Peer Networks and School Choice under Incomplete Information

Andrew Dustan*
Department of Economics
Vanderbilt University

May 6, 2015

Abstract

I present evidence that peer networks exert a strong influence on students' school choices, consistent with a model where students are risk-averse and thus prefer familiar schools. The allocation mechanism used by Mexico City's public high school choice system generates exogenous variation in older siblings' school assignment. The average effects of older sibling admission on the probabilities of choosing both the sibling's school and distinct but observably similar schools are large and positive, even when the siblings are too far apart in age to attend school together. This change in stated preferences affects admissions outcomes, including assignment to elite schools.

JEL Codes: I20, I25, O15

*E-mail: andrew.dustan@vanderbilt.edu. I thank Alain de Janvry, Elisabeth Sadoulet, and Fred Finan for their guidance and Ben Faber, Jeremy Magruder, and Ted Miguel for helpful comments. Participants at various UC Berkeley seminars, the Northeast Universities Development Consortium Conference, the Pacific Conference for Development, and several other seminars provided useful feedback. Roberto Peña Reséndiz supplied the COMIPEMS data and offered excellent insight into the school selection process, and Rafael de Hoyos afforded access to the ENLACE data. This research was funded in part by grants from the Institute of Business & Economic Research and the Weiss Family Fellowship through the Center for Effective Global Action.

1 Introduction

Many education systems allow students and their families some degree of choice in which school they will attend.[1] A key rationale for choice policies is that they allow students to leverage private information about student-school match quality—the interaction between school attributes and student preferences—by choosing the school that best caters to their own preferences and constraints. Students have incomplete information about schools, however, which may have profound effects on choice behavior. As a recent example, Hoxby and Avery (2012) observe that low-income high-achievers in the United States rarely apply to selective colleges, a phenomenon that they attribute partially to incomplete information about how well selective colleges would suit them.

If we think of the student as a Bayesian learner, then he uses information about a particular school in two ways. First, new information allows him to update his expectation of match quality with that school. This channel has been studied extensively in the school choice literature, reviewed below. Second, and thus far unstudied, is that information makes the student's belief about match quality more precise. If students are risk-averse, the uncertainty-reducing value of information makes students more likely to choose schools about which they are well-informed. A simple model of Bayesian learning about experience goods, following Roberts and Urban (1988) and Erdem and Keane (1996) and presented formally in the appendix, predicts that the expected effect of the first channel is zero for students with unbiased prior beliefs about match quality.[2] In contrast, the expected effect of the second channel is strictly positive because new information always increases the precision with which match quality is known. Thus new information about a school, in expectation, increases the probability with which students choose it. Furthermore, if that information is generalizable to observably similar schools, then demand is expected to rise for those schools as well.

If the quantity of information that a student has about each school is an important

[1]Recent work in economics analyzing school systems with formal choice mechanisms includes Abdulkadiroglu et al. (2012) and Dobbie and Fryer (2011) for the United States, Clark (2010) for the United Kingdom, Ajayi (2012) for Ghana, Lucas and Mbiti (2012) for Kenya, de Hoop (2012) for Malawi, Jackson (2010) for Trinidad and Tobago, Pop-Eleches and Urquiola (2013) for Romania, Lai et al. (2011) and Zhang (2012) for China, and Dustan et al. (2015) for Mexico, among others.

[2]Other studies relating to experience goods have used similar models, for example Johnson and Myatt (2006) and Crawford and Shum (2005).

determinant of choice, then the student's social network may be a crucial determinant of choice behavior because peers provide information about some schools but not others. Students learn about schools from older peers who attend them. Consequently, beliefs about match quality should be systematically more precise where the peer network is denser. This implies (on average) stronger preferences for schools attended by older peers, even if peers do not have a direct positive effect on match quality. Hoxby and Avery's (2012) observation regarding the application behavior of low-income high-achievers may be partially explained by a dependence of choice on information from peer networks, as they find that such students often "have only a negligible probability of meeting a... schoolmate from an older cohort who herself attended a selective college" (p. 2).

This paper shows that school-specific information originating from the peer network—specifically, older siblings—significantly affects school choice behavior and, as a consequence, admissions outcomes. I analyze the relationship between older sibling school assignment and younger sibling choice and admission outcomes using detailed student-level data for nearly a million students from fourteen years of Mexico City's public high school choice system. The assignment mechanism generates exogenous variation in the school assignment of older siblings, overcoming the well-known problems with causally interpreting correlated behavior in peer groups (Manski 1993 and 1995). This unified choice-based allocation system determines assignment priority solely on the basis of an exam score. A sharp regression discontinuity design is thus employed: given a group of older siblings who want to attend a certain school, some score barely high enough to be admitted and others score barely too low and must attend another school. This variation in assignment near the admission cutoff is used to identify the effects of older sibling assignment in both reduced form and discrete choice models of school choice.

The empirical results show that students prefer schools attended by their older siblings. Using the estimates from the discrete choice model, I find that students are willing to increase their round-trip commute by an average of eight kilometers per day in order to attend a school to which the older sibling was admitted, which is valued at $936 over the course of high school. This effect is not driven by the obvious explanation that it is convenient or beneficial for the family to have two children attending the same school. Older sibling admission to a particular high school increases the revealed preference for that school even when the siblings are far enough apart in age that the older sibling no longer attends high school. Furthermore, having an older sibling admitted to a school increases revealed preference for other campuses belonging to that school's subsystem, within which individual schools throughout the city share many attributes such as curriculum and vocational orientation. This suggests that students generalize the knowledge obtained about a peer's school when evaluating other schools within the same subsystem. There is also evidence, although not strictly causal, that revealed preference for a school increases much more when the older sibling experiences a positive academic outcome there. Taken together, these results support the view that students prefer schools about which they have more information and use information from their peer network to update beliefs about match quality.

The altered school choices induced by sibling admission lead to significant changes in the admissions outcomes of the younger sibling. Students are more likely to be admitted to their older siblings' assigned schools, as well as to a school within the same subsystem. Among the students least likely to apply to the set of Mexico City's elite public high schools, sibling admission to an elite school substantially increases the probability of both applying and gaining admission to elite schools, suggesting a role for information in encouraging qualified students to apply to high-quality schools.

Existing empirical literature on school choice under incomplete information does not incorporate risk aversion into student preferences. Hastings et al. (2009) provide a model of school choice where students trade off academic quality with attributes such as proximity. In their model, risk-neutral students optimize with respect to expected quality without regard for the precision of this belief. Empirical studies on the effect of information provision on school choice do not model risk aversion, either, because information is enriched for all schools simultaneously and thus does not induce between-school variation in the amount of information available to students. Hastings and Weinstein (2008), for example, demonstrate that providing information on test score aggregates to low-income families in the United States increased the likelihood of choosing high-performing schools. Related studies by Koning and van de Wiel (2010) in the Netherlands and Friesen et al. (2012) in Canada come to similar conclusions, while Mizala and Urquiola (2013) find no effect of publishing a quality measure in Chile.

Research quantifying the causal effects of peers on school choice is scarce, perhaps because it is difficult to isolate exogenous variation in the schools attended by a student's peer network. The sociology and education literatures have instead studied the effect of social learning on school choice in a qualitative framework. Most notably, Ball and Vincent (1998) find that, for primary schools in the United Kingdom, parents use their social networks (the "grapevine") to obtain specific, detailed information about schools and their likely fit for their own children. Ceja (2006) finds qualitative evidence that older siblings are an important source of information for Chicana students as they apply to college in the United States. The economics literature has carefully documented correlations, as in Hoxby and Avery (2012) and in Goodman et al. (2014), who find a correlation between college enrollment decisions of siblings in the United States. Closely related to this topic is Joensen and Nielsen (2015), who find that a shock to Danish teens' decision to enroll in advanced high school math and science coursework increased the probability that younger siblings did the same.

The remainder of the paper proceeds as follows. Section 2 describes the public high school choice system in Mexico City, showing that it provides a good context in which to empirically examine school choice under incomplete information. Section 3 explains the data. Section 4 gives the reduced form regression discontinuity method and results, while Section 5 lays out the discrete choice model and corresponding results. Section 6 provides validity checks for the empirical design and Section 7 concludes with policy recommendations.

2 High school choice in Mexico City

2.1 The COMIPEMS assignment mechanism

Prior to 1996, the nine major public high school subsystems in Mexico City controlled their own independent admissions processes.[3] Students applied to schools in one or more of these subsystems, waited to learn where they had been admitted, and then withdrew from all schools except their most-preferred one. In an effort to increase both the efficiency and transparency of this process, the subsystems formed the Comisión Metropolitana de Instituciones Públicas de Educación Media Superior (COMIPEMS) in 1996. Each year, COMIPEMS runs a unified, competitive admis-

[3]The discussion in this section draws on Dustan et al. (2015).

sions process that assigns students across Mexico City's public high schools on the basis of students' preferences and the results of a standardized exam.

The COMIPEMS admissions process is as follows.[4] In late January, students in ninth grade—the final year of middle school—receive informational materials about the admissions process. These materials include a list of all of their "educational options," which in most cases are schools but can also be specific tracks within schools, such as specific vocational education tracks in a technical school. Students then fill out a registration form, demographic survey, and list of up to 20 educational options, ranked in order of their preference. These forms must be submitted in late February or early March, depending on the student's family name. In June of that year, students take a standardized exam consisting of 128 multiple-choice questions, covering both subject-specific material from the public school curriculum and more general mathematical reasoning and language areas.

In July, the assignment process is carried out by the Centro Nacional de Evaluación para la Educación Superior (CENEVAL).[5] First, the school subsystems report the maximum number of seats available to incoming students. Second, all students who did not successfully complete middle school or scored below 31 of 128 points are discarded. Third, all remaining students are ordered by their exam score, from highest to lowest. Fourth, a computer program proceeds sequentially down the ranked list of students, assigning each student to his highest-ranked option that still has a seat remaining.[6] The process continues until all students are assigned, with the exception of students who scored too low to enter any of their listed options. Later in July, the assignment results are disseminated to students. Through 2011 this primarily happened in the form of a printed gazette sold at newsstands, although a system that sends personalized results via text message has become more popular over time. At that time, students who were eligible for assignment but were left unassigned during the computerized process because they scored too low for any of their choices may

[4] The timing of each step is given for the 2011 competition.

[5] CENEVAL is independent of COMIPEMS and its constituent school subsystems. This process is carried out by computer in the presence of representatives from all subsystems and external auditors from a large international accountancy firm.

[6] In the instance that two or more students have the same score and highest-ranked available option, but there are fewer remaining seats than the number of tied students, the assignment process pauses and representatives from the corresponding subsystem must decide to either admit all tied students or none of them.

choose a schooling option from those with seats remaining.

2.2 Student decision-making under the COMIPEMS mechanism

Students have considerable information about basic school attributes when they choose schools, but this information is generic rather than individually tailored. The subsystem membership of each school is known with certainty, and each subsystem has a well-formed public perception. There are two “elite,” university-affiliated subsystems: the Universidad Nacional Autónoma de México (UNAM) and the more technically-focused Instituto Politécnico Nacional (IPN). These are universally understood to be highly competitive, relatively rigorous, prestigious high schools that fill their student capacities before almost all non-elite options. Non-elite subsystems include those with traditional academic curricula and technical subsystems providing academic coursework combined with vocational training for careers such as auto repair and bookkeeping. Even within a subsystem, official information about school-level academic quality is available. Past cutoff scores—the score of the student admitted to the school’s final seat—for each school have been available on the COMIPEMS web site since 2005, and this site is actually browsed by many students because it allows them to easily complete most of the registration process online. Cutoff score and the mean score of admitted students are almost perfectly correlated, so students have access to an excellent proxy for mean peer ability. The combination of subsystem reputations and information about peer quality ensures that students are at least somewhat informed about general school attributes, though they may lack more specific details that affect the idiosyncratic match between the student and school.

When asked about the choice process, administrators and students almost universally claim that it is the student, rather than the parents, who decide on the schools they want to list and the order in which they are listed. Students often construct their rankings in the following way, similar to how United States students choose colleges (see Hoxby and Avery (2012), for example). First, they decide whether they would like to attend a high school in either or both of the two elite subsystems. If a student decides to apply within either or both subsystems, he lists some number of elite schools as his top choices. There are 30 elite schools (16 IPN and 14 UNAM), meaning that even within an elite subsystem, students face a wide variety of options. Following the elite schools, if any, he lists various non-elite schools (from about 600 options in most years), which offer a better chance of admission.[7]

Two aspects of the COMIPEMS assignment mechanism make the student's ranking quite informative about true preferences. First, the mechanism is equivalent to the deferred acceptance algorithm proposed by Gale and Shapley (1962), so it induces truth-telling by students.[8] Under such mechanisms it is never optimal to list a less-preferred school before a more-preferred school, regardless of the limit on how many options can be listed. Second, the ability to rank up to twenty options means that few students actually fill up their entire preference sheet; students generate a satisfactory choice portfolio without confronting the space constraint.[9] There is no strategic disadvantage to choosing a school at which the student has a small *ex ante* probability of admission, both because the number of options allowed is high and because the assignment algorithm does not punish students for ranking unattainable schools.

3 Data and sample construction

This section describes the Mexico City public high school admissions data and the sample construction that forms the basis for the regression discontinuity design.

3.1 Data description

This paper uses administrative data compiled by COMIPEMS for fourteen admissions cycles, from 1998 to 2011. For each student who registered for the exam, the database contains basic demographic information including the student's full name, date of birth, phone number, address, and a unique middle school identifier along with the grade point average attained there; the full list of up to 20 ranked school preferences; a context survey, completed by the student, including information about parental education, family composition, and other topics; and assignment results, including the student's exam score and the school assigned during the computerized allocation process.

---

[7]Most students live within a reasonable commuting distance of many schools, as illustrated in Appendix Figure A.1.

[8]See Dubins and Freedman (1981) and Roth (1982). This particular mechanism is referred to as a student-proposing deferred acceptance mechanism, which is discussed in Abdulkadiroglu and Sönmez (2010).

[9]Choosing the optimal portfolio of schools is a complex problem if listing choices is costly (e.g. time cost or opportunity cost due to a limited number of allowed choices), as mentioned by Ajayi (2012). Chade and Smith (2006) model a similar portfolio choice problem and derive its solution.

The data do not contain any explicit information on peer network structure and, since middle schools in Mexico City are quite large and neighborhoods are not geographically isolated, neither can be used to construct a useful proxy for the student's network. The data do, however, allow for the identification of siblings within a family, which is useful for two reasons. First, the strength of the peer relationship is likely to be strong compared to most classmates and neighbors. Second, the constant interactions between siblings within the home make it probable that the student learns a significant amount about the details of the school attended by his older sibling and how that school might fit his own tastes. The analysis is limited to sibling pairs where the older sibling attended a public middle school and the younger sibling was taking the exam in the final year of middle school. If the older sibling is observed taking the exam more than once, his first attempt is used.

To measure whether the older sibling graduated or dropped out of high school (a proxy for whether the peer signal transmitted to the younger sibling was good or bad), the COMIPEMS database is merged via national ID number (CURP) with a database from the national 12th grade exam, called the ENLACE Media Superior. This exam is only given to students who are on track to graduate at the end of the academic year, so it is a good proxy for graduation.[10] Unfortunately, this exam was only administered starting in the spring of 2008, and the database used in this paper contains results from 2008 to 2010, corresponding to students taking the COMIPEMS exam in 2005-2007. Thus the part of the analysis using this graduation measure is limited to younger siblings of these cohorts, which limits sample size. The larger and more demanded of the two elite subsystems, the UNAM, does not administer the ENLACE exam so graduation data is missing for students assigned there. This further limits the sample size when the graduation measure is used.

