Vanderbilt University Department of Economics Working Paper Series
Nashville, TN 37235-1819

Vanderbilt University Department of Economics Working Papers 15-00002

Flourish or Fail? The Risky Reward of Elite High School Admission in Mexico City

Andrew Dustan
Vanderbilt University

Alain De janvry
University of California at Berkeley

Elisabeth Sadoulet
University of California at Berkeley

Abstract

Winning admission to an elite school both promises modest rewards and imposes substantial risks on many students. Using variation in school assignment generated by the allocation mechanism, we find that admission to a system of elite public high schools in Mexico City raises end-of-high school test scores by an average of 0.17 standard deviations for the marginal admittee. On the other hand, for these students admission increases the probability of high school dropout by 9.5 percentage points. Students with weaker middle school grades and whose commute is lengthened by admission experience a larger rise in dropout probability, suggesting that the additional dropout risk is a result of both higher academic rigor and greater opportunity costs of attendance.

We thank Rafael de Hoyos, Miguel Székely, Roberto Peña and the advisers at the SEMS for their assistance with the data and understanding of the Mexico City school system.

Citation: Andrew Dustan and Alain De janvry and Elisabeth Sadoulet, (2015) ''Flourish or Fail? The Risky Reward of Elite High School Admission in Mexico City'', *Vanderbilt University Department of Economics Working Papers*, VUECON-15-00002

Contact: Andrew Dustan - andrew.dustan@vanderbilt.edu, Alain De janvry - alain@berkeley.edu, Elisabeth Sadoulet - esadoulet@berkeley.edu

Submitted: February 28. 2015 **Published:** March 28. 1015
URL: http://www.accessecon.com/Pubs/VUECON/VUECON-15-00002.pdf

1 Benefits and Risks of Attending an Elite School

Families often have some choice in where their children attend school, and all else equal, most families prefer a school of higher academic quality (see, e.g., Hastings, Kane, and Staiger 2009). Attending a "better" school, as defined by peer ability or school resources, is usually thought to benefit students academically. For example, a student may benefit from working with high-achieving and highly motivated peers and a better-funded school is able to afford more and better educational inputs. But there is also a risk to attending a better school, particularly if doing so means that the student is closer to the bottom of the school-specific ability distribution. The difficulty level of the coursework may prove too much for the student to handle. Teachers may teach mostly to the top of the class, leaving behind those who enter the school with a weaker academic background. Students experiencing such challenges may fail to complete their education at all, which is probably a much less desirable outcome than graduating from a lower-quality school.

This paper quantifies the trade-off between academic benefit and dropout risk facing students admitted to a subset of Mexico City's elite public high schools. Mexico City is ideal for this exercise for three reasons. First, there are large perceived disparities in public high school quality, with a well-identified group of "elite" schools standing above all others. This gives a natural definition of what an "elite" (or "better") school is. Second, nearly all public high schools in the city participate in a unified merit-based admissions system called COMIPEMS, using a standardized exam and students' stated preferences to allocate all students across schools. This mechanism allows us to credibly identify the impact of elite school admission on dropout probability and end-of-high school exam scores. Third, Mexico is characterized both by a high secondary school dropout rate and a significant estimated economic return to high school education, so the risk of dropping out is a first-order issue facing students. In our sample, about half of students who are assigned to a high school do not take the exit exam three years later. At the same time, young men with a high school diploma have 24% higher wages than those who only completed middle school (Campos-Vazquez 2013). Though this is not a causal relationship, it is suggestive that dropping out has a real cost for students.

A regression discontinuity design, made possible by the assignment mechanism, is used to discover whether students experience a change in dropout probability and in end-of-high school exam scores as a result of admission to an elite school, using their most-preferred non-elite school that would admit them as the counterfactual. We find that there is a clear trade-off for most marginally admitted students. Admission to an elite school raises their probability of high school dropout by 9.5 percentage points, compared to an average probability of 48%. Along with this substantial increase in dropout probability, elite school admission also results in an average gain of 0.17 standard deviations on the 12^{th} grade standardized exam, which comes mostly from gains in math. Students with lower middle school grade point averages experience larger increases in dropout probability, but there is no evidence that they experience a smaller boost in their exam scores from elite admission.

While a structural treatment of student preferences is not the subject of this paper, we present reduced-form evidence showing that students with lower performance in middle school choose elite schools less often, compared to neighboring high-performing students with the same entrance exam score. The paper's main findings offer one explanation for this result. Weak students may understand that elite school admission is a double-edged sword: while the expected academic benefit for graduates is positive, the increased chance of leaving high school without a diploma makes applying to an elite school a risky choice.

Beyond the pressure exerted on lower-achieving students, elite admission increases the opportunity cost of school attendance by inducing longer commutes. Most marginally admitted students commute a longer distance to elite schools than they would to their most-preferred alternative. Marginal admission to an elite school increases dropout probability more when admission results in a longer commute. The problem of travel distance for elite schools is not unique to Mexico City. For example, Abdulkadiroglu et al. (2014) find that students in New York City and Boston must travel farther to attend elite "exam high schools" than to their next-best option. We note, however, that commuting distance is but one factor affecting dropout risk–in our case, elite admission increases the probability of dropout even for students whose commute decreases due to admission.

Most previous studies on the effects of elite high school admission have focused on the impact on exam scores. Such studies typically analyze cases of merit-based admission systems, and use a sharp or fuzzy regression discontinuity design to estimate the effect of elite school admission on outcomes. Most have found zero or modest effects: Clark (2010) in the United Kingdom, Abdulkadiroglu et al. (2014) in Boston and New York, Lucas and Mbiti (2013) in Kenya, and Ajayi (2014) in Ghana all find zero or negligible impacts from elite high schools while Jackson (2010) and Pop-Eleches and Urquiola (2013) find a modest benefit of admission to high schools with higher-scoring peers in Trinidad and Tobago and Romania, respectively. Zhang (2012) exploits a randomized lottery for elite Chinese *middle* schools to show that elite admission has no significant impact on academic outcomes. Beyond the zero effect on exam scores, Dobbie and Freyer (2011) find that the New York elite high schools do not have an appreciable effect on long-run outcomes such as SAT score or college graduation. Estrada and Gignoux (2014) use a similar empirical strategy to ours with one year of COMIPEMS data and a separate survey (administered in a subsample of high schools) to estimate the effect of elite school admission on subjective expectations of the returns to higher education, finding that admission leads to higher expected returns. We will expand further on the relationship between their work and the present paper.

In a much different study, Duflo et al. (2011) randomly assigned Kenyan schools into a tracking regime where they divided their first grade classes by student ability. They find that while tracking is beneficial, there is no evidence that being in a class with better peers is the mechanism through which these benefits are manifested. We note that in the case of admission to competitive elite schools, admission results both in a more able peer group as well as a different schooling environment with resources, management, and culture that may be quite different from other public schools. Thus the effect of elite school admission is a reflection of both the peer and institutional channels, which regression discontinuity designs such as the present one cannot effectively disentangle.[1]

[1]Further studies on the impact of specific aspects of school quality on test scores include Dearden, Ferri, and Meghir (2002), Newhouse and Beegle (2006), Gould, Lavy, and Paserman (2004), Hastings, Kane, and Staiger (2006), Hastings and Weinstein (2008), Cullen, Jacob, and Levitt (2005 and 2006), and Lai et al. (2010).

The literature on the relationship between school quality and student dropout is sparser. Recent studies have mostly focused on the impacts of specific aspects of quality, randomly varying one aspect to see if it increased school attendance, which differs from the concept of dropout in that reduced attendance may not result in permanently abandoning schooling while dropout usually does. For example, Glewwe, Ilias, and Kremer (2010) find no effect of a teacher incentive pay scheme on student attendance in Kenyan public primary schools. More related to our study, de Hoop (2011) estimates the impact of admission to competitive, elite public secondary schools on dropout in Malawi. He finds that admission to such schools decreases dropout. This could be due to increased expected returns from an elite education inducing students to attend, or because the elite schools provide a more supportive environment. Our findings provide a stark contrast to these results, although in a much different economic and social context.

The rest of the paper is organized as follows. Section 2 gives a detailed overview of the Mexico City high school admissions system. Section 3 sets forth the method for identifying the effects of admission on outcomes. Section 4 describes the data and Section 5 gives the empirical results and several validity checks. Section 6 uses the results to rationalize revealed preference for elite schools. Section 7 concludes.

2 Mexico City public high school system and student enrollment mechanism

We first present the institutional environment in which Mexico City's students choose high schools, followed by background information on the elite schools and an explanation of how they differ from other available schooling options.

2.1 School choice in Mexico City

Beginning in 1996, the nine public high school subsystems in Mexico's Federal District and various municipalities in the State of Mexico adopted a competitive admissions process. This consortium of schools is known as the Comisión Metropolitana de Instituciones Públicas de Educación Media Superior (COMIPEMS). COMIPEMS was formed in response to the inefficient high school enrollment process at the time, in which students attempted to enroll in several schools simultaneously and then withdrew from all but the most-preferred school that had accepted them. The goal of COMIPEMS was to create a unified high school admissions system for all public high schools in the Mexico City metropolitan area that addressed such inefficiencies and increased transparency in student admissions.

Any student wishing to enroll in a public high school in the Mexico City metropolitan area must participate in the COMIPEMS admissions process. In February of the student's final year of middle school (grade nine), informational materials are distributed to students explaining the rules of the admissions system and registration begins. As part of this process, students turn in a ranked list of up to twenty high schools that they want to attend.[2] In June of that year, after all lists of preferred schools have been submitted, registered students take a comprehensive achievement examination. The exam has 128 multiple-choice questions worth one point each, covering a wide range of subject matters corresponding to the public school curriculum (Spanish, mathematics, and social and natural sciences) as well as mathematical and verbal aptitude sections that do not correspond directly to the curriculum.

After the scoring process, assignment of students to schools is carried out in July by the National Center of Evaluation for Higher Education (Ceneval), under the observation of representatives from each school subsystem and independent auditors. The assignment process is as follows. First, each school subsystem sets the maximum number of students that it will accept at each high school. Then, students are ordered by their exam scores from highest to lowest. Any student who scored below 31 points or failed to complete middle school is disqualified from participating.[3] Next, a computer program proceeds in descending order through the list of students, assigning

[2] Students actually rank programs, not schools. For example, one technical high school may offer multiple career track programs. A student may choose multiple programs at the same school. For simplicity we will use the term "school" to refer to a program throughout. No elite school has multiple programs at the same school, so this distinction is unimportant for the empirical analysis.

[3] This restriction was removed in 2013, after the period studied in this paper.

each student to his highest-ranked school with seats remaining when her turn arrives.[4] If by the time a student's turn arrives, all of his selected schools are full, he must wait until after the selection process is complete and choose from the schools with open slots remaining. This stage of the allocation takes place over several days, as unassigned students with the highest scores choose from available schools on the first day and the lowest scorers choose on the final days.

