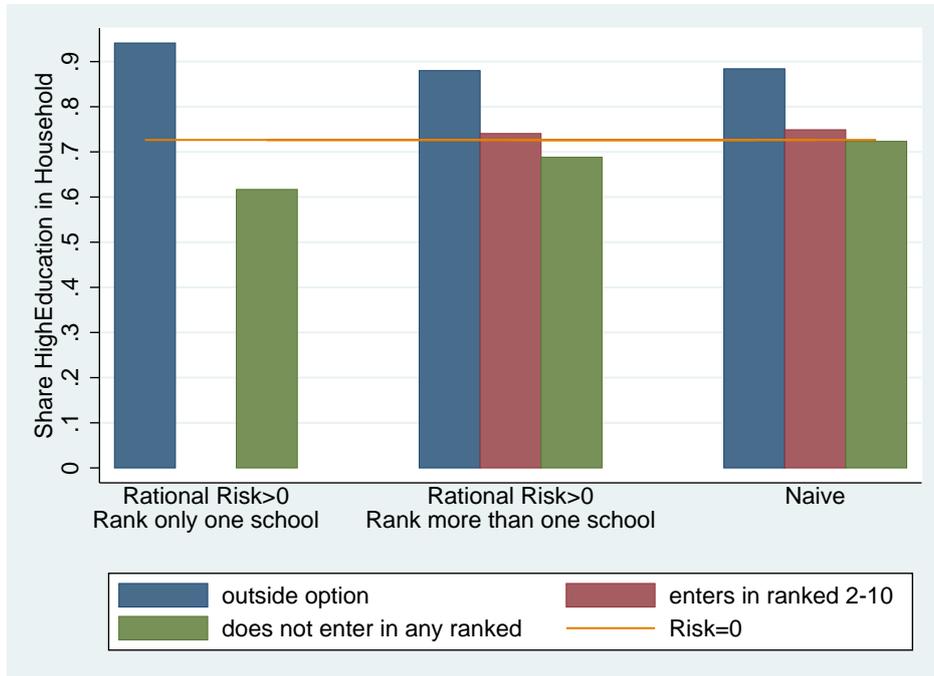


Figure A2: Household Education and Final Outcomes, outside option directly identified

Notes: Final outcomes as defined in Table 11. High Education is defined as completed secondary education or more using the census data. Population: baseline population who rank first a school in Barcelona, with  $Risk_{is1t} > 0$  who do not enter in school ranked first and that has been matched with school enrollment data. Source: Barcelona primary school applications data set matched with school enrollment data set and Catalan Population census, 2005-2010.