The demographic information is used to match siblings with each other in the following way. First, potential siblings are identified if they have the same paternal and maternal family names and either 1) have the same phone number or 2) live in the same postal code and attend the same middle school. From this pool of potential matches, sibling pairs are discarded if 1) the students state that they have different numbers of siblings; 2) the students do not report a birth order that makes them

[10]For more details on the ENLACE and how it relates to graduation, see Dustan et al. (2015).

the closest siblings in the family (e.g. first- and second-born);[11] 3) the students were born fewer than nine months apart or more than five years apart, the latter because it is unlikely that consecutive births five or more years apart represent a true match; or 4) the older student took the exam after the younger one. If one student matches with two potential older siblings, the match based on the shared phone number is used.

This matching process locates 455,375 sibling pairs in a population of 3,423,052 students. Columns 1 and 2 of Table 1 give a description of demographic, academic, and school choice variables for the full sample of students and for the matched younger siblings, respectively. The matched younger siblings are quite similar to the full sample. The average student ranks 9 school choices, which is similar across samples. Almost two thirds of students select a school in one of the two elite subsystems as their first option, but only one in five are admitted to one. On average, students choose a school almost 8 km away as their first option, measured as a straight line from the center of the student's home postal code to the school.[12] Siblings are, on average, 2.6 grade years apart and have fairly similar school preferences: 34% of sibling pairs select the same school as their first choice. Only 45% took the ENLACE exam, similar to the official graduation rate in Mexico City.

3.2 Overview of empirical strategy and sample definition

The COMIPEMS school assignment mechanism provides exogenous variation in older siblings' school assignment because, conditional on the older sibling's ranking of schools, his assignment depends solely on his exam score. This permits the use of a sharp regression discontinuity (RD) design, similar to RD designs used in prior work investigating the academic effects of school assignment in exam-based allocation regimes.[13] The basic idea behind this design is to define, for each school, the sample of older siblings who were either marginally admitted or marginally rejected from that school, and then compare the choices and outcomes of the younger siblings in the marginally admitted and marginally rejected groups. The rest of this subsection

[11] This is done so that the estimated effect of older sibling of admission does not include an indirect effect through the influence on a middle sibling's behavior

[12] Postal codes are very geographically specific in Mexico City. Students in the sample belong to more than 2,800 postal codes.

[13] See Pop-Eleches and Urquiola (2013), Abdulkadiroglu et al. (2012), Dobbie and Fryer (2011), Clark (2010), Jackson (2010), de Hoop (2012), and Dustan al. (2015).

gives the procedure for defining the "marginal" sample of older siblings for use in the RD analysis.

The assignment process results in hard cutoff scores for each school that filled all of its seats and thus had to reject some students; this cutoff is equal to the lowest score among all admitted students. Define this cutoff as c_j for school j. (The cutoff score for a given school varies across years, but for notational simplicity in the present discussion I assume there is only one year of data.) If school k is ranked before j on student i's preference list, including if j is unlisted, we write $k \succ j$. Denote the student's exam score as s_i. Then marginal students for school j are those who:

1. listed school j as a choice, such that all schools preferred to j had a higher cutoff score than j (otherwise assignment to j is impossible): $c_j < c_k, \forall k \succ j$, and

2. had a score sufficiently close to j's cutoff score to be within a given bandwidth w around the cutoff: $-w \leq s_i - c_j < w$.[14]

This marginal group includes students who were rejected from j ($s_i < c_j$) and those who scored high enough for admission ($s_i \geq c_j$). A student may belong to more than one school's sample of marginal students. Not all students scoring high enough for admission are actually assigned to j; some score sufficiently high for admission to some $k \succ j$. Figure 1 plots the probability of being assigned to the cutoff school as a function of $s_i - c_j$ and verifies that the jump in probability of admission to the cutoff school at the cutoff score is exactly 1. This probability falls monotonically with score to the right of the cutoff, as higher-scoring students are admitted to more-preferred schools.

One more restriction is placed on the sample, not to fulfill the assumptions of RD but to ease interpretation of the admission effect. Students who would be unassigned to any school upon missing j's cutoff by a point are omitted. Such students did not list any school with a cutoff score equal to or less than c_j. We do not know if the unassigned students later chose a school from those that did not fill up or if they did not enroll at all. Our focus is on the effect of a sibling being assigned to one school or another, rather than getting into any school or going unassigned.

[14] The second inequality is strict because the score variable is discrete, so this definition includes w score values too low to be admitted and w values high enough to be admitted.

4 Reduced form regression discontinuity analysis

The reduced form analysis provides clear, easily-interpreted evidence about the effects of sibling assignment on school choice and admission outcomes. Much of the logic from this analysis will be applied when estimating the discrete choice model as well.

4.1 Method

For all regressions in this paper, exam score is centered to be 0 at the school's cutoff score, which may be different in each year t: $\tilde{s}_{ijt} \equiv s_i - c_{jt}$. The basic RD specification for a single school j in year t is:

$$y_{ijt} = \delta_{jt} admit_{ijt} + f_{1jt}\left(\tilde{s}_{ijt}\right) + admit_{ijt} f_{2jt}\left(\tilde{s}_{ijt}\right) + \mu_{jt} + \varepsilon_{ijt}$$

where y_{ijt} is the outcome of interest, $admit_{ijt}$ is a dummy variable for whether $\tilde{s}_{ijt} \geq 0$,[15] $f_{1jt}\left(\tilde{s}_{ijt}\right)$ and $f_{2jt}\left(\tilde{s}_{ijt}\right)$ are polynomials in exam score approximating the unobservables that vary with score, and ε_{ijt} is an error term. In our case, y_{ijt} is an outcome for the younger sibling, such as choosing school j as his first option, while the explanatory variables are from the older sibling, since it is the admission outcome of the latter that is hypothesized to affect the choices of the former. The parameter δ_{jt} is the local average treatment effect of the older sibling's admission to j in year t on the younger sibling's outcome for older siblings close to the cutoff, compared to the counterfactual in which the older sibling is rejected from j and admitted to the most-preferred school that would actually accept him.

There are many schools and many exam years, so it is necessary to combine the information from all oversubscribed schools in order to make statements about the average effect of admission. To do this, I stack the RD samples of all oversubscribed school-years and estimate them jointly. It would be preferable to include different functions f_{1jt} and f_{2jt} for each school or school-year, similar to Abdulkadiroglu et al. (2012) who include different functions for each school. But the very large number of schools makes this infeasible in most specifications, so I include only one set of polynomials, as in Pop-Eleches and Urquiola (2013). Including cutoff school-year

[15] The variable $admit_{ijt}$ is equal to 1 both for students actually assigned to j and students who scored high enough to be admitted to a more-preferred school.

fixed effects, the stacked specification is:

$$y_{ijt} = \delta admit_{ijt} + f_1\left(\tilde{s}_{ijt}\right) + admit_{ijt} f_2\left(\tilde{s}_{ijt}\right) + \mu_{jt} + \varepsilon_{ijt} \tag{1}$$

The parameter δ is now the local average treatment effect of admission across all cutoff school-years.

Both local non-parametric and global parametric RD models are estimated. The non-parametric model is a local linear regression that weights observations using the edge kernel. Bandwidth selection is performed using the procedure proposed in Imbens and Kalyanaraman (2012). Presented as a complement to the non-parametric approach, the parametric model fits a fifth-order polynomial to the entire RD sample. Lee and Card (2008) show that when the running variable is discrete, as is the case here, standard errors should be clustered at the level of the running variable. This results in relatively few clusters in the present application, which I address with two approaches. First, I estimate standard errors accounting for clustering on both the older sibling and running variable dimensions, using a t-distribution with $(G-1)$ degrees of freedom for hypothesis testing, where G is the number of values of the running variable in the regression.[16] Second, I account for clustering with respect to the running variable using the wild cluster bootstrap from Cameron et al. (2008), implemented with Rademacher weights. Bootstrapped p-values for the coefficients of interest under the null hypothesis of zero effect are reported. These bootstrapped p-values are sometimes less conservative than those corresponding to the estimated standard errors, so inference is always based on the most conservative p-value.

4.2 Average effect of older sibling admission on school choice

The RD estimates give consistent causal evidence that students are more likely to apply to a school and to rank it highly if an older sibling was admitted there. Table 2 presents the estimated effects, accompanied by graphical evidence in Figure 2. Column 1 shows that for the local linear regression using the optimal bandwidth, older sibling admission to the cutoff school increases the probability of choosing the school as first choice by 7.6 percentage points. This estimate is large compared to the corresponding sample average of 17% choosing the cutoff school. Admission increases the probability of choosing the cutoff school as any choice by 9.8 percentage points

[16] Cameron and Miller (2015) provide guidance on estimating variances with few clusters.

(p.p.), compared to a sample mean of 59%. Effects on the probability of choosing the school immediately below the cutoff (the school to which a student is admitted if he misses admission to the cutoff school by one point) are similar: 4.7 p.p. for first choice probability and 10.5 p.p. for choosing the school at all. Estimates from the global specifications in columns 5 through 8 are almost identical to the local linear regression results, as will be the case for most results in this paper.

The sibling admission effect is persistent across schools in both elite and non-elite subsystems. Table 3 and corresponding Figure 3 divide the cutoff schools into elite and non-elite groups. The admission effect on first choice probability for elite cutoff schools is 11.3 p.p, while the effect on listing the cutoff school at all is 11.7 p.p. The effect for non-elite cutoff school is presented in columns 3 and 4. In column 3, the dependent variable is a dummy for whether the cutoff school was the younger sibling's first *non-elite* choice. This is because most students choose an elite school as their first choice, so that most adjustment in non-elite revealed preferences takes place lower in the choice list. The effect on first non-elite choice probability is 9.4 p.p., while the effect of choosing the cutoff school at all is 7.9 p.p.

The admission effect for most schools is positive. Figure 4 shows the distribution of estimated admission coefficients, obtained by estimating the RD specification separately for each school. Panel A gives the distribution of admission effects on first choice demand for elite schools only, which have large corresponding sample sizes and thus fairly precise estimated effects. All but two of the 30 schools have positive estimated effects of admission. Panel B gives the distribution of the effect on first non-elite choice for non-elite cutoff schools. Here, estimation error overstates the variance of the distribution substantially, such that the estimated effect of admission is negative for 30% of schools. To account for the estimation error, I estimate the true variance of the admission effects, following Aaronson et al. (2007).[17] Performing this correction and assuming a normal distribution of admission coefficients, it is estimated that 21% of non-elite schools have a negative admission effect. Hence it appears that the expected utility-increasing channels dominate for most schools.[18]

[17] This is done by subtracting the average estimation error from the variance of the estimated coefficients: $\mathrm{E}\left[\widehat{\boldsymbol{\delta}}'\widehat{\boldsymbol{\delta}}\right] - \mathrm{E}\left[\left(\boldsymbol{\delta}-\widehat{\boldsymbol{\delta}}\right)'\left(\boldsymbol{\delta}-\widehat{\boldsymbol{\delta}}\right)\right]$.

[18] One explanation for this result is that most of the uncertainty about a school comes from imprecise beliefs about idiosyncratic match quality rather than about the school's average level of

The effect of older siblings' school assignment on demand does not appear to be driven by a direct effect of sibling presence on match quality. The most obvious channel through which sibling assignment could affect match quality is that attending school together is convenient for the student or parent, for example in traveling to and from school or attending the same school functions. But the estimated effect of admission is nearly identical between siblings who are close enough in age to attend high school at the same time (two or fewer years apart) and siblings who are too far apart in age to attend contemporaneously. Table 4 shows this result. The estimated difference in admission effects between these two groups is precisely estimated to be close to zero: the point estimate of the differential is 0.6 p.p. with the 95% confidence interval bounded at 1.6 p.p. Thus it does not appear that students choose their siblings' schools simply because they want to attend the same school contemporaneously.[19]

Furthermore, the effect of admission is not confined to demand for the older sibling's school, suggesting that students apply what they learn about one school to others that are observably similar. Admission leads to the student ranking additional schools from the same subsystem, other than the older sibling's exact school. Table 5, accompanied by Figure 5, shows that this is the case. The sample definition here is different than in the previous analysis because it only considers students who would leave their subsystem if rejected from the cutoff school. The counterfactual to admission to the cutoff school in this case is admission to a school in a different subsystem. When the older sibling is admitted to the cutoff school, the younger sibling ranks on average 0.23 more schools in the same subsystem, excluding the older sibling's school, compared to a sample mean of 2.1 schools. The estimated effects are almost identical for the closely-spaced and far-apart sibling samples. Similar results are found for the effect of rejection from the cutoff school on demand for

attributes. In the former case, a similar proportion of students receiving a signal from a particular school would find out that it is better for them than expected and others would find out it is worse, while in the latter case, the surprise to beliefs for students within one school would be highly correlated and would thus result in negative demand effects for some schools.

[19]Furthermore, while I find statistically significant evidence of differential admission effects with respect to whether the siblings are of the same sex (Appendix Table A.1), the differentials are small (about 1 p.p.). If the effects were driven by the desire to attend school with a sibling, we might expect to see a much stronger effect for same-sex sibling pairs. Instead, it seems that boys learn from girls and vice versa.

schools belonging to the subsystem of the school immediately below the cutoff. The admission effect on subsystem demand and the persistence of admission effects for students far apart in age cannot be explained by a direct effect of sibling presence on match quality.

4.3 Effect on school assignment outcomes

Older sibling admission outcomes affect not only the stated preferences of younger siblings, but also their realized school assignments. Table 6, accompanied by Figure 6, shows the admission effect on the younger sibling's assignment. Admission to the cutoff school nearly doubles the probability of the younger sibling's assignment to the same school, an effect of 4.2 p.p. A similar result holds for the school immediately below the cutoff, with an admission effect of 4.1 p.p. Columns 3 and 4 limit the sample to observations where the older sibling has schools in different subsystems above and below the cutoff. Older sibling admission increases the probability of being assigned to any school within the same subsystem by 4.6 p.p.

Perhaps most interesting is the effect that older sibling admission to elite high schools has on younger sibling preference for, and assignment to, elite schools. Among the students least likely to apply to elite schools, older sibling admission has a large effect on both elite application and assignment rates. To show this, I first use a probit model to estimate the determinants of elite first choice among younger siblings of students who applied to elite schools and were rejected. Probit estimates are in Appendix Table A.2.[20] These estimates are then used to generate a predicted probability of elite first choice for all younger siblings of students who were either above or below the cutoff for elite admission. Table 7, Panel A shows estimates of the sibling admission effect on students falling within bins of the predicted elite first choice probability. Column 1 estimates a large 14.2 percentage point increase in the probability of choosing an elite school among students with less than a 60% predicted probability of doing so. As the predicted probability increases, the marginal effect of older sibling admission falls dramatically. Panel B shows similar results for the total number of elite schools selected. These results are consistent with the model: first,

[20]The most important predictor of elite preference is the proportion of the older sibling's middle school cohort expressing elite preference, which is exogenous to the older sibling's assignment outcome. This measure captures the preferences, constraints, and peer network effects in the older sibling's cohort, which are relevant for most younger siblings as they often attend the same middle school or live in the same neighborhoods as these students.

students with high baseline probabilities of choosing elite schools should mechanically have small marginal effects from sibling admission, and second, students with a low baseline probability of admission may have the noisiest prior beliefs about elite schools and thus change their behavior the most due to a new signal.