At the end of the final year of high school (grade twelve), students who are currently enrolled take a national examination called the Evaluación Nacional de Logro Académico en Centros Escolares (ENLACE), which tests students in Spanish and mathematics. This examination has no bearing on graduation or university admissions and the results have no fiscal or other consequence for high schools. It is a benchmark of student and school achievement and progress.

2.2 Elite subsystems

There are two elite high school subsystems in Mexico City, each affiliated with a prestigious national university. The Instituto Politécnico Nacional (IPN) is a university located in Mexico City that focuses on the sciences and engineering. It has 16 affiliated high schools in the city, also known for providing a rigorous education in math and science. The other elite subsystem is affiliated with the Universidad Nacional Autónoma de México (UNAM) and consists of 14 high school campuses. These schools do not stress quantitative coursework like the IPN, but rather offer a broader curriculum. Another important difference between the UNAM and IPN schools is that UNAM students who obtain a high enough grade point average are guaranteed admission to the university, while students from outside of the UNAM high schools must compete for university admission by way of a standardized entrance exam. There is an overwhelming public belief that the IPN and UNAM high schools are superior to the rest. For example, following the 2011 assignment

[4] In some cases, multiple students with the same score have requested the final seats available in a particular school, so that the number of students outnumbers the number of seats. When this happens, the representatives in attendance from the respective school subsystem must choose to either admit all of the tied applicants, slightly exceeding the initial quota, or reject all of them, taking slightly fewer students than the quota. The number of offered seats and the decisions regarding tied applicants are the only means by which administrators determine student assignment to schools; otherwise, assignment is entirely a function of the students' reported preferences and their scores. Neither seat quotas nor tie-breaking decisions offer a powerful avenue for strategically shaping a school's student body.

process, the major newspaper *El Universal* ran a story headlined "119 thousand students left out of the UNAM; Only 21 thousand middle school graduates win a spot at the IPN" (2011). [5]

The seven non-elite subsystems offer a range of educational options in their 265 campuses.[6] Some have traditional academic curricula, while others offer technical and vocational training. During the period of study, most technical and vocational schools required that students choose a track offered at the campus, so students actually faced 604 non-elite school-track choices. Figure 1 is a map of the available schools in the COMIPEMS zone, which consists of the Federal District and surrounding municipalities of the State of Mexico. While all but two of the elite schools are located in the Federal District, several of the UNAM schools and most of the IPN schools are located close to the State of Mexico and are within commuting distance of many students residing there.

While the UNAM schools are public in a sense, this subsystem refuses to administer the ENLACE exam and is legally able to do so because of its "autonomous" status.[7] The IPN, all other public subsystems, and many private schools administer the ENLACE, the latter doing so voluntarily. Because the ENLACE data provide the dependent variables for our analysis, only the effects of admission to IPN schools are examined in this paper. We will show in the data description how students attending IPN schools differ from those in the UNAM or non-elite schools, while in the empirical results we will see what bearing IPN admission has on the peer characteristics and commuting distance that students experience.

3 Regression discontinuity design and sample definition

The goal of this paper is to determine how much (marginal) admission to an IPN school changes students' probability of dropout and their end-of-high school exam scores, compared to the alter-

[5]The original title is "Fuera de la UNAM, 119 mil jóvenes; Sólo 21 mil egresados de la secundaria logran lugar en IPN."

[6]This discussion refers to the number of available schools in 2005. There have been minor changes since then.

[7]An additional difference between UNAM and other subsystems is that students selecting an UNAM school as their first choice during the COMIPEMS assignment process must take a version of the entrance exam written by UNAM, which is advertised to be equivalent to the standard version in content and difficulty.

native of admission to a non-elite school. Put another way, the econometric challenge is to estimate the effect on academic outcomes from admission to a school in an IPN subsystem instead of admission to the student's most-preferred non-elite choice, holding constant all student characteristics, observed and unobserved.

The COMIPEMS assignment mechanism permits a straightforward strategy for identifying the causal effect of IPN school admission on outcomes through a sharp regression discontinuity (RD) design. Each school that is oversubscribed (i.e., with more demand than available seats) accepts all applicants at or above some cutoff COMIPEMS exam score, and rejects all applicants scoring below that cutoff. This cutoff is set implicitly by the score of the student who obtains the final seat in that school during the sequential assignment process. If a student lists a particular school on his preference sheet and scores below the cutoff for each of his more-preferred schools, admission to that school is determined entirely by whether he scored at or above its cutoff score.[8] This generates a sharp discontinuity in the probability of admission (from 0 to 1) when the student's score reaches the cutoff.

The desired comparison is between IPN admission and non-elite admission. Thus we need to construct a sample of students such that assignment to "treatment" (admission to the IPN subsystem) depends solely on whether a student's COMIPEMS score exceeds a predetermined cutoff. To achieve this, we first identify, for each student, the minimum COMIPEMS exam score that the student could obtain and still be assigned to an IPN school. This student-specific IPN admission cutoff score is known because the student's stated preferences, combined with the cutoff scores for each school, fully determine the student's assignment for any point value of the COMIPEMS score.[9] If the IPN admission cutoff for a student is undefined because no COMIPEMS score would result in IPN assignment, then he is dropped from the sample.

In the sharp RD design employed here, a score exceeding the IPN admission cutoff implies treatment with probability of one. To obtain this outcome in the RD sample, we exclude any student

[8]The elite schools automatically reject all students with a grade point average below 7 out of 10. Very few students score high enough for admission and fail to meet this requirement.

[9]For example, assuming the student obtains a score of 70, the student's assignment would be his highest-ranked school that has a cutoff score of 70 or below.

who would be admitted to a non-IPN school for any point value exceeding the IPN admission cutoff. For example, a student might select an UNAM school with a cutoff score of 80 as his first choice and an IPN school with a cutoff of 70 as his second choice. In this case, COMIPEMS scores of 80 and above would lead to UNAM assignment while scores from 70 to 79 would lead to IPN assignment. Such students are excluded from the RD sample. This restriction implies that all students in the RD sample chose an IPN school as their most-preferred option, so we might think of the RD sample as consisting of students with a relatively strong preference for IPN schools.

Finally, we want to ensure that scoring below the IPN admission cutoff score leads to non-elite assignment. While by construction no score below this cutoff can result in IPN assignment, we exclude any student whose stated choices are such that he could obtain a score below the IPN admission cutoff and still be admitted to an UNAM school.[10] This could happen if, for example, the student's first choice was an IPN school with a cutoff of 80 and his second choice was an UNAM school with a cutoff of 70. These three sample restrictions–existence of an IPN admission cutoff score, no non-IPN school assignments possible above this cutoff, and only non-elite school assignments below this cutoff–result in an RD sample where the probability of elite (IPN) assignment is zero for all COMIPEMS scores below the IPN admission cutoff and one for all COMIPEMS scores above it.

Note that different scores above the student's IPN admission cutoff could result in assignment to different IPN schools–for example, a score of 70 may be enough for one requested IPN school, while a score of 75 would be sufficient for admission to a more-preferred IPN school. This does not pose a problem for the RD design because the treatment is defined as assignment to any IPN school, not only to the school that corresponds to the student's IPN admission cutoff. It will be useful at times in this paper to discuss this latter school, however, which we will refer to as the "cutoff school." Similarly, different COMIPEMS scores below the cutoff may result in assignment to various non-elite schools. We will refer to the school directly below the cutoff, i.e. the school

[10]There are two reasons for this restriction. First, the UNAM system is elite, and we want to estimate the impact of IPN admission versus the counterfactual of non-elite admission. Second, the UNAM is missing data on graduation and test scores, so we could not include these students in the sample even if we wanted to make this comparison.

assignment for a score one point below the IPN admission cutoff, as the "next-best" school. To summarize, each student is characterized by three things: his cutoff school (the lowest-cutoff IPN school he could attend, given his choices), his next-best school (the most-preferred non-elite school he could attend if he scored too low for IPN admission), and the cutoff score such that he would always be admitted to an IPN school if his COMIPEMS score were equal to or greater than this cutoff and would never be admitted to an IPN school if his COMIPEMS score were less than the cutoff.

For each student i in the RD sample in exam year t, we index the cutoff school by j. Following Abdulkadiroglu et al. (2014), we use a stacked nonparametric RD design that estimates, for students with a score close to the relevant cutoff, a single average admission effect over all cutoff schools while controlling for separate linear terms in the COMIPEMS score for each cutoff school. The estimating equation is:

$$Y_{ijt} = \delta admit_i + \gamma_{1j}\left(c_i - \underline{c}_{jt}\right) + \gamma_{2j}\left(c_i - \underline{c}_{jt}\right) admit_i + \mu_j + \varepsilon_{ijt} \tag{1}$$

where Y_{ijt} is the outcome of interest (dropout or ENLACE exam score), $c_i - \underline{c}_{jt}$ (the "centered" COMIPEMS score) is the difference between i's COMIPEMS score and j's cutoff score in year t, and $admit_i = 1$ if $c_i - \underline{c}_{jt} \geq 0$. The parameter of interest is δ, the local average treatment effect of being admitted to an IPN school instead of a non-elite school (Imbens and Lemieux 2008). This is an intention-to-treat effect since students do not necessarily attend a school in the subsystem to which they were admitted. But in practice, compliance is almost perfect. Of those in the RD sample who take the ENLACE exam, 99.8% of the students rejected from the IPN subsystem take the exam in a non-elite school, while 96.1% of ENLACE exam-takers who were admitted to an IPN school take the exam in an IPN school.

We use the bandwidth selection procedure suggested by Imbens and Kalyanaraman (2012) and, following the same authors, use the edge kernel in estimating the local linear regressions.[11] The

[11]The edge kernel is $K_h\left(c_i - \underline{c}_{jt}\right) = \mathbb{1}\left(\left|c_i - \underline{c}_{jt}\right| \leq h\right)\left(1 - \frac{c_i - \underline{c}_{jt}}{h}\right)$, where h is the bandwidth. We select the optimal bandwidth while omitting the cutoff fixed effects and using a single set of piecewise-linear terms instead of separate

running variable, centered COMIPEMS score, is discrete since the COMIPEMS exam is scored in one-point increments from 0 to 128. Per Lee and Card (2008), we cluster our standard errors at the level of the centered score in order to account for specification error in the local polynomials. Because there are relatively few clusters and analytic clustered standard errors may be downward-biased in this case, wild-cluster bootstrapped p-values are presented in addition to robust, unclustered standard errors (Cameron et al. 2008).

An advantage of the RD design is that it does not require any assumptions about the decision-making process by which students choose schools and whether their rankings of schools truly represent revealed preferences. Conditional on COMIPEMS score, the admitted and rejected students near a school's cutoff have the same expected characteristics, including preferences over schools. Even if students are trying to choose strategically or making mistakes in their selections, this behavior will not differ by admissions outcome near the cutoff. We can thus remain agnostic on the issue of the distribution of student preferences and the factors that influence them.