The consequent effect on elite *assignment* (Panel C) is almost 3.7 p.p. among those least likely to choose an elite school, compared to an estimated zero effect for those most likely to express elite preference. This represents an almost 23% increase in the probability of admission, compared to the rate of elite admission among those with older siblings missing elite assignment by one point (16%). This implies that there exist many students who are capable of elite admission but only apply when they are sufficiently exposed to elite schools through their peer network.

4.4 Effect of good versus bad surprises on school choice

A basic prediction of the social learning model is that a signal's impact on expected utility (and thus demand) depends on the sign and magnitude of the surprise to match quality. The sign and magnitude of surprises are unobserved by the econometrician, but one available proxy is an indicator for whether the older sibling graduates from high school or not. The logic for using this proxy is as follows. One contributor to dropout is a bad match between student and school. That is, there are students who will drop out from some schools but not others. Siblings are often similar in their preferences and abilities, so if the older sibling experiences a negative surprise to match quality (proxied by dropout), this suggests to the younger sibling that the school may not be a good match for him either.

Any estimates of differential admission effects with respect to dropout must be treated as suggestive rather than rigorously causal, because dropout is not randomly assigned (indeed, if it were, it would have no informational content for the student). Consider the following equation that will be estimated:

$$\begin{aligned} y_{ijt} = \delta\, admit_{ijt} + f_1\left(\tilde{s}_{ijt}\right) + admit_{ijt} f_2\left(\tilde{s}_{ijt}\right) + \mu_{jt} + \\ graduate_{ijt}\left\{\alpha\, admit_{ijt} + g_1\left(\tilde{s}_{ijt}\right) + admit_{ijt} g_2\left(\tilde{s}_{ijt}\right) + \nu_{jt}\right\} + \varepsilon_{ijt}, \end{aligned} \tag{2}$$

where $\widehat{\alpha}$ is equivalent to the result from estimating the simple RD equation separately for graduates and dropouts and then taking the difference of the estimated $\widehat{\delta}$ coefficients. If dropout were randomly assigned, then $\widehat{\alpha}$ would give the additional average effect of admission when the older sibling graduates. The problem arises when $cor\left(graduate_{ijt} \times admit_{ijt}, \varepsilon_{ijt}\right) \neq 0$, so that students who are *differentially* more or less likely to drop out when admitted to the cutoff school are systematically more or less likely to be emulated, or have family characteristics that affect the likelihood of choosing the cutoff school.

The empirical analysis addresses the potential issue of endogenous heterogeneous effects in three ways. First, it considers multiple samples and argues that the pattern in the results is consistent with the social learning model in which positive surprises affect demand for a school more positively than negative ones. Second, it controls for the older sibling's middle school grade point average, which is a significant predictor of high school dropout, and its interactions with admission and exam score. Finally, it may be that the sibling admission effect is heterogeneous with respect to the school's graduation rate or other school characteristics, not the sibling's individual graduation outcome. To control for this, separate admission coefficients are estimated for each cutoff school so that the admission-graduation interaction term gives the estimated heterogeneity due to sibling dropout conditional on cutoff school characteristics.

Keeping in mind the caveats associated with using graduation status to proxy for a surprise to match quality, as well as the data limitations in using the graduation data, Table 8 shows that the admission effect is heterogeneous with respect to older sibling dropout. Sample size is a problem, due to the fact that graduation data only exist for older siblings from the 2005-2007 cohorts and that graduation outcomes are missing for students at the UNAM schools, which are highly-demanded as first choices. This necessitates inclusion of all sibling pairs 1 to 5 years apart in age in the estimation sample. A sibling one year below his older sibling still has most of an academic year to learn about his sibling's school, since school begins in the early fall and preference listings are not due until February or March. Although graduation has not occurred yet for the siblings who are 1 or 2 years apart, in Mexico City most dropout occurs in the first year and it should be apparent early on whether match quality was good or bad.

The effect of admission on same-school demand is higher when the older sibling graduates. Panel A, column 1 gives the differential effect of admission on application to the cutoff school with respect to graduation status. On average, admission has a 2.3 p.p. higher impact on first choice preference for the cutoff school when the older sibling graduates.[21] The differential effect is illustrated in Appendix Figure A.2. Column 2 controls for an interaction between admission and older sibling GPA while estimating each uninteracted admission coefficient separately. The estimated differential effect declines slightly to 1.9 p.p.

In order to explore the issue of endogenous differential dropout, Panel B shows estimates of the impact on the first non-elite choice of students whose siblings were near the cutoff of a non-elite school. This, in part, addresses the possibility that students whose older siblings are more able to graduate in the cutoff school are more likely to choose better schools. In particular, we might worry that older siblings able to graduate from elite schools are from families with high academic expectations who push the younger sibling to apply as well. Focusing on the non-elite preferences of students with siblings at non-elite cutoffs, we are likely to mitigate this confounding factor to some degree. The differential effect here is large, 7.1 p.p. compared to a sample mean of 19%. Adding controls in column 2, the estimates remain almost identical.

The evidence for heterogeneity in the effect on demand for other schools in the same subsystem is consistent with the model as well. Panel C shows the differential impact on the number of other schools selected in the cutoff school's subsystem, restricting the sample to cases where the older sibling is at the margin of a subsystem. The point estimates for the differential effect are positive in the local linear and global parametric specifications, and statistically significant in the specifications that include controls and separate admission effects for each school. There is considerable variability between specifications, however. Thus, between the same-school and somewhat weaker subsystem effects, it appears that younger siblings react to signals from siblings with "good" and "bad" outcomes differently, learning about match quality and updating their choice behavior accordingly.

5 Discrete choice model of school choice

In this section, the basic RD design is extended to a discrete choice model of school choice. This approach follows from the social learning model where expected utilities are affected by informative peer signals. It also allows for a natural parameterization

[21]The estimates imply a smaller average effect of admission than did previous tables. This is because the UNAM cutoff schools are missing from the sample, and much of the admission effect on first choice demand comes from the elite UNAM and IPN subsystems.

of the impact of a peer signal: the change in willingness to travel to that school or another school in the same subsystem, which with further assumptions can then be translated into a willingness to pay measure.

5.1 Method

The RD design is incorporated into a discrete choice model by estimating the following equation, expressing expected utilities from each school as a function of school characteristics and older sibling assignment (suppressing time subscripts):

$$\begin{aligned} U_{ij}^{*} = \theta cut_{ij} + \delta \left(cut_{ij} \times admit_i \right) + f_1 \left(\tilde{s}_i \right) cut_{ij} + f_2 \left(\tilde{s}_i \right) \left(cut_{ij} \times admit_i \right) + \\ \underline{\theta} below_{ij} + \underline{\delta} \left(below_{ij} \times admit_i \right) + \underline{f}_1 \left(\tilde{s}_i \right) below_{ij} + \underline{f}_2 \left(\tilde{s}_i \right) \left(below_{ij} \times admit_i \right) + \\ \gamma dist_{ij} + \varepsilon_{ij} \end{aligned} \tag{3}$$

where $cut_{ij} = 1$ when student i's sibling is in school j's cutoff sample and 0 otherwise, $admit_i$ is a dummy for whether the older sibling meets the cutoff score, f_1 and f_2 are linear functions of centered exam score, $below_{ij}$ is a dummy for whether the school is immediately below the student's cutoff school, and $dist_{ij}$ is the distance between student and school. Allowing expected utility to be higher or lower for cutoff schools and the school assigned for scores immediately below the cutoff (through θ and $\underline{\theta}$, respectively), and for this expected utility to vary around the cutoff, δ and $\underline{\delta}$ capture only the discontinuous jumps in expected utility caused by the sibling crossing the cutoff score and being admitted to the school above the cutoff instead of the school immediately below.

Incorporating subsystems into the model is straightforward. For the M subsystems, let $X_j^1, \ldots, X_j^M$ be dummy variables equal to 1 if school j belongs to the corresponding subsystem and 0 otherwise. Define $cutsub_{ij}$ equal to 1 if j belongs to the cutoff school's subsystem and $belowsub_{ij}$ equal to 1 if j belongs to the "below" school's subsystem, 0 otherwise. Adding these variables into the RD specification, we have:

$$
\begin{aligned}
U^*_{ij} = \theta cut_{ij} + \delta \left(cut_{ij} \times admit_i\right) + f_1\left(\tilde{s}_i\right) cut_{ij} + f_2\left(\tilde{s}_i\right)\left(cut_{ij} \times admit_i\right) + \\
\underline{\theta} below_{ij} + \underline{\delta}\left(below_{ij} \times admit_i\right) + \underline{f}_1\left(\tilde{s}_i\right) below_{ij} + \underline{f}_2\left(\tilde{s}_i\right)\left(below_{ij} \times admit_i\right) + \\
\sum_{\ell=2}^{M} X_j^{\ell}\left(\pi^{\ell} + \eta^{\ell} cutsub_{ij} + \underline{\eta}^{\ell} belowsub_{ij}\right) + \\
cutsub_{ij}\left[\phi admit_i + h_1\left(\tilde{s}_i\right) + h_2\left(\tilde{s}_i\right) admit_i\right] + \\
belowsub_{ij}\left[\underline{\phi} admit_i + \underline{h}_1\left(\tilde{s}_i\right) + \underline{h}_2\left(\tilde{s}_i\right) admit_i\right] + \gamma dist_{ij} + \varepsilon_{ij}.
\end{aligned}
\tag{4}
$$

This specification allows marginal expected utilities to vary depending on whether the cutoff school belongs to j's subsystem, the older sibling's centered exam score, and whether the sibling exceeded the cutoff score (ϕ and ϕ, the coefficient of interest). The corresponding underlined coefficients are all analogous except that they apply to the subsystem of the school attended by the older sibling if he scores below the cutoff.

If we assume that ε_{ij} is distributed i.i.d. extreme value type I, then the parameters of this model can be estimated with a conditional logit, where the outcome variable is selecting the school as the first choice. It is more appropriate to estimate a nested logit where subsystems are the nests, so that idiosyncratic preferences may be correlated within a subsystem and thus the restrictive independence of irrelevant alternatives assumption need not apply across nests. The bandwidth for this nested logit is set to 7 points on either side of the cutoff, selected because it corresponds to the optimal bandwidth for the rectangular kernel in the reduced form first choice specification found in column 1 of Table 2. This is somewhat arbitrary, but the results are consistent for other choices of the bandwidth.

5.2 Results

The nested logit results have signs consistent with the reduced form estimates. Table 9 provides selected estimated parameters from the nested logit specification in equation 4. The mean COMIPEMS exam score of students admitted in the previous year, as well as the proportion of the older sibling's middle school cohort choosing the school, are included as covariates.

Assignment to the cutoff school increases expected utility from that school and reduces the expected utility from the school below the cutoff. We can interpret these effects as the average marginal effect of sibling admission on willingness to travel (WTT) to that school by taking the ratio of the admission coefficient to the distance coefficient. This calculation gives an increase in WTT of 2.9 km (0.346/0.118, $SE = 0.15$) for the school above the cutoff and 5.1 km (0.601/0.118, $SE = 0.24$) for the school below the cutoff.[22] This asymmetry may arise because the younger sibling has a less precise prior on match quality for the school below the cutoff, so that the peer signal will be weighted more heavily and thus the average change in expected utility will be higher. This is plausible; the older sibling, whose information set is correlated with that of his younger sibling, has already ranked this school as less preferred than the school above the cutoff. One of the possible reasons for this is greater uncertainty about match quality, in addition to differences in expected match quality.

When the older sibling is on the margin between one *subsystem* and another, admission to the subsystem above the cutoff increases WTT to all schools in that subsystem by 2.6 km (0.306/0.118, $SE = 0.20$). Admission to the system below the cutoff increases WTT to all schools in the below subsystem by 3.1 km (0.364/0.118, $SE = 0.23$).[23]

Column 2 restricts the sample to students 3 to 5 years apart, so that students do not attend high school at the same time. The estimated effects of admission decline slightly and remain strongly significant.

Evidence for heterogeneous effects of admission with respect to graduation is provided in column 3, which estimates equation 4 fully interacted with the graduation dummy variable. There is suggestive evidence for heterogeneous same-school and subsystem effects. The coefficients of interest are those giving the differential effect of admission by graduation status, labeled "Graduated × admission." While the differential effects of admission on WTT to the school and subsystem above the cutoff are small and insignificant, the differential effect of admission for the school below the cutoff is 3.6 km (0.417/0.116, $SE = 1.29$) higher when the sibling graduates and 2.5 km (0.287/0.116, $SE = 1.14$) higher for schools in the subsystem below the cutoff. To address the data issues with the graduation proxy sample (three years of data, UNAM high schools missing), column 4 estimates the effects of admission on first non-elite

[22]Standard errors for WTT effects are computed using the delta method.

[23]The total change in WTT for the cutoff school when the student is at the boundary of a subsystem is obtained by summing the effect of admission to the school with the effect for admission to the subsystem.

choice for students near non-elite cutoffs. In this sample, the heterogeneous effects of admission with respect to graduation are of the expected sign and statistically significant, for schools and subsystems both above and below the cutoff.

Additional assumptions allow for interpretation of the effect sizes as measures of marginal willingness to pay (WTP). Taking the average WTT effect between the schools above and below the cutoff (2.9 km and 5.1 km, respectively), we have a 4 km average increase in WTT due to sibling admission. But students must travel both to and from school, so this measure should be doubled to 8 km/day. Students in Mexico have 195 instructional days per year, so the annualized effect on WTT is $8 * 195 = 1560$ km/year. Translating this measure to travel time is difficult because students travel using a combination of subway, private bus, driving, and walking. Assuming that the average speed of travel over these modes during rush hour in Mexico City is 10 km/hour, then students are willing to spend $1560/10 = 156$ additional hours per year traveling as a result of sibling admission. According to the National Survey of Occupation and Employment (ENOE), the average urban teen wage is \$2/hour. Taking this as the average valuation of time for students in the estimation sample, the change in WTP due to sibling admission is \$312/year. High school is three years long in Mexico City, so the total effect on WTP is \$936. This is likely to be a conservative estimate because traveling farther may require paying an additional bus fare of about \$.50/day.

6 Validity checks

This section presents two standard checks for the validity of the RD design. The first check is for whether the density of the running variable (centered COMIPEMS score) suddenly increases or decreases when it crosses the cutoff, as suggested by McCrary (2008). This might occur if the younger siblings of rejected students were less likely to apply to high school, for example if rejected students were more likely to drop out of school and younger siblings followed that example. Another, less likely possibility is that admission induces behavior that makes it impossible to match siblings to each other, such as changing their phone number or middle school. Figure 7, Panel A shows the density of centered COMIPEMS score for the RD sample of older siblings (corresponding to column 1 of Table 2). There is no clear change in density across the threshold, and indeed the density is nearly uniform over this domain. Panel B gives a closer view of the density near the cutoff. Implementing

McCrary's formal test for a discontinuity in the density yields an estimated difference in log height of only 0.0003 ($SE = 0.007$) at the cutoff.

The second check is to repeat the reduced form RD regressions, this time using exogenous student characteristics as the dependent variables. Imbens and Lemieux (2008) propose this as a way of verifying that exogenous characteristics do not suddenly change at the cutoff (which would call into question whether the endogenous variable would be balanced in the absence of a treatment effect). In order to jointly test that the admission coefficient is zero for all tested exogenous characteristics of the older sibling, seemingly unrelated regression (Zellner (1962)) is used. Table 10 shows the results, failing to reject that the admission coefficients are equal to zero ($p = 0.52$). The point estimates are quite precise as well, ruling out even fairly small covariate imbalances. Thus both checks yield support for the validity of the RD design.