4 Data description

The data used in this paper come from two sources, both obtained from the Subsecretariat of Secondary Education of Mexico: the registration, scoring, and assignment data for the 2005 and 2006 COMIPEMS entrance examination processes, and the scores from the 2008, 2009, and 2010 12$^{\text{th}}$ grade ENLACE exams.[12] The COMIPEMS dataset includes all students who registered for the exam, with their complete ranked listing of up to twenty high school preferences, basic background information such as middle school grade point average and gender, exam score out of 128 points, and the school to which the student was assigned as a result of the assignment process. It also includes student responses to a multiple choice demographic survey turned in at the time of

sets for each cutoff school. Because the fixed effects and additional linear terms have very little explanatory power in most of these regressions, omitting them has little effect on the selected bandwidth. Having selected the bandwidth, we estimate equation 1 including the fixed effects and cutoff school-specific linear terms.

[12]The 2010 data is used in order to match students from the 2006 COMIPEMS cohort who took four years to complete high school instead of three.

registration for the exam.

The ENLACE dataset consists of exam scores for all students who took the test in Spring 2008 (the first year that the 12^{th} grade ENLACE was given), 2009, or 2010. The scores for both the math and Spanish sections are reported as a continuous variable, reflecting the weighting of raw scores by question difficulty and other factors. We normalize the scores by subtracting off the year-specific mean score for all examinees in public high schools within the COMIPEMS geographic area and dividing by the year-specific standard deviation from this same sample. The ENLACE scores are matched with the 2005 and 2006 COMIPEMS-takers by using the Clave Única de Registro de Población (CURP), a unique identifier assigned to all Mexican citizens. Matching is performed by name and date of birth if no CURP match is found and, following that, further matching is performed on name and assigned school. The matching rate of ENLACE takers to their COMIPEMS scores is nearly 100% and will be discussed further in section 5.5.

We limit the sample to applicants who graduated from a public middle school in Mexico City in the year that they took the COMIPEMS exam. The IPN schools are highly-demanded among these students. For every seat available in an IPN school, 1.9 students list an IPN school as their first choice. Every IPN school is oversubscribed. Figure 2 shows the distribution of cutoff scores for all oversubscribed schools. Panel (a) shows that, along with the UNAM schools, the IPN schools have far higher cutoff scores than the vast majority of non-elite schools. Panel (b) weights the cutoff schools by the number of students in attendance, showing that nearly all students in a high-cutoff school are in the IPN or UNAM subsystems.

Table 1 presents summary statistics for the sample of all students, the subsamples of students who attended the IPN, UNAM, and non-elite systems, and students meeting the criteria for inclusion in the RD sample. Students attending IPN schools are quite different from those at non-elite schools. IPN's student body has higher average COMIPEMS exam scores (88.0 points vs. 57.7), grade point (8.54/10 vs. 7.96/10), parental education (11.4 years vs. 9.8), family income (5,210 pesos/month vs. 3,850), and ENLACE exam scores (1.12 normalized score vs. -0.18).[13] Students

[13]There is no binding test score ceiling for either exam. Score ceilings present a problem for academic gains because there is no way for students with the highest score to demonstrate progress. The COMIPEMS exam intentionally

commute an average 4.33 kilometers farther to IPN schools than non-elite options.[14] Another notable contrast is that while 2/3 of IPN students are male, fewer than half of students in the non-elite systems are. This is due to higher preference for the IPN schools among males, perhaps because of the polytechnic focus of the curriculum. On the other hand, IPN students are similar to students from the UNAM schools on most dimensions, including COMIPEMS score, middle school GPA, and family background. Again, though, the IPN student body is more male-dominated than the UNAM.

The RD sample is described in column 5. There are 41,075 students who meet these criteria. As expected, the mean characteristics for this group fall between the IPN and non-elite samples. How much did each restriction on the RD sample affect the sample size? We start by discarding students who could not be assigned to an IPN school for *any* possible COMIPEMS score; 76,738 students remain. Dropping students who would be assigned to a non-IPN school for some COMIPEMS scores above the IPN admission cutoff eliminates 26,348 students. Of these, 26,161 were dropped because some COMIPEMS scores above the cutoff would result in UNAM assignment. Finally, 9,315 students are dropped because they would be assigned to an UNAM school for some COMIPEMS scores below the IPN cutoff.

It is clear from Table 1 that many COMIPEMS exam takers do not take the ENLACE. We will present evidence in section 5.3 that this is almost entirely due to student dropout rather than some other feature of the data. For the moment, we treat non-taking as dropout and show in Table 2 that dropout is predicted both by academic ability and IPN admission. Column 1 shows that, in the cross-section, COMIPEMS exam score and middle school grade point average (GPA) are negatively correlated with dropout. Particularly striking is the GPA coefficient, showing that a one standard deviation (0.82) increase in GPA predicts a 14 percentage point decrease in dropout probability. Parental education is negatively correlated with dropout as well, but the magnitude of the coefficient is very small compared to those of COMIPEMS and GPA. Column 2 adds high

avoids a ceiling in order to sort students during assignment.

[14] Distance is computed as the straight-line distance from the centroid of the student's postal code to the location of the assigned school.

school fixed effects and shows that these relationships are similar within a high school. Column 3 adds a control for commuting distance, which is missing in about 14% of cases due to an inability to match students' reported postal codes with geographical coordinates. Here we see that commuting distance positively predicts dropout: a 10 km increase in commute predicts a 2.7 percentage point increase in dropout probability. Column 4 shows that, conditional on listing an IPN school as one's first choice, dropout is much higher for students admitted to IPN schools than for those admitted to non-elite schools. This correlation does not have a causal interpretation, however, because unobservable student attributes could affect both selection into an IPN school and dropout probability. The next section uses the RD design to establish the causal IPN admission-dropout relationship.

5 Effects of elite school admission

This section uses the RD design outlined in Section 3 to estimate the effect of marginal admission to an IPN school on the probability of dropping out of high school and, conditional on taking the ENLACE exam, on the exam score obtained. Because we lack individual-level data on graduation, taking the ENLACE exam is used as a proxy for graduation. Only students on track to graduate at the end of the school year are registered to take the exam. We present evidence in section 5.5 that this is a good proxy, in particular that schools do not strategically administer this exam. Thus the only sample used from this point forward is the RD sample as defined in section 3.

5.1 School characteristics and commute

Before presenting the effects of IPN admission on dropout and test scores, we show that admission results in students attending a school with drastically more able peers while also commuting a longer distance to reach school. Table 3 and corresponding Figure 3 show the results from estimating the local linear version of Equation 1 with peer characteristics and commute distance as the dependent variables.[15] On average, marginal IPN admission implies assignment to a school where peers scored 20.3 COMIPEMS points (more than one standard deviation) higher than the next-best school. Peers also have, on average, middle school GPAs 0.52 points (0.62 standard deviations) higher than the next-best school and have parents with 1.2 additional years of education. Students also experience longer commutes due to IPN admission, traveling 4.5 km farther in each direction, nearly 50% more than the RD sample average. Thus IPN admission, on average, exposes students to much "better" peers while requiring a longer commute.

5.2 Probability of dropout

Marginal admission to an IPN school significantly increases the probability of dropout. Figure 4 illustrates this graphically, plotting the dropout rate in a 20 point window around the IPN admission cutoff. Table 4 confirms this finding, reporting the average effect of admission on dropout estimated using Equation 1 for the optimal bandwidth (column 1). The estimated dropout effect is large, 9.5 percentage points compared to a dropout rate of about 44 percentage points in the RD sample.[16] This result is robust across different bandwidth selections: estimates using half (column 2) and double (column 3) the optimal bandwidth are 9.2 and 11.0 percentage points, respectively. We note that the optimal bandwidth is 15.3 COMIPEMS points, somewhat less than one standard deviation of this score in the RD sample over which this bandwidth is computed (18.49 points). Use of the edge kernel puts more weight on data near the cutoff, so 55% of the summed weights come from observations within 5 points of the cutoff score.

The increase in dropout is accompanied by a higher rate of delayed high school completion, as shown in column 4. The dependent variable in this regression is a dummy equal to one if the student either dropped out (did not take the ENLACE) or took the ENLACE more than three years after participating in the admissions process, indicating a delay of one or more years. The

[15] Results from local quadratic regressions are similar for these and all other regressions in the paper. Appendix C contains tables reproducing all results using local quadratic specifications.

[16] Our estimates for the effect of admission on dropout are larger than those found in Estrada and Gignoux (2014). Appendix Table A1 and its accompanying text give insight into these differences, but in brief, we view the difference in results as coming from differences in the samples used rather than from a difference in methods.

estimated effect of IPN admission on dropout or delay is 12.5 percentage points, three percentage points higher than the estimated impact on dropout alone.

There is important heterogeneity behind the average effect of IPN admission on dropout. Figure 5 and corresponding Table 5 present these results, which are estimated using the following equation:

$$\begin{aligned} Y_{ijt} = \delta admit_i + \gamma_{1j}\left(c_i - \underline{c}_{jt}\right) + \gamma_{2j}\left(c_i - \underline{c}_{jt}\right) admit_i + \mu_j + \\ z_{ijt}\left[\tilde{\delta} admit_i + \tilde{\gamma}_{1j}\left(c_i - \underline{c}_{jt}\right) + \tilde{\gamma}_{2j}\left(c_i - \underline{c}_{jt}\right) admit_i + \tilde{\mu}_j\right] + \varepsilon_{ijt} \end{aligned} \tag{2}$$

where z_{ijt} is a dummy variable representing some dimension of heterogeneity in the admission effect. The point estimate of $\tilde{\delta}$ is identical to the difference between admission coefficients obtained from estimating equation 1 separately for each value of z_{ijt}.

Students with a low middle school GPA (below the sample mean among IPN students) experience a 6.9 percentage point higher increase in dropout probability than those students with a high GPA (column 1). This suggests that an important driver of dropout for (marginal) IPN students is the academic difficulties that accompany being a relatively weak student in a demanding school. Column 2 fails to find any heterogeneity with respect to parental education level, although the confidence interval contains both large positive and negative values.

Column 3 gives results for the differential effects with respect to changes in commuting distance.[17] The "change in commute" variable is constructed by subtracting the commuting distance to the next-best school from the commuting distance to the IPN cutoff school. We then partition the sample based on whether the student's commute would become longer or shorter as a result of IPN admission. About 71% of students in the RD sample have a longer commute as a result of IPN admission. For both cases, the probability of dropout increases, but students induced into a longer commute experience a larger increase (0.128 vs. 0.058). We note that the robust standard errors are unexpectedly more conservative than the wild-cluster bootstrapped p-values in this case,

[17]For this regression, students who would go unassigned during the computerized assignment process and would have to choose later from schools that had not filled yet are excluded.

although even robust standard errors indicate statistical significance of each coefficient at the 10% level.

Column 4 repeats the commute distance differential construction exercise, except that now the differential is with respect to the mean COMIPEMS scores of the incoming high school cohort. We do not find evidence that results are driven by admission for students who experience an above-average increase in peer quality as a result of admission, although again the confidence interval permits substantial heterogeneity on this dimension.

These results make clear that dropout is systematically related to IPN admission and its interaction with academic ability as proxied by middle school GPA. Students admitted to an IPN school are on average more likely to drop out and thus less likely to take the ENLACE, such that even after conditioning on COMIPEMS score, IPN admittees taking the ENLACE have higher middle school GPAs. To show this, we estimate the following equation for each of the student characteristics x_{ijtk}:

$$x_{ijtk} = \phi_k admit_i + \beta_{1k}\left(c_i - \underline{c}_{jt}\right) + \beta_{2k}\left(c_i - \underline{c}_{jt}\right) admit_i + \mu_j + \varepsilon_{ijtk} \quad (3)$$

If x_k is balanced across the cutoff, then $\hat{\phi}_k$ should be close to zero. Table 6, Panel (a) and accompanying Figure 6 give estimates at the time of assignment (prior to dropout), where we expect balance. Of the seven covariates tested, none are found to change discontinuously at the cutoff. When estimating the equations jointly using seemingly unrelated regression and performing a joint test for discontinuities, we fail to reject the null hypothesis of no discontinuity ($p = 0.58$). Panel (b), however, shows that within the sample of ENLACE takers middle school GPA is unbalanced (about $1/10$ standard deviations higher for admitted students) as well as parental education and hours studied. The p-value for the joint test of discontinuities is 0.01. Hence dropout among marginally admitted students is not only higher than among the rejected, but it is also heterogeneous with respect to student characteristics. This differential dropout may bias upward estimates of the IPN admission effect on ENLACE exam scores if the additional dropout is among the students who would have the lowest ENLACE scores. We will need to bound the estimated ENLACE effects to account for this possibility.

5.3 ENLACE exam performance

We now turn to the effect of IPN admission on the standardized ENLACE exam score. We first ignore the differential dropout issue raised in the previous section and then bound the effects while accounting for dropout. Using all observed scores, Figure 7, Panel (a) suggests that there is a significant, positive effect of IPN admission on ENLACE score. Panels (b) and (c) show that this effect comes almost entirely from improved math scores. This result may be unsurprising given that IPN schools focus heavily on mathematics, engineering, and the sciences in their curriculum. Table 7 reports the RD estimates of these relationships for the optimal bandwidth (column 1) and both half and double this bandwidth (columns 2 and 3, respectively). Again, the results are robust to the choice of bandwidth: the local linear results range from an effect of 0.16 to 0.17 standard deviations of the composite (math and Spanish) score. The estimated effects on math scores range from 0.22 to 0.25 standard deviations, while the Spanish estimates are positive but statistically insignificant.

We address the potential for bias due to differential dropout in two ways.[18] First, we apply the sharp bounds approach proposed by Lee (2009) to the RD design. In the context of a randomized controlled trial, the Lee bounds process begins by estimating the degree of differential attrition between treatment and control groups, trimming observations from the group (treatment or control) with lower attrition in order to balance the attrition rates. Trimming is accomplished either by dropping observations with the highest values of the outcome variable (to obtain a lower bound on the treatment effect) or with the lowest values (to obtain an upper bound). Estimation of the original relationship of interest is then carried out using the trimmed sample in order to obtain upper and lower bounds on the treatment effect. In order to apply this procedure to an RD design, we assume that the dropout effect is constant within the selected bandwidth. This allows us to trim the same proportion of rejected students for each value of the centered COMIPEMS score,

[18]The high rate of dropout in the sample makes Horowitz and Manski (2000) nonparametric bounds uninformative.

since excess dropout was among the admitted students. Because we are interested in a lower bound for the treatment effects given the apparent positive selection into ENLACE-taking, we trim the worst-performing students for each value of centered COMIPEMS score. We then carry out the RD estimation procedure with the trimmed sample. Standard errors are bootstrapped, where each repetition includes the dropout effect stage, the subsequent trimming based on the estimated differential dropout, and the final estimation of the lower bound.

Despite the extreme approach of trimming the worst-performing students, the point estimates for the admission effect on the composite and math scores are both positive. The estimated lower bound for the composite score is modest, 0.044 ($SE = 0.032$) with a p-value of 0.16, while for math it is larger and strongly significant: 0.12 ($SE = 0.035$). The Spanish bound is negative: -0.11 ($SE = 0.038$), as expected since the original point estimate was small. We take this as strong evidence of a positive math score effect and weaker evidence for a positive effect on the composite score.

A second approach is to estimate the effect of admission on the probability of taking the ENLACE exam and obtaining at least a pre-specified score on the exam. This is equivalent to imputing an arbitrarily negative score for non-takers and estimating the effect of admission on the probability of exceeding a particular ENLACE score. Figure 8 shows the estimated admission coefficients from local linear regressions, plotted over the range of possible ENLACE scores. Panel (a) plots the composite (math and Spanish) score and shows the admission effect becoming positive at a score of 0.3, although this difference remains statistically insignificant for all but scores of 1.4-1.5. The math score effects in Panel (b), on the other hand, are positive beginning at a score of 0 and are significant for scores of the range 0.7-2.1. As expected, the Spanish effects are negative for most scores, although the point estimates do become positive at a score of 1.3. Hence, particularly for math, the results are consistent with elite admission increasing the probability of graduating with a high ENLACE score while simultaneously increasing dropout and therefore decreasing the probability of graduating with a low score.

The dropout results showed striking heterogeneity, in particular with respect to middle school

GPA and changes in commuting distance. We repeat this exercise for ENLACE score in Table 8, interpreting with caution because these estimates may be biased due to the differential dropout that has been documented thus far.[19] The only significant source of heterogeneity here is with respect to changes in commuting distance: the effect of admission is estimated to be 0.285 for students whose commutes decrease and $0.285 - 0.170 = 0.115$ for students whose commutes increase. If longer commutes were simply leading to dropout among the worst students, then we would expect to see larger ENLACE effects among those induced to commute farther. Instead, the evidence suggests that the longer commute mitigates some of the academic benefits from attending an IPN school. It is also possible that students facing different changes in commutes due to admission are different in other ways so that IPN admission has differential effects for some reason other than travel distance.

5.4 Effects of admission to a higher-cutoff IPN school

In order to gain further insight into why IPN admission affects dropout and test scores, we briefly investigate the effects of being admitted to a higher-cutoff IPN school, compared to the counterfactual of admission to a lower-cutoff IPN school. We begin with the already-described RD sample and identify, separately for each IPN school, the corresponding school-specific sample. The sample for IPN school A consists of students whose counterfactual assignment is to A for COMIPEMS score equal to A's cutoff score and whose counterfactual assignment is to another IPN school for the COMIPEMS score one below A's cutoff. These are students who, very near the cutoff score, are either barely admitted to A or barely rejected and sent to a different IPN school. Having constructed such a sample for each IPN school, we stack the school-specific samples and estimate Equation 1. Because students may belong to more than one school's sample, we cluster standard errors at the student level.

Table 9 begins by showing that admission to a higher-cutoff IPN school results in a somewhat different peer group: the mean peer COMIPEMS score is 4.7 points higher (compared to the 20

[19] Results for the math and Spanish scores are in Appendix Figure B1.

point jump from non-IPN to IPN schools), while mean peer middle school GPA is 0.12 points greater and peers' parents have on average 0.31 years more of education. On average, students commute 2.3 km less due to admission, in contrast with the increased commute due to admission at the IPN/non-IPN boundary. The point estimate for the effect of admission on dropout is 2 percentage points, but the 95% confidence interval ranges from -1.8 to 5.8 percentage points. Thus it is unclear how admission to a "better" IPN school affects dropout probability, except that we can rule out effects as large as those from the non-IPN to IPN jump. On the other hand, the estimated admission effect on ENLACE scores is 0.075 standard deviations and is significantly different from zero. It seems that students do benefit marginally from attending a higher-cutoff IPN school, at least in terms of ENLACE performance.

5.5 Validity checks

Here we present two sets of validity checks to address potential concerns with the results. First, support for the validity of the RD design is given. Second, support is given for the assertion that the dropout-related results in this paper are indeed due to IPN students leaving school at a higher rate, rather than a data issue.

There is no a priori reason to think that the RD design might be invalid. Because the school-specific cutoff scores are determined in the process of the computerized assignment process, monitored by school subsystem representatives and independent auditors, there is no opportunity for student scores to be manipulated in order to push particular students from marginal rejection to marginal admission. Nevertheless, Figure 9 provides graphical evidence of the design's validity, showing the distribution of centered COMIPEMS scores for students in the RD sample. Panel (a) shows the entire density, while Panel (b) zooms in on a smaller window around the cutoff. There is no visual evidence for a jump in the density of COMIPEMS score to one side of the cutoff or the other. We test formally for bunching in the density, following McCrary (2008). The p-value for this test is 0.90, in agreement with the visual evidence presented.

As further support for the RD design, we recall the balance of baseline covariates across the admission cutoff shown in Figure 6 and Table 6, Panel (a). The lack of a discontinuity in these covariates suggests that students were unable to sort into or out of IPN admission, as we would expect given the computerized assignment process.

There is substantial evidence that the difference in ENLACE taking rate between students admitted to and rejected from the IPN is due to students dropping out of school, rather than a data problem or the rate at which 12th graders in IPN schools take the ENLACE exam. The difference cannot be due to a lower rate of success in matching ENLACE takers from IPN schools to their COMIPEMS score. Of all ENLACE takers in IPN schools in 2010, 99% are matched successfully to their COMIPEMS scores in 2005, 2006, or 2007. Another possibility that we can dismiss is that the IPN is selectively administering the exam to its best 12th graders. Although the ENLACE is taken at the end of the school year, schools must report the full roster of students in their final academic year to the Secretariat of Education so that all of those students can be programmed to take the exam. The ratio of actual exam takers to those programmed in the fall is nearly identical between the IPN and non-IPN schools (81%). Thus differential exam taking would have to be sufficiently premeditated to 1) fail to register low-ability students in the Fall and 2) systematically prevent the unregistered students from showing up at the exam. The exam is given by proctors from outside of the school. Administrators who run the ENLACE express doubt that a school subsystem would go through this trouble, especially when considering that ENLACE scores are not used to allocate resources or to incentivize or punish educators. Finally, because the ENLACE dataset used in this paper includes years 2008 through 2010, it captures COMIPEMS takers from 2005 who took four or five years to graduate, and COMIPEMS takers from 2006 who took four years to graduate, instead of the standard three years. The differential exam taking rate, then, cannot be explained by students taking longer to graduate in the IPN schools but not dropping out.