7 Conclusion

Older siblings' school assignments strongly affect students' stated preferences and admissions outcomes, a finding that speaks to the role that peer networks play in overcoming incomplete information. What policy lessons can be taken from this result? One lesson is that aggregate school-level information is not a perfect substitute for the more subjective, individually-tailored information that students currently obtain from their networks. Match quality for the average student may already be known in the population, but idiosyncratic match quality is uncertain. Providing individualized information on match quality is not a trivial task for individual schools (as in the case of colleges) or public school systems. One approach already being undertaken at the tertiary level is to deploy the school's alumni network to connect with prospective students, providing them with personalized information through informal meetings and repeated electronic communication. But, as pointed out in Hoxby and Turner (2013), such labor-intensive interventions are expensive. Recruitment offices also have a role to play if they can provide the kinds of individual-specific information desired by students, but again, this may be expensive. Public school systems face the challenge of providing individualized information about all member schools. Furnishing printed material containing data beyond school-level aggregates is one way to begin, as in Hoxby and Turner (2013) in the context of college choice where information on net costs is the predominant barrier.

The findings give mixed support for school assignment systems that rely heavily on student choice. On one hand, it appears that the correlation observed by Hoxby and Avery (2012) is indeed causal, at least in this context: students with a low concentration of peers attending a particular school or set of schools are less likely to apply there, when under full information they might have applied. But this is also an endorsement of allowing choice, because it acknowledges a key rationale for its existence: students have access to a wealth of relevant individual-specific information, some from their peer networks, that administrators likely do not know. School choice allows students to put this information to work in the matching process. Creative policies that augment the information set of students in disadvantaged peer networks may help to retain the positive features of choice systems while lowering the informational barriers that reduce their effectiveness.

References

Aaronson, Daniel, Lisa Barrow, and William Sander. 2007. "Teachers and Student Achievement in the Chicago Public High Schools." Journal of Labor Economics 25(1):95-135.

Abdulkadiroglu, Atila, Joshua D. Angrist, and Parag A. Pathak. 2011. "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools." NBER Working Paper No. 17264.

Abdulkadiroglu, Atila, and Tayfun Sönmez. 2010. "Matching Markets: Theory and Practice." Prepared for the Econometric Society World Congress, China.

Ajayi, Kehinde F. 2012. "School Choice and Educational Mobility." Unpublished working paper, Boston University.

Ball, Stephen J., and Carol Vincent. 1998. "'I Heard It on the Grapevine': 'Hot' Knowledge and School Choice." British Journal of Sociology of Education 19(3): 377-400.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2008. "Bootstrap-Based Improvements for Inference with Clustered Errors." Review of Economics and Statistics 90(3): 414-427.

Cameron, A. Colin, and Douglas L. Miller. 2015. "A Practitioner's Guide to Cluster-Robust Inference." Journal of Human Resources 50(2).

Ceja, Miguel. 2006. "Understanding the Role of Parents and Siblings as Information Sources in the College Choice Process of Chicana Students." Journal of College Student Development 47(1): 87-104.

Chade, Hector, and Lones Smith. 2006. "Simultaneous Search." Econometrica 74(5): 1293-1307.

Clark, Damon. 2010. "Selective Schools and Academic Achievement." The B.E. Journal of Economic Analysis & Policy (Advances) 10(1), Article 9.

Crawford, Gregory S., and Matthew Shum. 2005. "Uncertainty and Learning in Pharmaceutical Demand." Econometrica 73(4): 1137-1173.

de Hoop, Jacobus. 2011. "Selective Schools and Education Decisions: Evidence from Malawi." Unpublished working paper, Free University of Amsterdam.

Dustan, Andrew, Alain de Janvry, and Elisabeth Sadoulet. 2015. "Flourish or Fail? The Risky Reward of Elite High School Admission in Mexico City." Unpublished working paper.

Dobbie, Will, and Roland G. Fryer, Jr. 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.

Dubins, L. E., and D. A. Freedman. 1981. "Machiavelli and the Gale-Shapley Algorithm." The American Mathematical Monthly 88(7): 485-495.

Erdem, Tülin, and Michael P. Keane. 1996. "Decision-Making under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets." Marketing Science 15(1): 1-20.

Friesen, Jane, Mohsen Javdani, Justin Smith, and Simon Woodcock. 2012. "How Do School 'Report Cards' Affect School Choice Decisions?" Canadian Journal of Economics 45(2): 784-807.

Gale, David, and Lloyd S. Shapley. 1962. "College Admissions and the Stability of Marriage." The American Mathematical Monthly 69(1): 9-15.

Goodman, Joshua, Michael Hurwitz, Jonathan Smith, and Julia Fox. 2015. "The Relationship Between Siblings' College Choices: Evidence from One Million SAT-Taking Families." Unpublished working paper.

Hastings, Justine S., Thomas J. Kane, and Douglas O. Staiger. 2009. "Heterogeneous Preferences and the Efficacy of Public School Choice." NBER Working Paper No. 12145.

Hastings, Justine, and Jeffrey Weinstein. 2008. "Information, School Choice, and Academic Achievement: Evidence from Two Experiments." Quarterly Journal of Economics 123(4): 1373-1414.

Hoxby, Caroline M., and Avery, Christopher. 2012. "The Missing 'One-Offs': The Hidden Supply of High-Achieving, Low Income Students." NBER Working Paper No. 18586.

Hoxby, Caroline M., and Sarah Turner. 2013. "Expanding College Opportunities for High-Achieving, Low Income Students." SIEPR Discussion Paper No. 12-014.

Imbens, Guido, and Karthik Kalyanaraman. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," *Review of Economic Studies* 79 (2012), 933-959.

Imbens, Guido W., and Thomas Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." Journal of Econometrics 142(2): 615-635.

Jackson, Kirabo. 2010. “Do Students Benefit From Attending Better Schools?: Evidence From Rule-based Student Assignments in Trinidad and Tobago.” The Economic Journal 142(549): 1399-1429.

Joensen Juanna Schrøter, and Helena Skyt Nielsen. 2015. “Peer Effects in Math and Science.” Unpublished working paper.

Johnson, Justin P., and David P. Myatt. 2006. “On the Simple Economics of Advertising, Markting, and Product Design.” American Economic Review 96(3): 756-784.

Koning, Pierre, and Karen van de Wiel. 2010. “Ranking the Schools: How Quality Information Affects School Choice in the Netherlands.” IZA Discussion Paper No. 4984.

Lai, Fang, Elisabeth Sadoulet, and Alain de Janvry. 2009. “The Adverse Effects of Parents’ School Selection Errors on Academic Achievement: Evidence from the Beijing Open Enrollment Program.” Economics of Education Review 28(4): 485-496.

Lai, Fang, Elisabeth Sadoulet, and Alain de Janvry. 2011 “The Contributions of School Quality and Teacher Qualifications to Student Performance: Evidence from a Natural Experiment in Beijing Middle Schools.” Journal of Human Resources 46(1): 123-53.

Lee, David S., and David Card. 2008. “Regression Discontinuity Inference with Specification Error.” Journal of Econometrics 142(2): 655-674.

Lee, David S., and Thomas Lemieux. 2010. “Regression Discontinuity Designs in Economics.” Journal of Economic Literature 48(2), 281-355.

Lucas, Adrienne M., and Isaac M. Mbiti. 2012. “The Determinants and Consequences of School Choice Errors in Kenya.” American Economic Review: Papers and Proceedings 102(3): 283-288.

Manski, Charles F. 1993. “Identification of Endogenous Social Effects: The Refection Problem.” Review of Economic Studies 60(3): 531-42.

Manski, Charles F. 1995. *Identification Problems in the Social Sciences*. Harvard University Press.

McCrary, Justin. 2008. “Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test.” Journal of Econometrics 142(2): 698-714.

Mizala, Alejandra, and Miguel Urquiola. 2013. “School Markets: The Impact of Information Approximating Schools’ Effectiveness.” Journal of Development Economics 103: 313-335.

Pallais, Amanda. 2009. “Small Differences that Matter: Mistakes in Applying to College.” Revise and resubmit, Journal of Labor Economics.

Pop-Eleches, Cristian, and Miguel Urquiola. 2013. “Going to a Better School: Effects and Behavioral Responses.” American Economic Review, forthcoming.

Roberts, John H., and Glen L. Urban. 1988. “Modeling Multiattribute Utility, Risk, and Belief Dynamics for New Consumer Durable Brand Choice." Management Science 34(2): 167-185.

Roth, Alvin E. 1982. "Incentive Compatibility in a Market With Indivisible Goods." Economics Letters 9(2): 127-132.

Train, Kenneth E. 2009. *Discrete Choice Methods with Simulation.* Cambridge University Press.

Zellner, Arnold. 1962. "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias." Journal of the American Statistical Association 57(298): 348-368.

Zhang, Hongliang. 2012. "The Mirage of Elite Schools: Evidence from Lottery-based School Admissions in China." Unpublished working paper, Chinese University of Hong Kong.

Tables

Table 1: Summary statistics for full and sibling samples

	(1) All students	(2) Matched younger sibling sample	p-value for equality of means
Male	0.50	0.48	0.00
Maximum of mother's and father's education (years)	10.32	10.46	0.00
	(3.57)	(3.39)	
Number of siblings	2.23	2.33	0.00
	(1.49)	(1.28)	
Birth order (1 is first-born)	2.19	2.68	0.00
	(1.38)	(1.05)	
Hours studied per week	4.81	4.90	0.00
	(3.16)	(3.17)	
Middle school grade point average (of 10)	8.01	8.14	0.00
	(0.83)	(0.82)	
Number of schools ranked	9.24	8.92	0.00
	(3.93)	(3.57)	
Elite school as first choice	0.61	0.64	0.00
Assigned to an elite school	0.20	0.21	0.00
COMIPEMS examination score	62.90	62.19	0.00
	(19.20)	(18.83)	
Distance from student's home to first choice school (km)	7.73	7.82	0.00
	(6.32)	(6.15)	
Distance from student's home to first non-elite choice school (km)	6.23	5.94	0.00
	(5.47)	(5.04)	
Grade year difference between siblings		2.57	
		(1.28)	
Siblings chose same first choice school		0.34	
Graduated high school (for non-UNAM students in 2005-2007 cohorts)	0.39	0.45	0.00
Observations	3,423,052	455,375	

Note. Standard deviations in parentheses. Statistics in column 2 are for all younger siblings in the matched sample, described in the text.

Table 2: Effect of older sibling admission on younger sibling choice

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Specification	Local linear				Global			
Dependent variable	Cutoff school is first choice	Cutoff school is any choice	School below cutoff is first choice	School below cutoff is any choice	Cutoff school is first choice	Cutoff school is any choice	School below cutoff is first choice	School below cutoff is any choice
Score ≥ cutoff	0.076***	0.098***	-0.047***	-0.105***	0.078***	0.099***	-0.048***	-0.107***
	(0.0026)	(0.0030)	(0.0017)	(0.0034)	(0.0029)	(0.0035)	(0.0017)	(0.0037)
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Observations	340,255	393,423	318,386	356,435	1,084,137	1,139,075	1,139,075	1,139,075
Adjusted R-squared	0.164	0.247	0.033	0.186	0.140	0.223	0.038	0.166
Mean of dependent variable	0.167	0.588	0.039	0.451	0.153	0.553	0.036	0.420
Bandwidth	9.2	10.0	8.1	8.7	--	--	--	--

Note. Dependent variables are dummy variables pertaining to the choice of the younger sibling. Regressions include cutoff school-year fixed effects and polynomials in older sibling's centered exam score. Local linear model weights observations using the edge kernel; global model uses the rectangular kernel. Standard errors accounting for clustering at the older sibling and centered score levels are in parentheses. Stars for statistical significance are based on t-tests using (G-1) degrees of freedom, where G is the number of points of support of the centered score. Bootstrapped p-values accounting for clustering at the centered score level are in brackets.

* p<0.10, ** p<0.05, *** p<0.01

Table 3: Effect of older sibling admission on younger sibling choice, disaggregated by type of cutoff school

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Specification	Local linear				Global			
Sample (cutoff school type)	Elite	Elite	Non-elite	Non-elite	Elite	Elite	Non-elite	Non-elite
Dependent variable	Cutoff school is first choice	Cutoff school is any choice	Cutoff school is first non-elite choice	Cutoff school is any choice	Cutoff school is first choice	Cutoff school is any choice	Cutoff school is first non-elite choice	Cutoff school is any choice
Score ≥ cutoff	0.113***	0.117***	0.094***	0.079***	0.120***	0.123***	0.096***	0.079***
	(0.0045)	(0.0042)	(0.0041)	(0.0043)	(0.0047)	(0.0048)	(0.0048)	(0.0055)
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Observations	157,709	177,343	163,424	197,596	566,312	568,242	517,614	570,833
Adjusted R-squared	0.133	0.143	0.130	0.248	0.101	0.122	0.126	0.240
Mean of dependent variable	0.250	0.705	0.194	0.473	0.228	0.662	0.176	0.443
Bandwidth	8.1	9.3	9.1	10.1	--	--	--	--

Note. Dependent variables are dummy variables pertaining to the choice of the younger sibling. Regressions include cutoff school-year fixed effects and polynomials in older sibling's centered exam score. Local linear model weights observations using the edge kernel; global model uses the rectangular kernel. Standard errors accounting for clustering at the older sibling and centered score levels are in parentheses. Stars for statistical significance are based on t-tests using (G-1) degrees of freedom, where G is the number of points of support of the centered score. Bootstrapped p-values accounting for clustering at the centered score level are in brackets.

* p<0.10, ** p<0.05, *** p<0.01

Table 4: Effect of older sibling admission on younger sibling's preference for same school, heterogeneity by age difference of siblings

	(1)	(2)	(3)	(4)
Specification	Local linear		Global	
Dependent variable	Cutoff school is first choice	Cutoff school is any choice	Cutoff school is first choice	Cutoff school is any choice
Score ≥ cutoff	0.080***	0.101***	0.082***	0.104***
	(0.0036)	(0.0040)	(0.0039)	(0.0047)
	[0.00]	[0.00]	[0.00]	[0.00]
(Score ≥ cutoff) × (Siblings 3+ years apart)	-0.006	-0.005	-0.008	-0.008
	(0.0052)	(0.0059)	(0.0057)	(0.0068)
	[0.47]	[0.52]	[0.25]	[0.20]
Observations	340,255	393,423	1,084,137	1,139,075
Adjusted R-squared	0.163	0.253	0.140	0.228
Mean of dependent variable	0.167	0.588	0.153	0.553
Bandwidth	9.2	10.0	--	--

Note. Regressions include cutoff school-year fixed effects and polynomials in older sibling's centered exam score, all interacted with the (Siblings 3+ years apart) dummy. Local linear model weights observations using the edge kernel; global model uses the rectangular kernel. Standard errors accounting for clustering at the older sibling and centered score levels are in parentheses. Stars for statistical significance are based on t-tests using (G-1) degrees of freedom, where G is the number of points of support of the centered score. Bootstrapped p-values accounting for clustering at the centered score level are in brackets.