As with any study using a RD approach, there may be some skepticism in extrapolating the effects for marginal students to the rest of the sample. This would be a particular concern if there were few students near the margin compared to the total population of IPN students. The nature of the assignment mechanism, however, tends to bunch students near the cutoff of the school to which they are admitted, since a modestly higher score would often lead to admission to a more-preferred school. Similarly, many of the students admitted to the IPN *subsystem* are only a few points away from rejection to a non-IPN school. In fact, 34% of students admitted to an IPN school are within 7 COMIPEMS points of falling out of the IPN subsystem, while more than half are within 12 points of the boundary. The standard deviation of COMIPEMS score in the full sample is 17.95 and the within-school standard deviation for IPN students is 7.19, implying that a significant portion of IPN students are not far from the margin of the IPN subsystem.

6 Preference for elite schools

Students with lower GPAs are less likely to apply to elite schools. The findings in this paper offer one way of rationalizing this empirical regularity. Students with a weak academic background face a less desirable dropout risk-academic reward trade-off and may respond rationally by choosing to avoid it altogether. This should be particularly true for students who are likely to gain admission to an elite school only at the margin.

To show that conditional on COMIPEMS score, high-achieving students are more likely to list an elite school as their first choice, the following local linear regressions are estimated for all observations within a 2-point bandwidth of each COMIPEMS point value c:

$$elite_{imtc} = \alpha_{mtc} + \beta_c COMIPEMS_i + \theta_c GPA_i + \varepsilon_{imtc}, \tag{4}$$

where $elite_{imt}$ is a dummy variable equal to 1 if student i in year t from municipality/delegation m chose an elite school as her first choice, and GPA_i is middle school GPA. The municipality/delegation of residence of the student is added to control for the possible unequal geographic access to elite schools. The parameters of interest are the θ_c's, which measure the marginal effect (though not a causal relationship) of GPA on elite school preference only for students with $COMIPEMS_i$ near c. Figure 10 graphs these coefficients and shows that for all values of COMIPEMS score above 65 points, i.e., that are high enough to gain admission to the least-competitive elite school, a higher GPA is correlated with higher rates of elite school preference. At a COMIPEMS score of 80, students with a 9.0 GPA are 15 percentage points more likely to select an elite school than those with a 7.0 GPA. This is a large difference, indicating that among students living in the same municipality or delegation and with the same possibility of admission to elite schools as a result of their COMIPEMS score, those with a lower GPA are much less likely to list an elite school as a first choice. The less favorable risk-reward tradeoff facing these students offers one way to explain this result.

7 Discussion

This paper used Mexico City's high school allocation mechanism to identify the effects of admission to a subset of its elite public schools, relative to their non-elite counterparts. At least for marginally admitted students, elite schools present an important trade-off. Elite admission appears to positively affect student test scores, increasing end-of-high school exam scores by 0.17 standard deviations under the assumption that dropout does not bias the estimated effect. However, admission is found to significantly increase the probability of dropping out of school. Students with relatively low middle school GPAs and who are induced to commute farther are especially affected, suggesting that elite schools are too challenging or far away for some students and they either fail out or elect to leave school. Allowing for bias due to differential dropout lowers this estimate, but the results are quite robust when examining the potential effects of this bias. In particular, students' math scores seem to improve significantly with attending an IPN school. The fact that this trade-off is, in expectation, worse for those from weaker academic backgrounds offers one possible explanation for the lower rate at which qualified students with low GPAs apply to elite high schools.

The existence of this trade-off between academic benefit and dropout probability highlights an important educational policy issue in Mexico. The current configuration of the high school education system does not facilitate lateral transfers of students between school subsystems, which are run by numerous entities at the local, state, and national level. Students who find that their current school is a bad fit cannot easily switch to a school that balances academic rigor, curriculum, and other characteristics to their taste, unless they drop out of school entirely and attempt to begin anew elsewhere. The Comprehensive High School Education Reform (RIEMS) represents somewhat of an attempt to rectify this by imposing a (partial) common curriculum, but this reform has faced delays and political opposition and its future remains in question. Such rigidity in the current system may explain why the academic benefit-dropout trade-off is so strong in this paper in comparison to studies in other countries. Our result highlights the value of flexibility in choice-based admissions systems so that the consequences of a "bad" choice can be mitigated, provided that lateral transfers to more competitive schools are not allowed as a means of gaming the current system.

References

"Fuera de la UNAM, 119 mil jóvenes." *El Universal* 28 July 2011. Retrieved from `http://search.proquest.com/docview/879688183`.

Abdulkadiroglu, Atila, Joshua Angrist, and Parag Pathak, "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools," *Econometrica* 82 (2014), 137-196.

Ajayi, Kehinde, "Does School Quality Improve Student Performance? New Evidence from Ghana," Unpublished working paper (2014).

Altonji, Joseph, Todd Elder, and Christopher Taber, "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy* 113 (2005), 151-184.

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

Campos-Vázquez, Raymundo M., "Why Did Wage Inequality Decrease in Mexico after NAFTA?," *Economía Mexicana NUEVA EPOCA* 22 (2013), 245-278.

Clark, Damon, "Selective Schools and Academic Achievement," *The B.E. Journal of Economic*

Analysis & Policy (Advances) 10 (2010), Article 9.

Cullen, Julie, Brian Jacob, and Steven Levitt, "The Impact of School Choice on Student Outcomes: An Analysis of the Chicago Public Schools," *Journal of Public Economics* 89 (2005), 729-760.

Cullen, Julie, Brian Jacob, and Steven Levitt, "The Effect of School Choice on Participants: Evidence from Randomized Lotteries," *Econometrica* 74 (2006): 1191-1230.

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

Dearden, Lorraine, Javier Ferri, and Costas Meghir, "The Effect of School Quality on Educational Attainment and Wages," *The Review of Economics and Statistics* 84 (2002), 1-20.

Dobbie, Will, and Roland Fryer, Jr., "Exam High Schools and Academic Achievement: Evidence from New York City," NBER Working Paper 17286 (2011).

Duflo, Esther, Pascaline Dupas, and Michael Kremer, "Peer Effects and the Impacts of Tracking: Evidence from a Randomized Evaluation in Kenya." *American Economic Review* 101 (2011), 1739–74.

Estrada, Ricardo, and Jeremie Gignoux, "Benefits to Elite Schools and the Formation of Expected Returns to Education: Evidence from Mexico City," Working Paper 2014-06, Paris School of Economics (2014).

Glewwe, Paul, Nauman Ilias, and Michael Kremer, "Teacher Incentives," *American Economic Journal: Applied Economics* 2 (2010), 205-227.

Gould, Eric, Victor Lavy, and Daniele Paserman, "Immigrating to Opportunity: Estimating the Effect of School Quality Using a Natural Experiment on Ethiopians in Israel," *Quarterly Journal of Economics* 119 (2004), 489-526.

Hastings, Justine, Thomas Kane, and Douglas Staiger, "Preferences and Heterogeneous Treatment Effects in a Public School Choice Lottery," NBER Working Paper No. 12145 (2009).

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

Horowitz, Joel, and Charles Manski, "Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data," *Journal of the American Statistical Association* 95 (2000), 77-84.

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

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

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

Lai, Fang, Elisabeth Sadoulet, and Alain de Janvry "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 (2011), 123-53.

Lee, David, "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects," *Review of Economic Studies* 76 (2009), 1071-1102.

Lee, David S., and David Card, "Regression Discontinuity Inference with Specification Error," *Journal of Econometrics* 142 (2008), 655-674.

McCrary, Justin, "Manipulating the Running Variable in the Regression Discontinuity Design: A Density Test," *American Economic Review* 97 (2008), 318-353.

Newhouse, David, and Kathleen Beegle, "The Effect of School Type on Academic Achievement: Evidence from Indonesia," *Journal of Human Resources* 41 (2006): 529-557.

Pop-Eleches, Cristian, and Miguel Urquiola, "Going to a Better School: Effects and Behavioral Responses," *American Economic Review* 103 (2013), 1289-1324.

Zhang, Hongliang, "The Mirage of Elite Schools: Evidence from Lottery-based School Admissions in China," Mimeo (2013).

Figures

Figure 1: Map of COMIPEMS-participating high schools in the metropolitan Mexico City area