* p<0.10, ** p<0.05, *** p<0.01

Table 5: Effect of older sibling admission on number of other schools chosen in cutoff subsystem

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Specification	Local linear				Global			
Dependent variable	Schools chosen in subsystem of cutoff school	Schools chosen in subsystem of cutoff school	Schools chosen in subsystem of school below cutoff	Schools chosen in subsystem of school below cutoff	Schools chosen in subsystem of cutoff school	Schools chosen in subsystem of cutoff school	Schools chosen in subsystem of school below cutoff	Schools chosen in subsystem of school below cutoff
Score ≥ cutoff	0.233***	0.246***	-0.201***	-0.207***	0.235***	0.251***	-0.192***	-0.200***
	(0.0182)	(0.0240)	(0.0158)	(0.0217)	(0.0230)	(0.0301)	(0.0223)	(0.0302)
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
(Score ≥ cutoff) × (Siblings 3+ years apart)		-0.026		0.016		-0.029		0.013
		(0.0363)		(0.0312)		(0.0458)		(0.0440)
		[0.14]		[0.59]		[0.19]		[0.73]
Observations	224,648	224,648	274,039	274,039	541,561	541,561	541,561	541,561
Adjusted R-squared	0.170	0.179	0.044	0.049	0.153	0.158	0.038	0.040
Mean of dependent variable	2.069	2.069	1.603	1.603	2.024	2.024	1.600	1.600
Bandwidth	12.0	12.0	14.6	14.6	--	--	--	--

Note. Dependent variables are the number of school in the respective subsystems chosen by the younger sibling, excluding either the cutoff school or the school immediately below the cutoff. Samples are limited to older siblings for whom rejection from the cutoff school results in admission to a school in a different subsystem. Regressions include cutoff school-year fixed effects and polynomials in older sibling's centered exam score, all interacted with the (Siblings 3+ years apart) dummy. Local linear model weights observations using the edge kernel; global model uses the rectangular kernel. Standard errors accounting for clustering at the older sibling and centered score levels are in parentheses. Stars for statistical significance are based on t-tests using (G-1) degrees of freedom, where G is the number of points of support of the centered score. Bootstrapped p-values accounting for clustering at the centered score level are in brackets.

* p<0.10, ** p<0.05, *** p<0.01

Table 6: Effect of older sibling admission on younger sibling assignment outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Specification	Local linear				Global			
Dependent variable	Assigned to cutoff school	Assigned to school below cutoff	Assigned to school in subsystem of cutoff school	Assigned to school in subsystem below cutoff	Assigned to cutoff school	Assigned to school below cutoff	Assigned to school in subsystem of cutoff school	Assigned to school in subsystem below cutoff
Score ≥ cutoff	0.042***	-0.041***	0.046***	-0.047***	0.044***	-0.043***	0.050***	-0.047***
	(0.0015)	(0.0018)	(0.0038)	(0.0031)	(0.0021)	(0.0021)	(0.0046)	(0.0048)
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Observations	630,611	429,768	207,198	318,614	1,139,075	1,139,075	541,561	541,561
Adjusted R-squared	0.032	0.021	0.035	0.017	0.042	0.017	0.057	0.015
Mean of dependent variable	0.082	0.066	0.211	0.197	0.072	0.052	0.206	0.185
Bandwidth	16.6	10.6	11.1	18.1	--	--	--	--

Note. Dependent variables are dummy variables pertaining to the admission outcome of the younger sibling. Samples for subsystem admission outcomes are limited to older siblings for whom rejection from the cutoff school results in admission to a school in a different subsystem. Regressions include cutoff school-year fixed effects and polynomials in older sibling's centered exam score. Local linear model weights observations using the edge kernel; global model uses the rectangular kernel. Standard errors accounting for clustering at the older sibling and centered score levels are in parentheses. Stars for statistical significance are based on t-tests using (G-1) degrees of freedom, where G is the number of points of support of the centered score. Bootstrapped p-values accounting for clustering at the centered score level are in brackets.

* p<0.10, ** p<0.05, *** p<0.01

Table 7: Effect of older sibling admission to an elite school on younger sibling elite school choice

Panel A. Effect on selecting an elite school as first choice

	(1)	(2)	(3)
Baseline predicted probability of selecting elite school as first choice	[0, .6]	(.6, .8]	(.8, 1]
Score ≥ cutoff	0.142***	0.076***	0.032***
	(0.0180)	(0.0094)	(0.0061)
	[0.00]	[0.00]	[0.00]
Observations	12,410	33,507	40,603
Adjusted R-squared	0.038	0.022	0.013
Mean of dependent variable	0.635	0.789	0.910
Bandwidth	19.3	16.2	13.9

Panel B. Effect on total number of elite schools selected

	(1)	(2)	(3)
Baseline predicted probability of selecting elite school as first choice	[0, .6]	(.6, .8]	(.8, 1]
Score ≥ cutoff	0.649***	0.378***	0.211***
	(0.0976)	(0.0767)	(0.0585)
	[0.00]	[0.00]	[0.07]
Observations	12,869	27,705	50,432
Adjusted R-squared	0.046	0.047	0.057
Mean of dependent variable	2.705	3.999	5.516
Bandwidth	19.7	12.7	18.0

Panel C. Effect on assignment to an elite school

	(1)	(2)	(3)
Baseline predicted probability of selecting elite school as first choice	[0, .6]	(.6, .8]	(.8, 1]
Score ≥ cutoff	0.037**	0.035***	0.004
	(0.0143)	(0.0093)	(0.0089)
	[0.06]	[0.00]	[0.57]
Observations	13,849	37,030	48,066
Adjusted R-squared	0.032	0.023	0.021
Mean of dependent variable	0.196	0.230	0.313
Bandwidth	21.7	18.4	17.4

Note. Sample is composed of older siblings near the cutoff of an elite school who will be assigned to a non-elite school if they score below the cutoff. Columns partition this sample by the younger sibling's predicted probability of selecting an elite school as first choice, predicted on the sample of students with older siblings below the elite cutoff, using the probit model in Table A2. Regressions include cutoff school-year fixed effects and first-order polynomials in older sibling's centered exam score. Observations are weighted using the edge kernel. Standard errors accounting for clustering at the older sibling and centered score levels are in parentheses. Stars for statistical significance are based on t-tests using (G-1) degrees of freedom, where G is the number of points of support of the centered score. Bootstrapped p-values accounting for clustering at the centered score level are in brackets.
* p<0.10, ** p<0.05, *** p<0.01

Table 8: Differential effect of older sibling admission on school choice by graduation outcome

Panel A. Effect on first choice

	(1)	(2)	(3)	(4)
Specification	Local linear		Global	
Score ≥ cutoff	0.014***	(One per	0.020***	(One per
	(0.0043)	school)	(0.0077)	school)
	[0.00]		[0.01]	
(Score ≥ cutoff) × (Older sibling graduated)	0.023***	0.019***	0.023*	0.025**
	(0.0064)	(0.0068)	(0.0117)	(0.0109)
	[0.01]	[0.00]	[0.07]	[0.03]
Observations	121,218	121,218	158,266	158,266
Adjusted R-squared	0.081	0.083	0.076	0.079
Mean of dependent variable	0.086	0.086	0.084	0.084
Bandwidth	23.7	23.7	--	--

Panel B. Effect on first non-elite choice

	(1)	(2)	(3)	(4)
Specification	Local linear		Global	
Score ≥ cutoff	0.043***	(One per	0.051***	(One per
	(0.0100)	school)	(0.0134)	school)
	[0.00]		[0.03]	
(Score ≥ cutoff) × (Older sibling graduated)	0.071***	0.073***	0.067***	0.081***
	(0.0142)	(0.0153)	(0.0193)	(0.0175)
	[0.01]	[0.00]	[0.00]	[0.01]
Observations	56,118	56,118	121,974	121,974
Adjusted R-squared	0.125	0.126	0.120	0.122
Mean of dependent variable	0.192	0.192	0.173	0.173
Bandwidth	11.7	11.7	--	--

Panel C. Effect on number of other schools chosen in same subsystem

	(1)	(2)	(3)	(4)
Specification	Local linear		Global	
Score ≥ cutoff	0.202***	(One per	0.246***	(One per
	(0.0480)	school)	(0.0816)	school)
	[0.00]		[0.13]	
(Score ≥ cutoff) × (Older sibling graduated)	0.116*	0.153**	0.074	0.189*
	(0.0660)	(0.0693)	(0.1146)	(0.1041)
	[0.04]	[0.01]	[0.46]	[0.05]
Observations	54,579	54,579	80,483	80,483
Adjusted R-squared	0.182	0.188	0.180	0.187
Mean of dependent variable	1.863	1.863	1.885	1.885
Bandwidth	20.5	20.5	--	--

Note. Samples exclude all students at an UNAM school cutoff or who would attend an UNAM school if rejected from the cutoff school, since the UNAM schools have no graduation data available. Sample in Panel B is limited to sibling pairs where the older sibling is at the cutoff of a non-elite school. Sample in Panel C is limited to sibling pairs where the older sibling would be admitted to a school in a different subsystem upon rejection. Regressions include graduation-cutoff school-year fixed effects and polynomials in older sibling's centered exam score interacted with the graduation dummy. Columns 2 and 4 also include one admission coefficient per cutoff school, de-meaned older sibling's GPA, a first-order piecewise polynomial in the interaction between older sibling's GPA and centered test score, and the interaction between GPA and graduation. Local linear model weights observations using the edge kernel; global model uses the rectangular kernel. Standard errors accounting for clustering at the older sibling and centered score levels are in parentheses. Stars for statistical significance are based on t-tests using (G-1) degrees of freedom, where G is the number of points of support of the centered score. Bootstrapped p-values accounting for clustering at the centered score level are in brackets.

* $p<0.10$, ** $p<0.05$, *** $p<0.01$

Table 9: Nested logit estimates of school choice model

		(1)	(2)	(3)	(4)
Sample		All siblings	Siblings 3-5 years apart	Graduation proxy available	Graduation proxy available
Dependent variable		First choice	First choice	First choice	First non-elite choice
School above cutoff	Constant	0.454***	0.382***	0.522***	0.648***
		(0.015)	(0.021)	(0.069)	(0.049)
	Score ≥ cutoff	0.346***	0.332***	0.186**	0.131**
		(0.018)	(0.026)	(0.084)	(0.059)
	Graduated			0.065	-0.019
				(0.087)	(0.063)
	Graduated × Score ≥ cutoff			0.003	0.181**
				(0.106)	(0.079)
School below cutoff	Constant	0.766***	0.646***	0.684***	0.752***
		(0.020)	(0.028)	(0.083)	(0.065)
	Score ≥ cutoff	-0.601***	-0.522***	-0.296***	-0.385***
		(0.029)	(0.041)	(0.113)	(0.088)
	Graduated			0.147	0.196**
				(0.101)	(0.081)
	Graduated × Score ≥ cutoff			-0.417***	-0.262**
				(0.149)	(0.116)
Subsystem of school above cutoff	Fixed effects	YES	YES	YES	YES
	Score ≥ cutoff	0.306***	0.302***	0.190**	0.059
		(0.024)	(0.036)	(0.092)	(0.079)
	Graduated × fixed effects			YES	YES
	Graduated × Score ≥ cutoff			0.080	0.253**
				(0.123)	(0.108)
Subsystem of school below cutoff	Fixed effects	YES	YES	YES	YES
	Score ≥ cutoff	-0.364***	-0.339***	-0.076	-0.071
		(0.027)	(0.040)	(0.099)	(0.082)
	Graduated × fixed effects			YES	YES
	Graduated × Score ≥ cutoff			-0.287**	-0.207*
				(0.132)	(0.113)
Distance to school	Constant	-0.118***	-0.116***	-0.116***	-0.130***
		(0.001)	(0.001)	(0.002)	(0.003)
	Graduated			0.000	0.000
				(0.001)	(0.002)
Mean COMIPEMS of school	Constant	0.034***	0.033***	0.030***	0.040***
		(0.000)	(0.000)	(0.001)	(0.001)
	Graduated			-0.002***	-0.002***
				(0.001)	(0.001)
Proportion of older sib MS cohort choosing as first choice	Constant	2.400***	2.227***	2.436***	5.093***
		(0.025)	(0.035)	(0.070)	(0.206)
	Graduated			0.350***	-0.484*
				(0.080)	(0.249)
	Intra-nest correlation parameter (λ)	0.483***	0.467***	0.455***	0.450***
		(0.004)	(0.006)	(0.008)	(0.010)
	Students	263,431	121,202	46,861	36,349

Note. Results are from a nested logit model, with subsystem as the nest, for students within 7 points of the cutoff of a school. Specifications include dummy variables for school subsystem and interactions of these dummy variables with 1) an indicator for whether the school above the cutoff belongs to that subsystem and 2) an indicator for whether the school below the cutoff belongs to that subsystem. Also included are first-order piecewise polynomials in school above cutoff, school below cutoff, subsystem above cutoff, and subsystem below cutoff. In columns 3 and 5, every variable is interacted with the "graduated" dummy variable. Standard errors accounting for clustering at the older sibling level are in parentheses.

*** p<0.01, ** p<0.05, * p<0.1.

Table 10: Test for balance of older sibling covariates

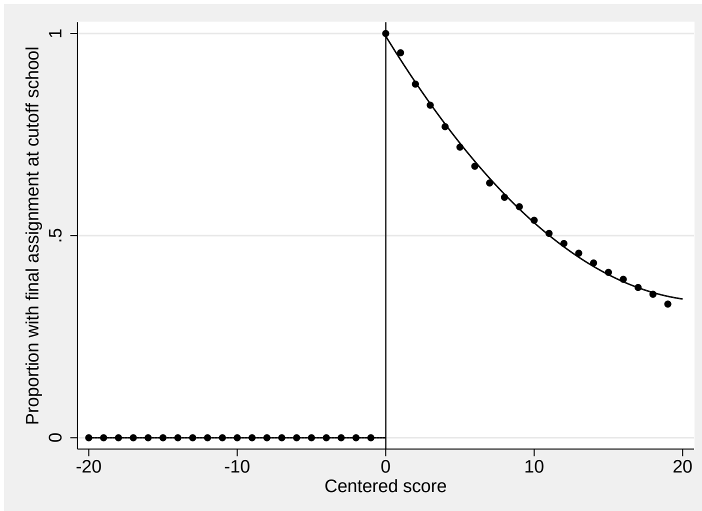
Dependent variable	(1) Parental Education	(2) Male	(3) Hours studied	(4) Middle school GPA	(5) Number of siblings	(6) Birth order (1=oldest)
Score ≥ cutoff	0.004	-0.002	0.021	0.004	0.011	0.009
	(0.0195)	(0.0024)	(0.0208)	(0.0040)	(0.0071)	(0.0059)
	[0.89]	[0.49]	[0.56]	[0.26]	[0.06]	[0.05]
Observations	461,454	717,351	406,606	599,433	468,740	467,701
Adjusted R-squared	0.104	0.074	0.074	0.201	0.057	0.033
Mean of dependent variable	10.79	0.44	5.17	8.16	2.20	1.58
S.D. of dependent variable	3.33	0.50	3.25	0.80	1.18	0.98
Bandwidth	16.3	19.6	13.7	16.0	16.1	16.4
p-value for joint significance of Score ≥ cutoff coefficients						0.52

Note. Dependent variable corresponds to the older sibling. Regressions include cutoff school-year fixed effects and first-order polynomials in older sibling's centered exam score. Observations are weighted using the edge kernel. Standard errors accounting for clustering at the older sibling and centered score levels are in parentheses. Stars for statistical significance are based on t-tests using (G-1) degrees of freedom, where G is the number of points of support of the centered score. Bootstrapped p-values accounting for clustering at the centered score level are in brackets.

* p<0.10, ** p<0.05, *** p<0.01

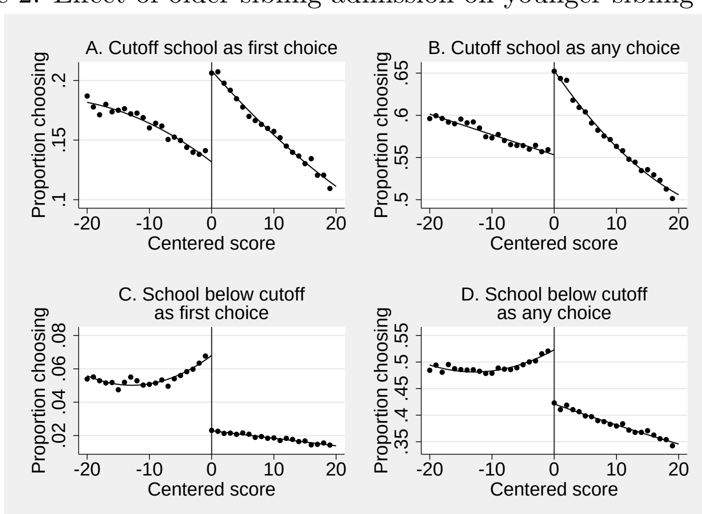
Figures

Figure 1: Verification of sharp discontinuity in admission probability due to assignment rule



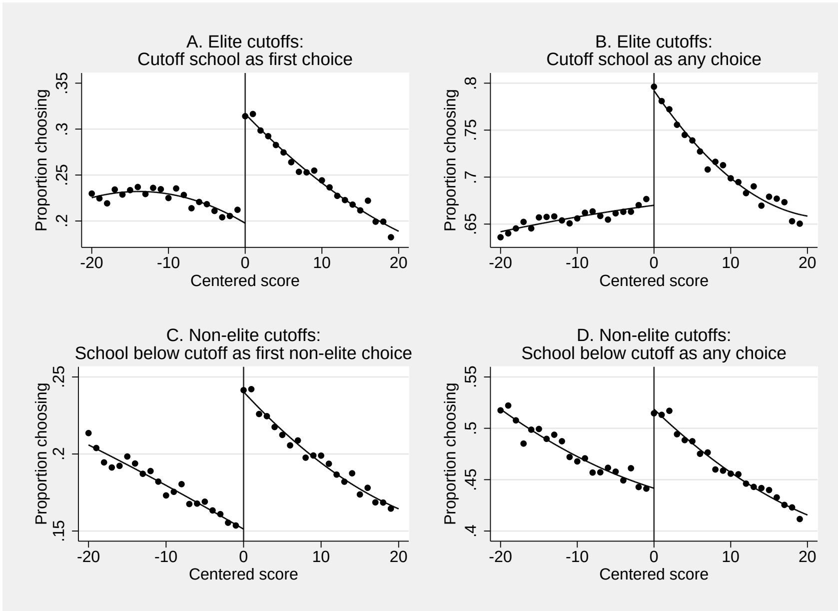
Note. Variable on vertical axis is proportion of older siblings assigned to the cutoff school. Variable on horizontal axis is older sibling's COMIPEMS exam score, centered to be 0 at the corresponding cutoff score. Fitted lines are from a quadratic fit.

Figure 2: Effect of older sibling admission on younger sibling choice



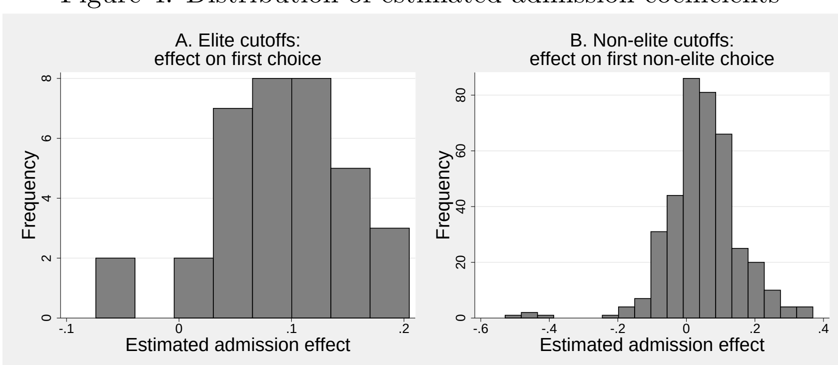
Note. Proportion on vertical axis pertains to the younger siblings. Variable on horizontal axis is older sibling's COMIPEMS exam score, centered to be 0 at the corresponding cutoff score. Fitted lines are from a quadratic fit.

Figure 3: Effect of older sibling admission on younger sibling choice, disaggregated by type of cutoff school



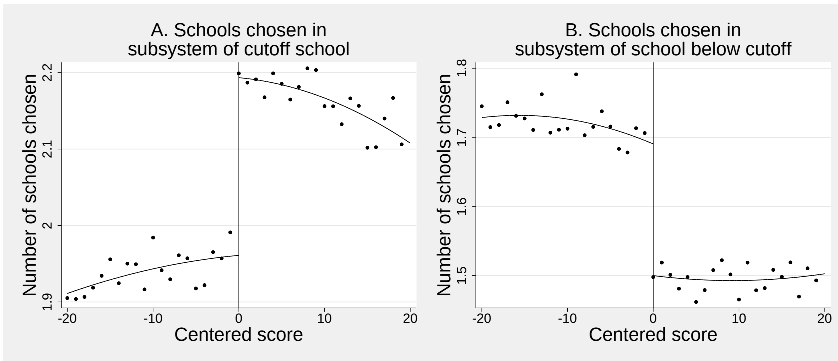
Note. Proportion on vertical axis pertains to the younger sibling. Variable on horizontal axis is older sibling's COMIPEMS exam score, centered to be 0 at the corresponding cutoff score. Fitted lines are from a quadratic fit.

Figure 4: Distribution of estimated admission coefficients



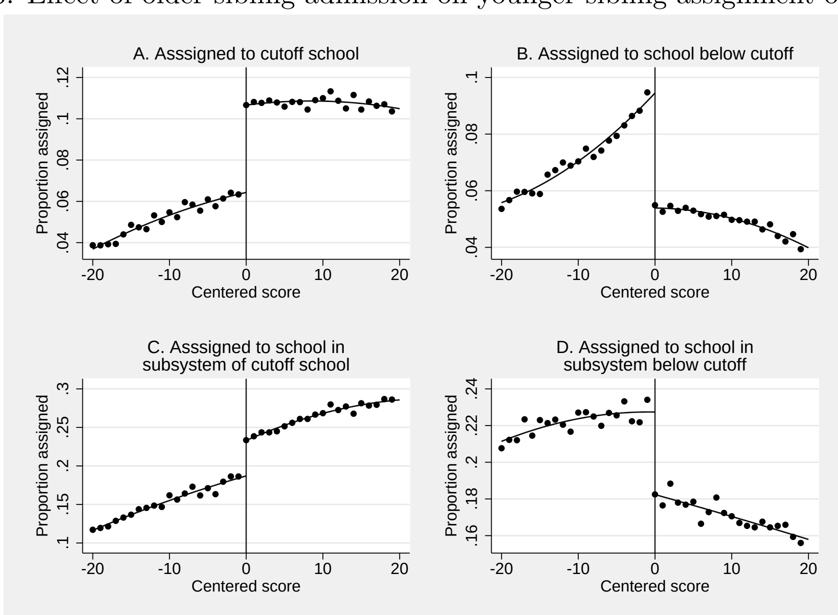
Note. Histograms are of estimated coefficients on older sibling admission, estimated from separate local linear regressions for each cutoff school at the IK optimal bandwidth, where the dependent variable is a dummy variable equal to 1 if the younger sibling chose the cutoff school as his first choice. Panel A plots the coefficients for elite cutoff schools. Panel B plots the coefficients for non-elite cutoff schools.

Figure 5: Effect of older sibling admission on number of other schools chosen in cutoff subsystem



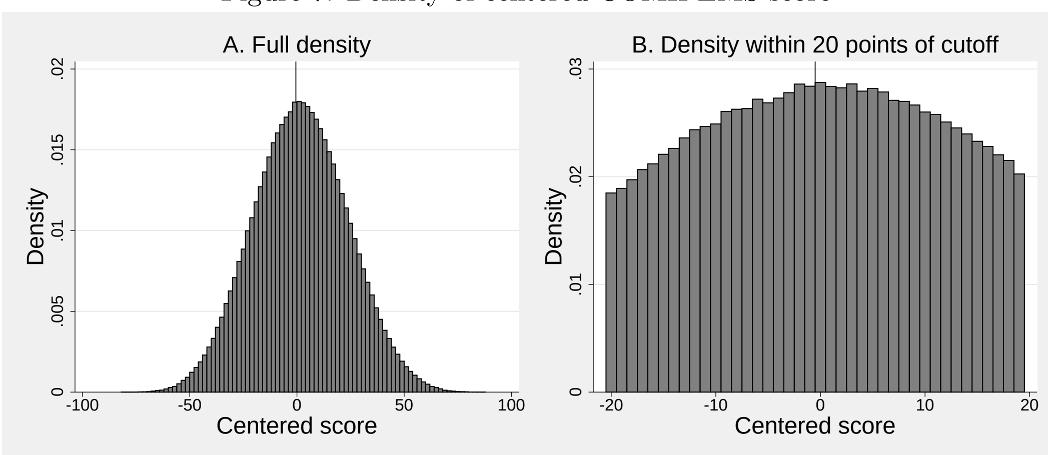
Note. Variable on vertical axis is number of schools selected by the younger sibling in the subsystem to which the older sibling's cutoff school belongs, excluding the cutoff school. Sample is limited to older siblings for whom the school immediately below the cutoff belongs to a different subsystem than the cutoff school, so that marginal rejection results in assignment to another subsystem. Variable on horizontal axis is older sibling's COMIPEMS exam score, centered to be 0 at the corresponding cutoff score. Fitted lines are from a quadratic fit.

Figure 6: Effect of older sibling admission on younger sibling assignment outcomes



Note. Proportion on vertical axis pertains to the admissions outcome of the younger sibling. Variable on horizontal axis is older sibling's COMIPEMS exam score, centered to be 0 at the corresponding cutoff score. Fitted lines are from a quadratic fit.

Figure 7: Density of centered COMIPEMS score



Note. Histogram is of COMIPEMS score for students near a cutoff. Scores are centered so that they are 0 at the cutoff score.

Appendix: For Online Publication

Model of school choice

This section extends a model of school choice from Hastings et al. (2009) by incorporating incomplete information, risk aversion, and learning from peers. In my model, the utility from attending each school is uncertain because of incomplete information about student-school match quality. Risk-averse students revise their beliefs about utilities by receiving informative signals about match quality from peers. The setup is similar to models of consumer demand for experience goods, in particular Roberts and Urban (1988) and Erdem and Keane (1996), where consumers are uncertain about product quality and revise their beliefs due to word-of-mouth or informative advertising. Students may also gain productive knowledge about schools from their peers, which allows them to obtain higher utility from attending the peer's school. This latter advantage can be thought of in a similar way to the effect of learning on technology adoption, as in Foster and Rosenzweig (1995).[24] In this case, students are unsure of how to use the school "technology" to build human capital but learn from peers about how to do so optimally.

This model produces testable hypotheses about how students react to new information about specific schools. First, the model predicts that the average impact of new information on same-school expected utility is positive. This is a prediction about the average effect of new information over all students and schools in the population, not a prediction that the average effect will be positive for each school. Second, the model predicts that the impact of new information depends on how positive or negative the signal was. Finally, these effects are predicted to apply, to a lesser degree, to other schools that are observably similar to the school about which the information was received.

General setup

The student's problem is to maximize expected utility by choosing one school to attend from his choice set. Here I abstract from the problem of portfolio construction and focus on the first choice. This is reasonable if one thinks that the first listed option is the student's most-preferred school, a modest assumption given the large number

[24] Foster, Andrew D., and Mark R. Rosenzweig. 1995. "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture." Journal of Political Economy 103(6): 1176-1209.

of options that a student is allowed to list in order to diversify and choose safety schools.

Student i's utility from school $j \in J$ is a function of student-school match quality:

$$U_{ij} = U(X_{ij}\boldsymbol{\beta}_i) = U\left(\left(\bar{X}_j + \widetilde{X}_{ij}\right)\boldsymbol{\beta}_i\right)$$

where match quality is expressed as the sum of student-school attributes in the vector X_{ij} weighted by the student-specific vector of preference parameters $\boldsymbol{\beta}_i$. The attribute vector is decomposed into two terms: $\bar{X}_j$ is the average level in the population and $\widetilde{X}_{ij}$ is the student-specific deviation from this level. An example of a student-school attribute is academic fit, which is on average higher at some schools than others, but also has a student-specific component that depends on how well the school caters to the student's particular learning style and ability level.

The student knows the relative weights $\boldsymbol{\beta}_i$ he puts on each attribute. If he also knows X_{ij}, and if he is risk-neutral with respect to match quality, so that $U(X_{ij}\boldsymbol{\beta}_i) = X_{ij}\boldsymbol{\beta}_i$, this model is nearly identical to the one in Hastings et al. (2009). In that case, the student chooses school j if it provides the highest match quality out of all schools in the choice set: $X_{ij}\boldsymbol{\beta}_i > X_{ik}\boldsymbol{\beta}_i \; \forall k \neq j \in J$.[25]

Incomplete information about match quality

Incomplete information about match quality is modeled by making it so that the student imperfectly observes student-school attributes. He does not observe $\bar{X}_j$ or $\widetilde{X}_{ij}$, but he knows the distributions from which each is drawn:

$$\bar{X}_j \sim \mathcal{N}\left(\bar{X}_j^0, \Sigma_{\bar{X}_j}\right), \quad \widetilde{X}_{ij} \sim \mathcal{N}\left(\widetilde{X}_{ij}^0, \Sigma_{\widetilde{X}_{ij}}\right).$$

For simplicity of exposition, the covariance matrices $\Sigma_{\bar{X}_j}$ and $\Sigma_{\widetilde{X}_{ij}}$ are assumed to be diagonal, and $\bar{X}_j$ and $\widetilde{X}_{ij}$ are assumed to be mean independent. Thus X_{ij} is distributed normally with mean $X_{ij}^0 = \bar{X}_j^0 + \widetilde{X}_{ij}^0$ and diagonal covariance matrix with $(\ell, \ell)^{\text{th}}$ entry $1/\tau_{\ell ij}^0$.[26]

Because X_{ij} is unknown, a risk-neutral student chooses j if it maximizes *expected* match quality: $\mathrm{E}_0\left[X_{ij}\boldsymbol{\beta}_i\right] > \mathrm{E}_0\left[X_{ik}\boldsymbol{\beta}_i\right] \; \forall k \neq j \in J$, where the 0 subscript indicates

[25]Hastings et al. (2009) do not explicitly model uncertainty, but they do say that uncertainty about an attribute would lead to a lower effective weight being placed on it.

[26]I assume that for any two schools j and k, X_{ij} and X_{ik} are mean independent.

that the expectation is formed solely on the basis of the match quality distributions. Incomplete information about match quality (in particular, about mean quality $\overline{X}_j$) is sufficient to predict the results from Hastings and Weinstein (2008), where giving information about school-level average test scores to students increased the weight that students placed on test scores when choosing schools.[27]

Risk aversion and returns to productive knowledge

I now introduce two channels through which information will positively affect expected utility: returns to productive knowledge and risk aversion with respect to match quality.

I parameterize the returns to productive knowledge in a simple way, adding a term $r_j(n_{ij})$ to the utility function, where n_{ij} is the level of i's knowledge about school j. The marginal return to knowledge is strictly positive so that $r_j' > 0$. Examples of productive knowledge are knowing which teachers are the best to take or being aware of an after-school tutoring program.