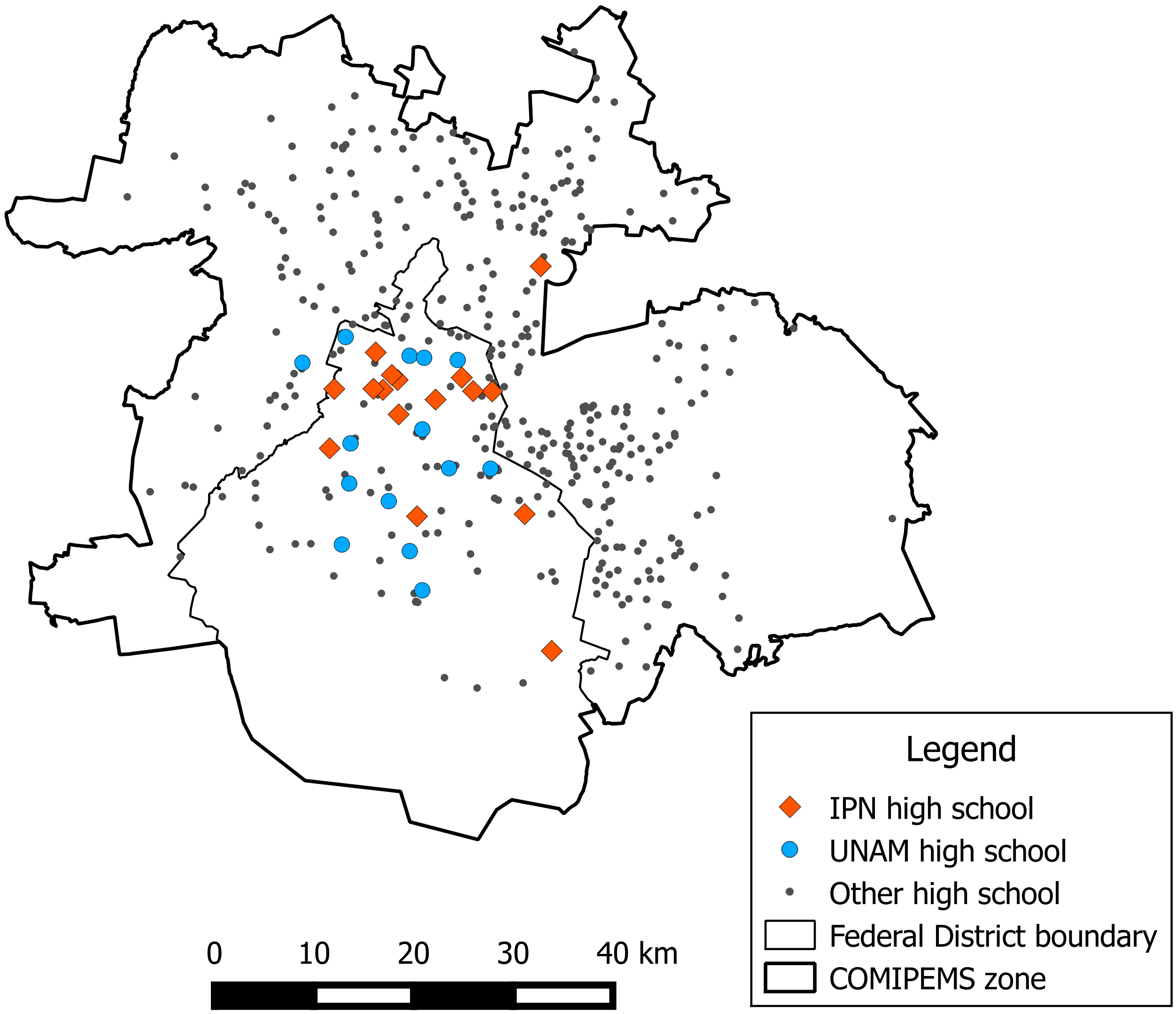
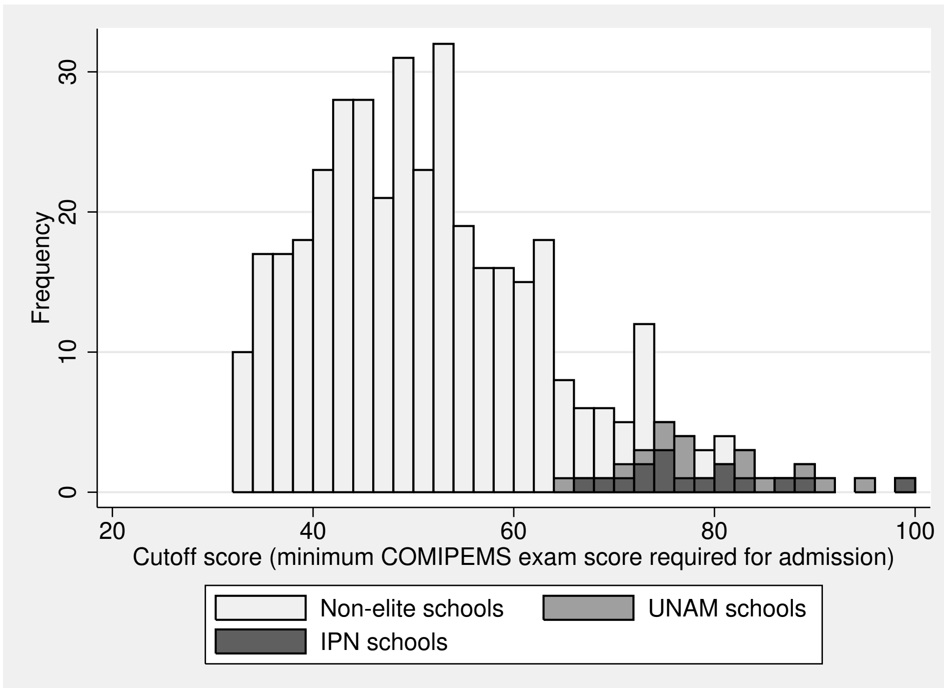
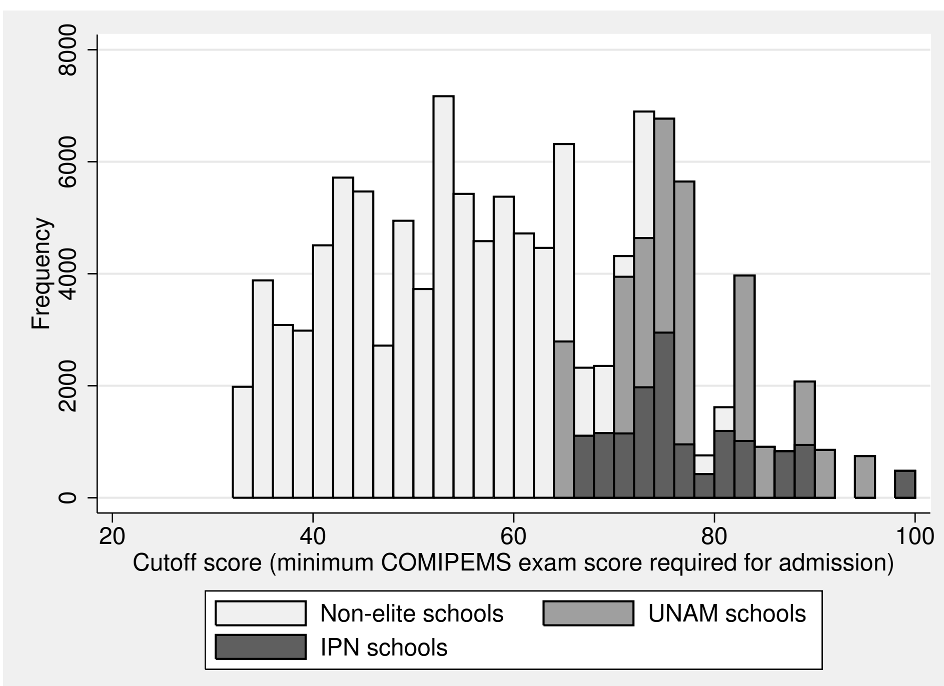


Figure 2: Distribution of admission cutoff scores for oversubscribed schools, 2005 exam year

(a) Unweighted



(b) Weighted by number of assigned students

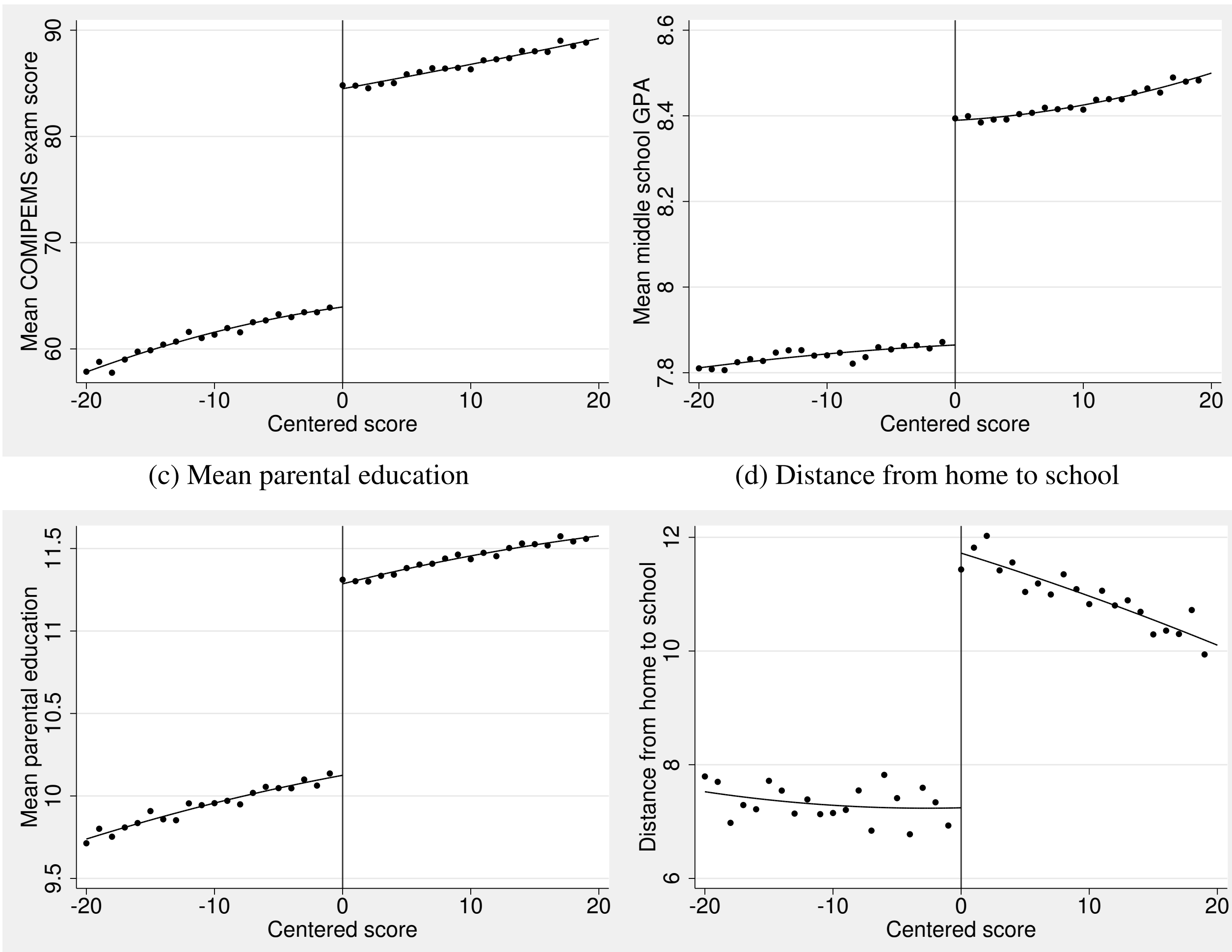


Note. Elite schools are those belonging to the IPN and UNAM subsystems. Panel (b) weights oversubscribed schools by the number of students assigned, so that the mass represents the number of students attending schools with the indicated cutoff score.

Figure 3: Effect of IPN admission on school characteristics experienced by student

(a) Mean COMIPEMS exam score

(b) Mean middle school GPA

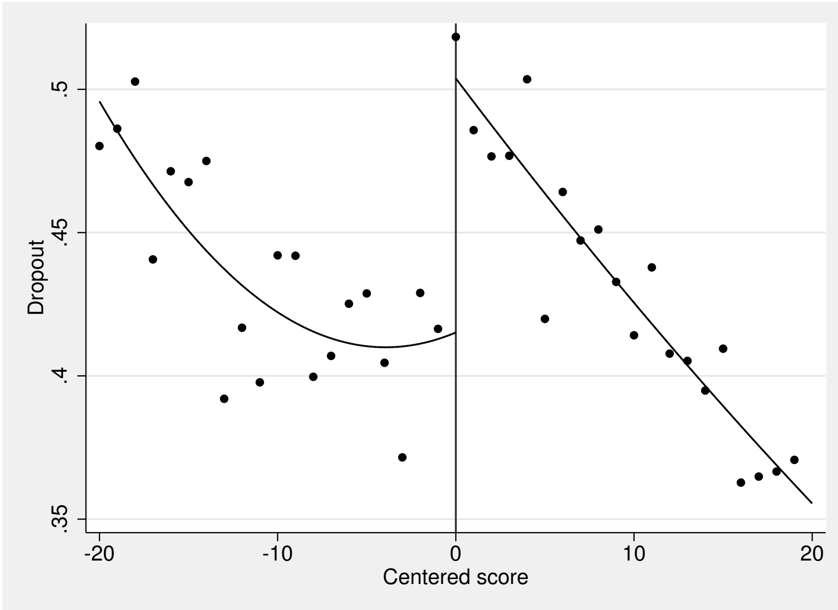


(c) Mean parental education

(d) Distance from home to school

Note. Plots are for students belonging to the regression discontinuity sample defined in the text.

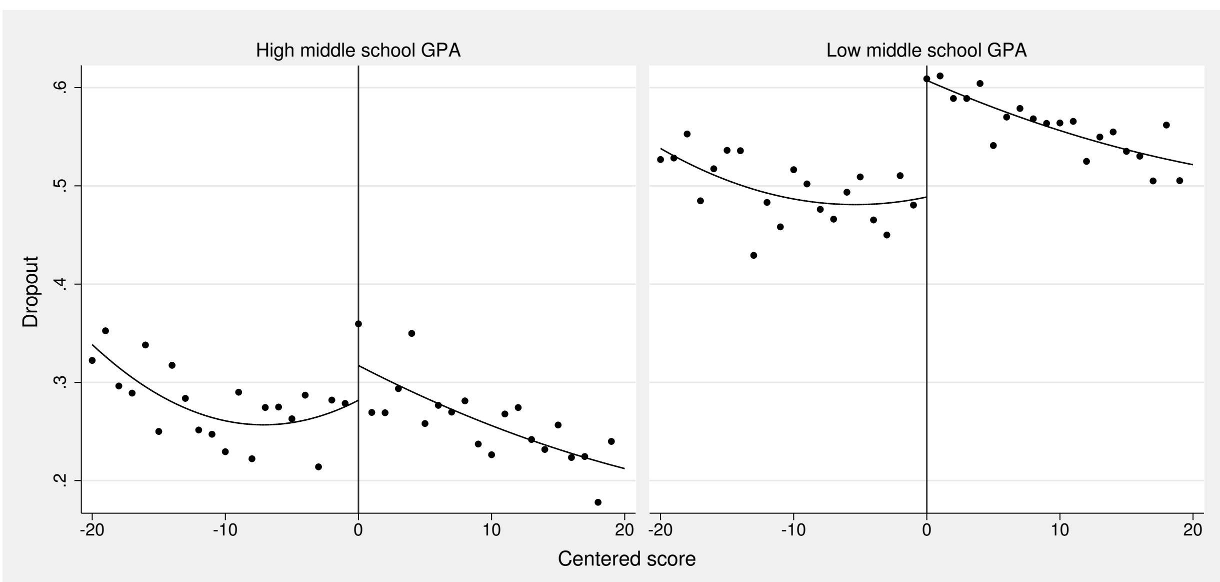
Figure 4: Effect of IPN admission on dropout probability



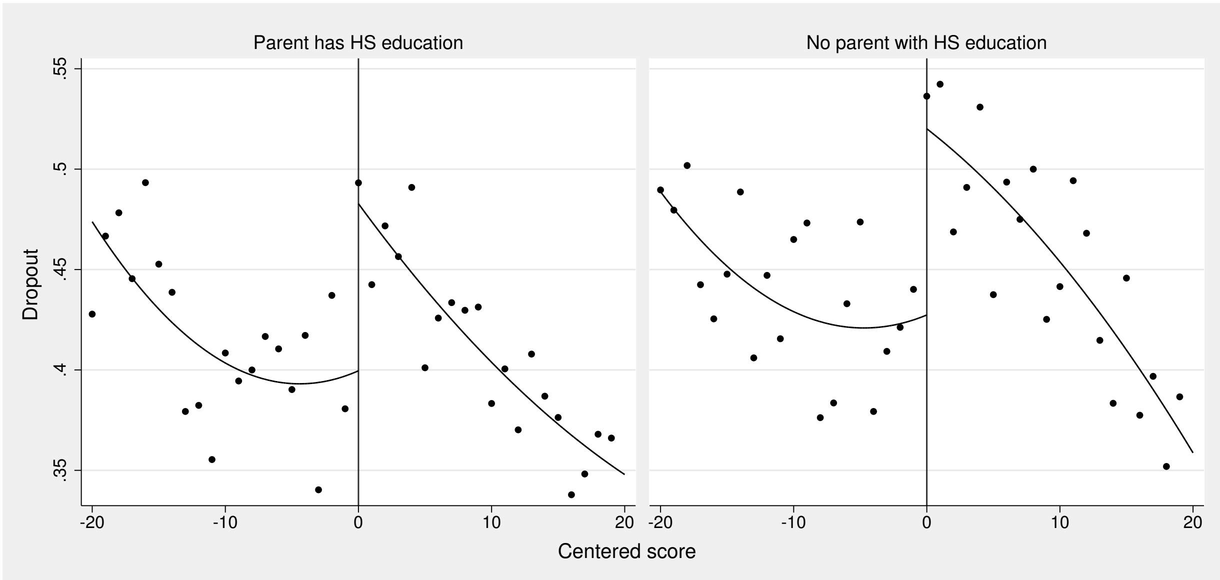
Note. Dropout is defined as not taking the ENLACE exam. Plot is for students belonging to the regression discontinuity sample defined in the text.