Allowing the student to be risk-averse will address a troubling result from the risk-neutral model. Risk neutrality implies that the relative precision with which match quality is known does not affect choice. That is, presented with a choice between two schools of equal expected match quality but where one's match is known with complete certainty and the other with uncertainty, the student will be indifferent between them. A risk-averse student will prefer the school where match quality is known with certainty.

To model risk aversion, I allow utility to be concave in match quality. Following Roberts and Urban (1988), I use exponential utility:

$$U_{ij} = -exp\left\{-\rho X_{ij}\boldsymbol{\beta}_i + r_j\left(n_{ij}\right)\right\}$$

where ρ, the coefficient of risk aversion, is assumed to be positive. Due to exponential utility and the joint normal distribution of X_{ij}, expected utility from school j in the absence of additional information can be written in terms of the mean and variance (or precision) of the prior distribution of match quality, as well as the return to

[27]Intuitively, students were choosing on the basis of both signal and noise about test scores, and the information intervention allowed students to choose on the basis of a stronger signal.

productive knowledge:[28]

$$U^*_{0ij} = \mathrm{E}_0\left[X_{ij}\boldsymbol{\beta}_i\right] - \frac{\rho}{2} Var\left(X_{ij}\boldsymbol{\beta}_i\right) + r_j\left(n^0_{ij}\right)$$
$$= X^0_{ij}\boldsymbol{\beta}_i - \frac{\rho}{2}\sum_{\ell}\frac{\beta^2_{\ell i}}{\tau^0_{\ell ij}} + r_j\left(n^0_{ij}\right). \tag{1}$$

where $\beta^2_{\ell i}/\tau^0_{\ell ij}$ is the variance of the distribution of match quality from attribute ℓ. The student optimizes with respect to both the mean and variance of match quality, so schools are now "penalized" when beliefs about them are noisier. He also values productive knowledge. He chooses the school j that provides the highest expected utility of all available schools: $U^*_{0ij} > U^*_{0ik}\ \forall k \neq j \in J$.

Effect of peer information

When student i's peer attends school j, he gives two pieces of information. First, he provides productive knowledge about school j, so that the new level of knowledge is higher: $n^1_{ij} > n^0_{ij}$. Second, the student improves on his prior belief about match quality by receiving informative signals about student-school attributes X_{ij}. This information comes in the form of an unbiased, noisy signal about each attribute:

$$P_{ij} = X_{ij} + \boldsymbol{\varepsilon}_{ij}, \quad \boldsymbol{\varepsilon}_{ij} \sim \mathcal{N}\left(0, \Sigma_{P_{ij}}\right),$$

where $\Sigma_{P_{ij}}$ is diagonal with entries $1/\tau^P_{\ell ij}$. The signals received are about student-school attributes for student i, not the peer.[29] The idea is that social interactions with the peer allow i to learn more about the school and infer something about how much he will benefit from different aspects of it.

The student uses this new information to update his expected utility from attending school j. Because the prior and signal are both distributed normally and because the covariance matrix for each is diagonal, the form of the posterior distribution of each student-school attribute is simple:

[28] The full expression for expected utility is $\mathrm{E}_0\left[U_{ij}\right] = -exp\left\{-\rho\left(X^0_{ij}\boldsymbol{\beta}_i - \frac{\rho}{2}\sum_{\ell}\frac{\beta^2_{\ell i}}{\tau^0_{\ell ij}} + r_j\left(n^0_{ij}\right)\right)\right\}$, but since this is strictly monotonically increasing in the terms in braces, this is equivalent to optimizing with respect to equation 1.

[29] This is in contrast with Roberts and Urban (1988), in which only quality for the peer is observed.

$$X_{\ell ij}^{1} \sim \mathcal{N}\left(\frac{\tau_{\ell ij}^{0} X_{\ell ij}^{0} + \tau_{\ell ij}^{P} P_{\ell ij}}{\tau_{\ell ij}^{0} + \tau_{\ell ij}^{P}},\ \frac{1}{\tau_{\ell ij}^{0} + \tau_{\ell ij}^{P}}\right)$$

The posterior distribution of each attribute is a precision-weighted average of the prior and signal. The expected utility from j is now

$$U_{1ij}^{*} = \widehat{X}_{ij}^{1} \boldsymbol{\beta}_i - \frac{\rho}{2} \sum_{\ell} \frac{\beta_{\ell i}^{2}}{\left(\tau_{\ell ij}^{0} + \tau_{\ell ij}^{P}\right)} + r_j\left(n_{ij}^{1}\right) \tag{2}$$

where $\widehat{X}_{ij}^{1}$ is the mean of the posterior distribution of X_{ij}^{1}. To see how the peer signals affected expected utility, compare equations 1 and 2:

$$U_{1ij}^{*} - U_{0ij}^{*} = \left(\widehat{X}_{ij}^{1} - X_{ij}^{0}\right) \boldsymbol{\beta}_i + \frac{\rho}{2} \sum_{\ell} \frac{\beta_{\ell i}^{2} \tau_{\ell ij}^{P}}{\tau_{\ell ij}^{0}\left(\tau_{\ell ij}^{0} + \tau_{\ell ij}^{P}\right)} + \left(r_j\left(n_{ij}^{1}\right) - r_j\left(n_{ij}^{0}\right)\right) \tag{3}$$

The change in expected utility comes from three sources. The first term is the change in expected match quality. This quantity may be positive or negative depending on the content of the peer signal. Students may learn that the school is a better or worse match for them than they had guessed. The second term is the change in expected utility resulting from the lower variance in the posterior distribution of match quality. This quantity is unambiguously positive. The increased knowledge about match quality works in the school's favor because the risk-averse student is now more certain about how good the match is. The third term is the change in the utility from productive knowledge, which is also positive.

This result gives rise to two testable hypotheses, derived at the end of this section:

Hypothesis 1: *The expected effect of peer information on U_{ij}^{*}, taken over all students i and schools j, is positive:* $\mathrm{E}_{ij}\left[U_{1ij}^{*} - U_{0ij}^{*}\right] > 0$.

This is the key testable hypothesis of the model that distinguishes it from models without channels through which information strictly increases expected utility. It says that, on average, receiving peer information about a school increases the expected utility from attending there. Intuitively, the signal is sometimes better than the student's prior belief and sometimes it is worse, but the average effect on expected match quality is zero. On the other hand, the reduction in uncertainty about match quality and the increase in productive knowledge always work in the school's favor.

Note that the expected effect may be positive for certain schools and negative for others, because mean quality $\overline{X}_j$ is drawn from a random distribution. This hypothesis is about the expected effect over all schools.

Hypothesis 2: *All else equal, the change in expected utility from j depends positively on how favorable the peer signal about match quality from j was:* $\frac{\partial\left(U^*_{1ij}-U^*_{0ij}\right)}{\partial P_{ij}\boldsymbol{\beta}_i} > 0$.

This hypothesis simply says that when the student receives a relatively good (i.e. high) signal about the match quality from a school, he is more likely to choose that school than if he had received a relatively bad (low) signal.

Shared attributes across schools

Students may know that the level of an attribute is shared across schools. In the empirical setting studied here, schools are divided into subsystems that share important attributes such as curriculum and vocational orientation. In this case, learning about one school in the subsystem also yields useful information about all other schools in the same subsystem. (Likewise, productive knowledge about one school might be applicable to other schools in the subsystem. I will not model this because it is now obvious that this channel will operate identically to the learning about shared attributes channel.) In order to model the shared attributes in a simple way, we can maintain all prior assumptions of the model and additionally assume that for school j in subsystem s, match quality is expressed as $X_{ijs}\boldsymbol{\beta}_i + \mu_{is}$, where $\mu_{is} \equiv \bar{\mu}_s + \widetilde{\mu}_{is}$. The average component of subsystem match quality is distributed $\bar{\mu}_s \sim \mathcal{N}\left(\bar{\mu}^0_s, \sigma^2_s\right)$ and the student-specific component is distributed $\widetilde{\mu}_{is} \sim \mathcal{N}\left(\widetilde{\mu}^0_{is}, \eta^2_{is}\right)$, and $1/\tau^{\mu}_{is} \equiv \sigma^2_s + \eta^2_{is}$. In addition to the signal P_{ij} about unshared attributes, the student receives a signal about the shared attribute:

$$q_{is} = \mu_{is} + \xi_{is}, \quad \xi_{is} \sim \mathcal{N}\left(0, 1/\tau^q_{is}\right).$$

When the student receives a signal about school j in subsystem s, he can update his expected utility from a different school k in the same subsystem:

$$U^*_{1iks} - U^*_{0iks} = \left(\widehat{\mu}^1_{is} - \mu^0_{is}\right) + \frac{\rho}{2}\frac{\tau^q_{is}}{\tau^{\mu}_{is}\left(\tau^{\mu}_{is} + \tau^q_{is}\right)} \tag{4}$$

where $\widehat{\mu}^1_{is}$ is the mean of the posterior distribution of the shared attribute and μ^0_{is} is the mean of the prior. This assumption of a shared attribute produces two additional hypotheses, derived at the end of the section:

Hypothesis 3: *The expected effect of peer information on the expected utility from any other school in the same subsystem is positive: indexing the peer's school by j and fixing another school k_j in j's subsystem s_j,* $\mathrm{E}_{ij}\left[U^*_{1ik_js_j} - U^*_{0ik_js_j}\right] > 0$.

On average, receiving a signal about a school increases the expected utility from attending other schools in the same subsystem. The intuition is the same as for Hypothesis 1. Surprises about the match quality from j's subsystem are also surprises about the match quality for all other schools in the subsystem. The surprises cancel out when we average across all schools and students. There is always a reduction in uncertainty about match quality from j's subsystem, which increases expected utility from attending schools in the subsystem.

Hypothesis 4: *Suppose the student receives a peer signal about school j in subsystem s. All else equal, the change in expected utility from school k in subsystem s depends positively on how favorable the peer signal about subsystem match quality was:* $\frac{\partial\left(U^*_{1iks} - U^*_{0iks}\right)}{\partial q_{is}} > 0$.

The more positive a surprise to the match quality for j's subsystem, the larger is the increase in expected utility from other schools in the same subsystem.

Proofs of model hypotheses

Hypothesis 1: $\mathrm{E}_{ij}\left[U^*_{1ij} - U^*_{0ij}\right] > 0$.

Proof: Equation 3 gives the expected change, over all students and schools, in expected utilities when a signal is received. The increase in productive knowledge r_j clearly increases expected utility, so I suppress the r_j terms here. This expectation is:

$$\begin{aligned}\mathrm{E}_{ij}\left[U^*_{1ij} - U^*_{0ij}\right] &= \mathrm{E}_{ij}\left[\left(\widehat{X}^1_{ij} - X^0_{ij}\right)\boldsymbol{\beta}_i + \frac{\rho}{2}\sum_{\ell}\frac{\beta^2_{\ell i}\tau^P_{\ell ij}}{\tau^0_{\ell ij}\left(\tau^0_{\ell ij} + \tau^P_{\ell ij}\right)}\right] \\ &= \mathrm{E}_{ij}\left[\widehat{X}^1_{ij}\boldsymbol{\beta}_i\right] - \mathrm{E}_{ij}\left[X^0_{ij}\boldsymbol{\beta}_i\right] + \frac{\rho}{2}\sum_{\ell}\mathrm{E}_{ij}\left[\frac{\tau^P_{\ell ij}}{\tau^0_{\ell ij}\left(\tau^0_{\ell ij} + \tau^P_{\ell ij}\right)}\right].\end{aligned}$$

From the definition of $\widehat{X}^1_{ij}$:

$$\begin{aligned}
\mathrm{E}_{ij}\left[\widehat{X}^1_{ij}\boldsymbol{\beta}_i\right] &= \mathrm{E}_{ij}\left[\sum_{\ell}\beta_{\ell i}\frac{\tau^0_{\ell ij}X^0_{\ell ij}+\tau^P_{\ell ij}P_{\ell ij}}{\tau^0_{\ell ij}+\tau^P_{\ell ij}}\right] \\
&= \sum_{\ell}\left\{\mathrm{E}_{ij}\left[\frac{\tau^0_{\ell ij}}{\tau^0_{\ell ij}+\tau^P_{\ell ij}}\beta_{\ell i}X^0_{\ell ij}\right]+\mathrm{E}_{ij}\left[\frac{\tau^P_{\ell ij}}{\tau^0_{\ell ij}+\tau^P_{\ell ij}}\beta_{\ell i}P_{\ell ij}\right]\right\} \\
&= \sum_{\ell}\left\{\mathrm{E}_{ij}\left[\frac{\tau^0_{\ell ij}}{\tau^0_{\ell ij}+\tau^P_{\ell ij}}\beta_{\ell i}X^0_{\ell ij}\right]+\mathrm{E}_{ij}\left[\frac{\tau^P_{\ell ij}}{\tau^0_{\ell ij}+\tau^P_{\ell ij}}\beta_{\ell i}\left(X_{\ell ij}+\varepsilon_{\ell ij}\right)\right]\right\} \\
&= \sum_{\ell}\left\{\mathrm{E}_{ij}\left[\frac{\tau^0_{\ell ij}}{\tau^0_{\ell ij}+\tau^P_{\ell ij}}\beta_{\ell i}X^0_{\ell ij}\right]+\mathrm{E}_{ij}\left[\frac{\tau^P_{\ell ij}}{\tau^0_{\ell ij}+\tau^P_{\ell ij}}\beta_{\ell i}\left(X_{\ell ij}\right)\right]\right\} \\
&= \sum_{\ell}\left\{\mathrm{E}_{ij}\left[\frac{\tau^0_{\ell ij}}{\tau^0_{\ell ij}+\tau^P_{\ell ij}}\beta_{\ell i}X^0_{\ell ij}\right]+\mathrm{E}_{ij}\left[\frac{\tau^P_{\ell ij}}{\tau^0_{\ell ij}+\tau^P_{\ell ij}}\beta_{\ell i}X^0_{\ell ij}\right]\right\} \\
&= \sum_{\ell}\mathrm{E}_{ij}\left[\beta_{\ell i}X^0_{\ell ij}\right] \\
&= \mathrm{E}_{ij}\left[X^0_{ij}\boldsymbol{\beta}_i\right].
\end{aligned}$$

Substituting this result back into the original equation, we have:

$$\begin{aligned}
\mathrm{E}_{ij}\left[U^*_{1ij}-U^*_{0ij}\right] &= \mathrm{E}_{ij}\left[\widehat{X}^1_{ij}\boldsymbol{\beta}_i\right]-\mathrm{E}_{ij}\left[X^0_{ij}\boldsymbol{\beta}_i\right]+\frac{\rho}{2}\sum_{\ell}\mathrm{E}_{ij}\left[\frac{\tau^P_{\ell ij}}{\tau^0_{\ell ij}\left(\tau^0_{\ell ij}+\tau^P_{\ell ij}\right)}\right] \\
&= \mathrm{E}_{ij}\left[X^0_{ij}\boldsymbol{\beta}_i\right]-\mathrm{E}_{ij}\left[X^0_{ij}\boldsymbol{\beta}_i\right]+\frac{\rho}{2}\sum_{\ell}\mathrm{E}_{ij}\left[\frac{\tau^P_{\ell ij}}{\tau^0_{\ell ij}\left(\tau^0_{\ell ij}+\tau^P_{\ell ij}\right)}\right] \\
&= \frac{\rho}{2}\sum_{\ell}\mathrm{E}_{ij}\left[\frac{\tau^P_{\ell ij}}{\tau^0_{\ell ij}\left(\tau^0_{\ell ij}+\tau^P_{\ell ij}\right)}\right] > 0
\end{aligned}$$

where the inequality holds because the τ and ρ terms are all positive by definition. ■

Hypothesis 2: All else equal, $\frac{\partial\left(U^*_{1ij}-U^*_{0ij}\right)}{\partial P_{ij}\boldsymbol{\beta}_i} > 0$.

Proof: Treating $\boldsymbol{\beta}_i$ and X^0_{ij} as fixed:

$$\frac{\partial\left(U^*_{1ij}-U^*_{0ij}\right)}{\partial P_{ij}\boldsymbol{\beta}_i}=\sum_{\ell}\frac{\partial\left(U^*_{1ij}-U^*_{0ij}\right)}{\partial P_{\ell ij}\beta_{\ell i}}=\sum_{\ell}\frac{\partial U^*_{1ij}}{\partial P_{\ell ij}\beta_{\ell i}}$$
$$=\sum_{\ell}\frac{\partial}{\partial P_{\ell ij}\beta_{\ell i}}\left(\beta_{\ell i}\frac{\tau^0_{\ell ij}X^0_{\ell ij}+\tau^P_{\ell ij}P_{\ell ij}}{\tau^0_{\ell ij}+\tau^P_{\ell ij}}\right)=\sum_{\ell}\frac{\tau^P_{\ell ij}}{\tau^0_{\ell ij}+\tau^P_{\ell ij}}>0$$

where the inequality holds because the τ terms are positive. ■

Hypothesis 3: indexing the peer's school by j and fixing another school k_j in j's subsystem s_j, $\mathrm{E}_{ij}\left[U^*_{1ik_js_j}-U^*_{0ik_js_j}\right]>0$.

Proof: This is almost identical to the proof for Hypothesis 1, except that the student is only receiving information about the shared attribute μ_{is}. From equation 4, again excluding the effect of productive knowledge, the expectation of the change in expected utility from any other school in the same subsystem is:

$$\mathrm{E}_{ij}\left[U^*_{1ik_js_j}-U^*_{0ik_js_j}\right]=\mathrm{E}_{ij}\left[\left(\widehat{\mu}^1_{is}-\mu^0_{is}\right)\right]+\mathrm{E}_{ij}\left[\frac{\rho}{2}\frac{\tau^q_{is}}{\tau^\mu_{is}\left(\tau^\mu_{is}+\tau^q_{is}\right)}\right].$$

Using the steps from the proof of Hypothesis 1, we have that $\mathrm{E}_{ij}\left[\widehat{\mu}^1_{is}\right]=\mathrm{E}_{ij}\left[\mu^0_{is}\right]$. So:

$$\begin{aligned}&\mathrm{E}_{ij}\left[\left(\widehat{\mu}^1_{is}-\mu^0_{is}\right)\right]+\mathrm{E}_{ij}\left[\frac{\rho}{2}\frac{\tau^q_{is}}{\tau^\mu_{is}\left(\tau^\mu_{is}+\tau^q_{is}\right)}\right]\\&=\mathrm{E}_{ij}\left[\left(\mu^0_{is}-\mu^0_{is}\right)\right]+\mathrm{E}_{ij}\left[\frac{\rho}{2}\frac{\tau^q_{is}}{\tau^\mu_{is}\left(\tau^\mu_{is}+\tau^q_{is}\right)}\right]\\&=\mathrm{E}_{ij}\left[\frac{\rho}{2}\frac{\tau^q_{is}}{\tau^\mu_{is}\left(\tau^\mu_{is}+\tau^q_{is}\right)}\right]>0\end{aligned}$$

where the inequality holds because the τ and ρ terms are all positive. ■

Hypothesis 4: Suppose that schools j and k are in the same subsystem s. Then all else equal, $\frac{\partial\left(U^*_{1iks}-U^*_{0iks}\right)}{\partial q_{is}}>0$.

Proof: Treating μ_{is}^0 as fixed:

$$\frac{\partial\left(U_{1iks}^{*}-U_{0iks}^{*}\right)}{\partial q_{is}}=\frac{\partial U_{1iks}^{*}}{\partial q_{is}}=\frac{\partial\widehat{\mu}_{is}^{1}}{\partial q_{is}}=\frac{\partial}{\partial q_{is}}\left(\frac{\tau_{is}^{\mu}\mu_{is}^{0}+\tau_{is}^{q}q_{is}}{\tau_{is}^{\mu}+\tau_{is}^{q}}\right)=\frac{\tau_{is}^{q}}{\tau_{is}^{\mu}+\tau_{is}^{q}}>0$$

where the inequality holds because the τ terms are positive. ■

Additional tables and figures

Table A.1: Differential effect of older sibling admission on school choice by sex pairing of siblings

	(1)	(2)	(3)	(4)	(5)	(6)
Specification	Local linear			Global		
Sample	All cutoffs	Non-elite cutoffs	Students changing subsystem at cutoff	All cutoffs	Non-elite cutoffs	Students changing subsystem at cutoff
Dependent variable	Cutoff school is first choice	Cutoff school is first non-elite choice	Schools chosen in subsystem of cutoff school	Cutoff school is first choice	Cutoff school is first non-elite choice	Schools chosen in subsystem of cutoff school
Score ≥ cutoff	0.069***	0.084***	0.251***	0.074***	0.089***	0.241***
	(0.0036)	(0.0057)	(0.0261)	(0.0029)	(0.0049)	(0.0239)
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
(Score ≥ cutoff) × (Same-sex sibling)	0.012**	0.016*	-0.024	0.008***	0.011***	0.004
	(0.0052)	(0.0081)	(0.0361)	(0.0013)	(0.0022)	(0.0116)
	[0.00]	[0.01]	[0.50]	[0.00]	[0.01]	[0.54]
Observations	340,194	163,372	224,620	1,084,111	517,565	541,571
Adjusted R-squared	0.161	0.129	0.149	0.137	0.124	0.140
Mean of dependent variable	0.167	0.194	2.069	0.153	0.176	2.024
Bandwidth	9.2	9.0	12.0	--	--	--

Note. Regressions include cutoff school-year fixed effects and polynomials in older sibling's centered exam score, all interacted with the same-sex sibling dummy. Samples in columns 3 and 6 are limited to older siblings for whom rejection from the cutoff school results in admission to a school in a different subsystem. Local linear model weights observations using the edge kernel; global model uses the rectangular kernel. Standard errors accounting for clustering at the older sibling and centered score levels are in parentheses. Stars for statistical significance are based on t-tests using (G-1) degrees of freedom, where G is the number of points of support of the centered score. Bootstrapped p-values accounting for clustering at the centered score level are in brackets.

* p<0.10, ** p<0.05, *** p<0.01

Table A.2: Predictors of choosing elite school as first choice

Specification	Probit
Dependent variable	Elite first choice
Parental education (years)	0.011***
	(0.0004)
Male	-0.002
	(0.0026)
Proportion of older sibling's middle school with elite first choice	0.451***
	(0.0086)
log(Older sibling's middle school cohort size)	0.016***
	(0.0021)
Distance from closest elite school (km)	-0.002***
	(0.0004)
Observations	98,395
Mean of dependent variable	0.760

Note. Estimates are average marginal effects from a probit regression. Sample consists of students whose older siblings were below the cutoff of an elite school and were assigned to a non-elite school as a result. Specification also includes younger sibling exam year fixed effects. Standard errors accounting for clustering at the older sibling level are in parentheses.

* p<0.10, ** p<0.05, *** p<0.01

Figure A.1: Map of COMIPEMS zone of Mexico City

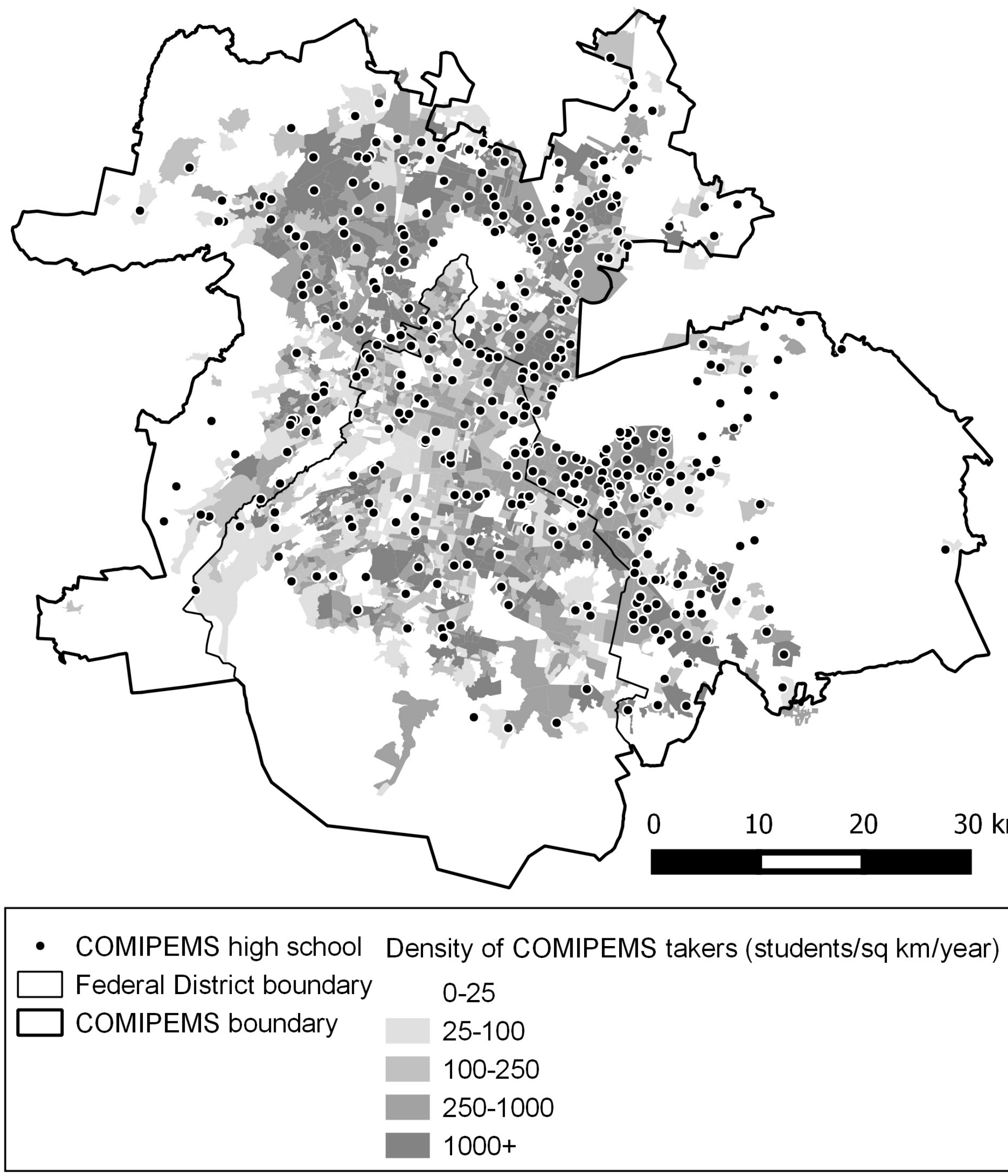
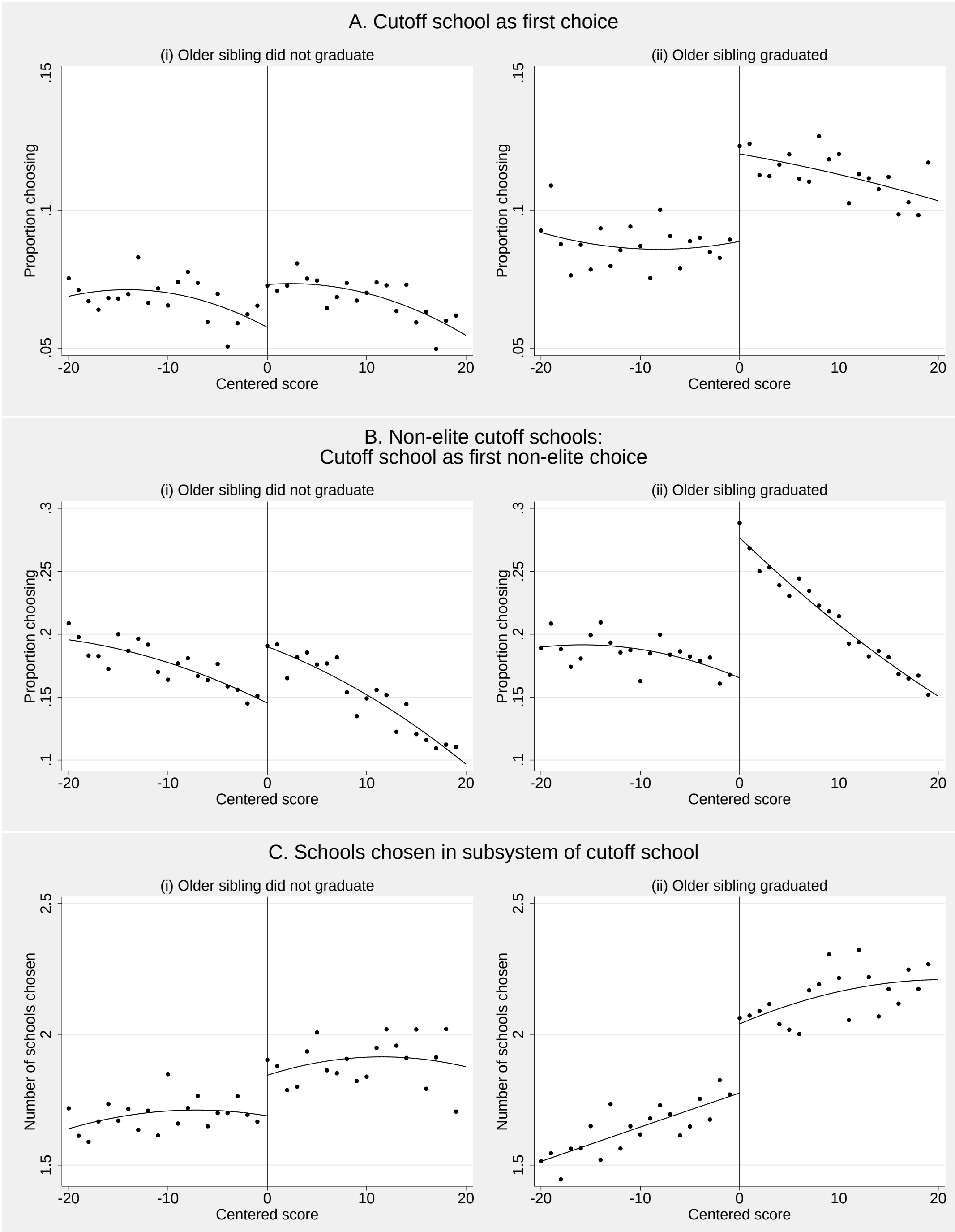


Figure A.2: Effect of older sibling admission, by graduation outcome



Note. Vertical axis pertains to the choice of the younger sibling. Number of schools chosen in Panel C excludes the cutoff school. Sample in Panel C is limited to older siblings for whom the school immediately below the cutoff belongs to a different subsystem than the cutoff school, so that marginal rejection results in assignment to another subsystem. "Graduation" is proxied by an indicator of whether the older sibling took the 12^{th} grade standardized exam. Variable on horizontal axis is older sibling's COMIPEMS exam score, centered to be 0 at the corresponding cutoff score. Fitted lines are from a quadratic fit.