Figure 5: Differential effect of IPN admission on dropout

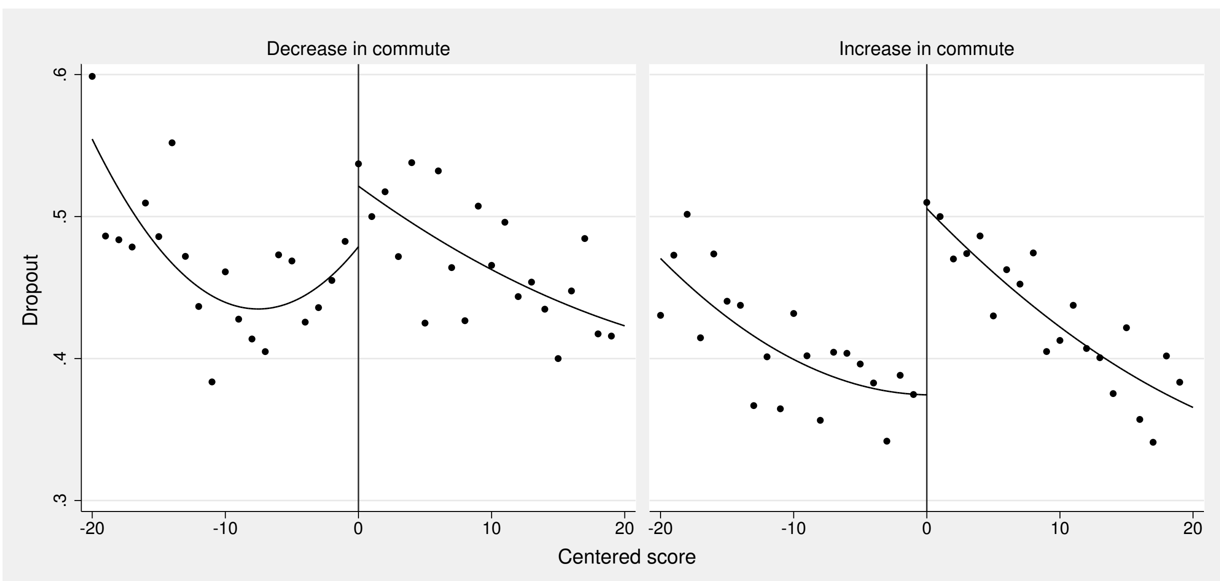
(a) Student middle school GPA



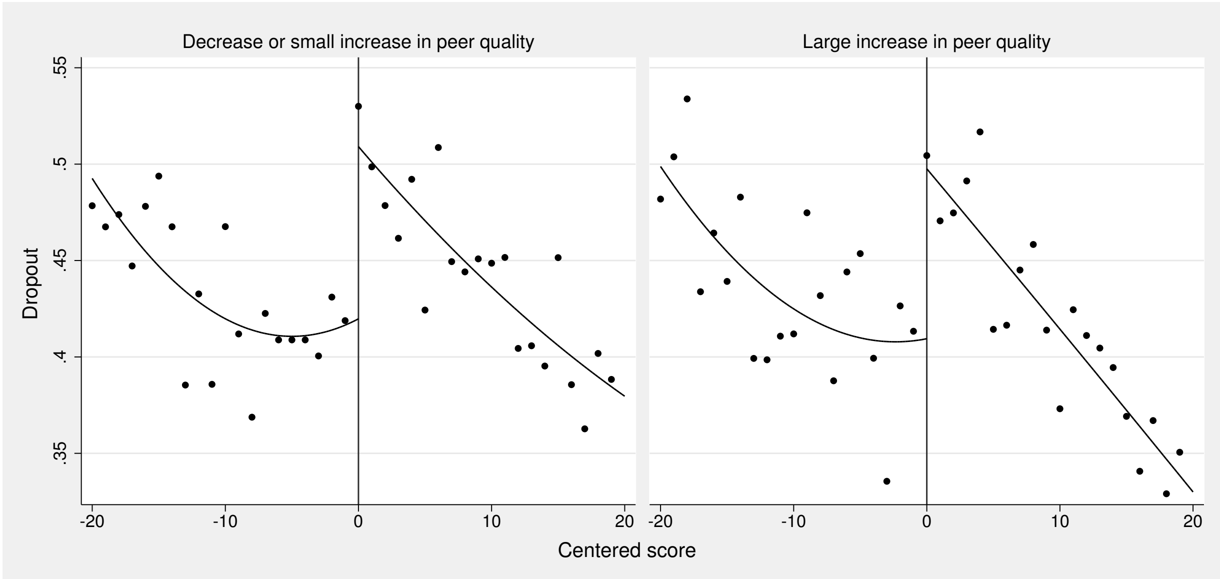
(b) Parental education



(c) Change in commuting distance vs. next-best school

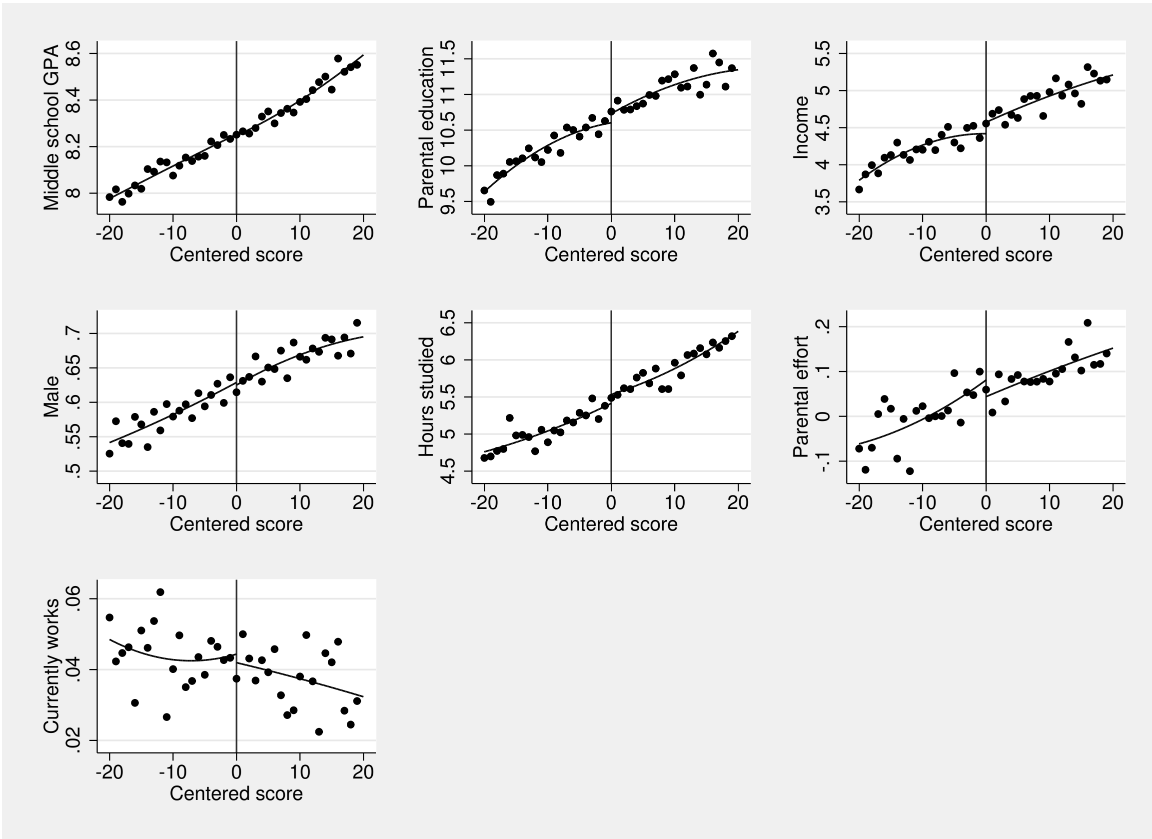


(d) Change in peer quality (mean COMIPEMS score) vs. next-best school



Note. Dropout is defined as not taking the ENLACE exam. Plot is for students belonging to the regression discontinuity sample defined in the text. Low middle school GPA is defined as a GPA below the median among IPN students. Change in commuting distance is the difference in the distance between home and the IPN cutoff school and the school that would be assigned if the student scored one point below the cutoff. This quantity is defined only for students with a such a school on their preference lists. Change in mean COMIPEMS score is the difference in the IPN cutoff school's mean COMIPEMS exam score for that year's assigned students and those of the school that would be assigned if the student scored one point below the cutoff. This quantity is defined only for students with a such a school on their preference lists. A large increase in peer quality is defined as being above the median change in peer quality in the regression discontinuity sample.

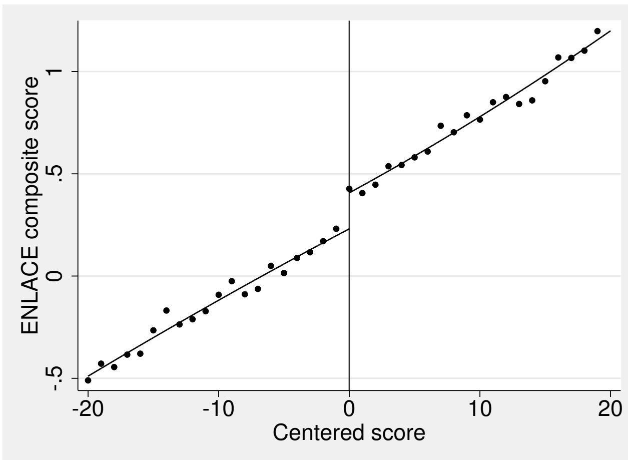
Figure 6: Balance of baseline covariates with respect to IPN admission



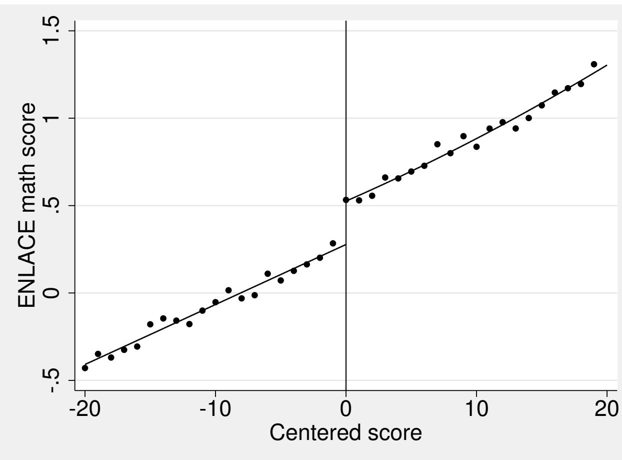
Note. Dependent variables indicated on the vertical axes. Plots are for students belonging to the regression discontinuity sample defined in the text.

Figure 7: Effect of IPN admission on end-of-high school ENLACE exam scores

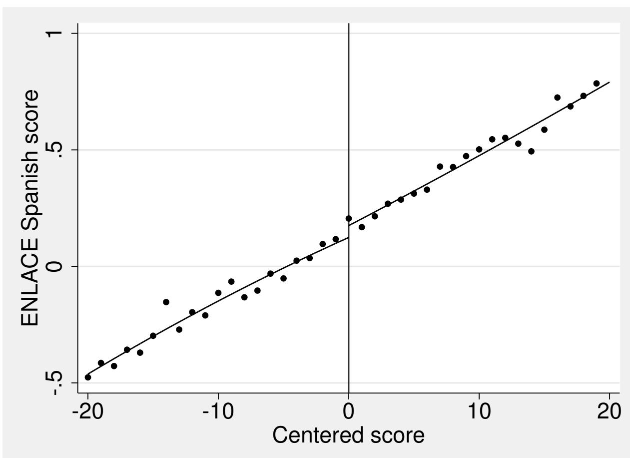
(a) Composite score (math & Spanish)



(b) Math score



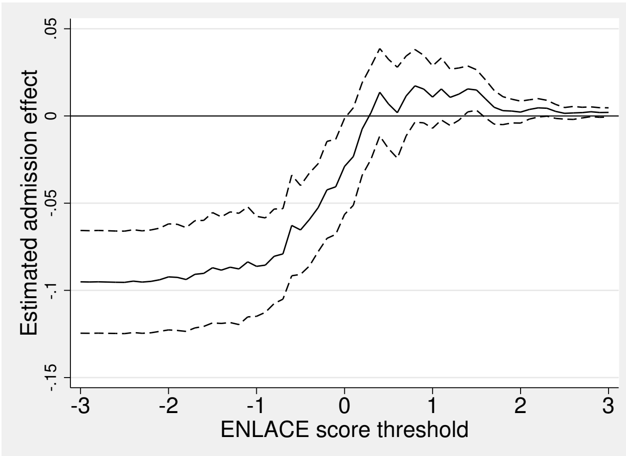
(c) Spanish score



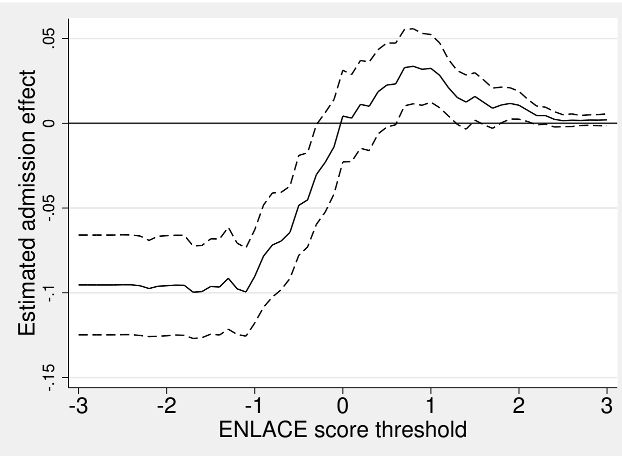
Note. Plot is for students belonging to the regression discontinuity sample defined in the text.

Figure 8: Effect of IPN admission on probability of taking ENLACE and scoring above thresholds

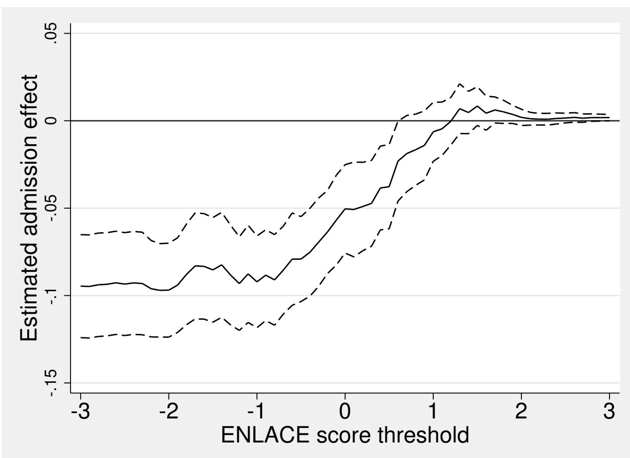
(a) Composite score (math & Spanish)



(b) Math score

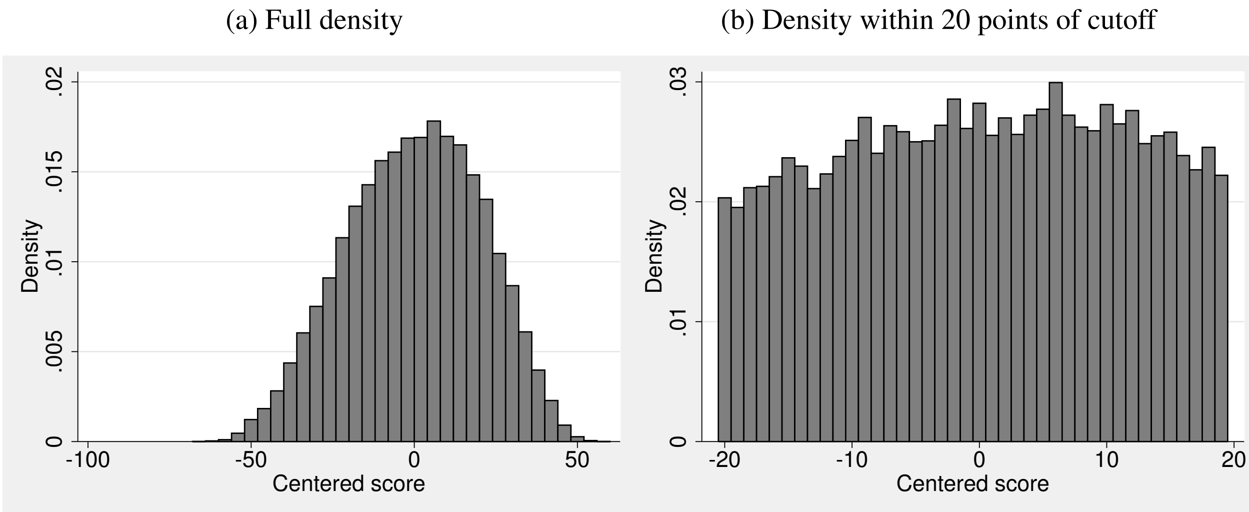


(c) Spanish score



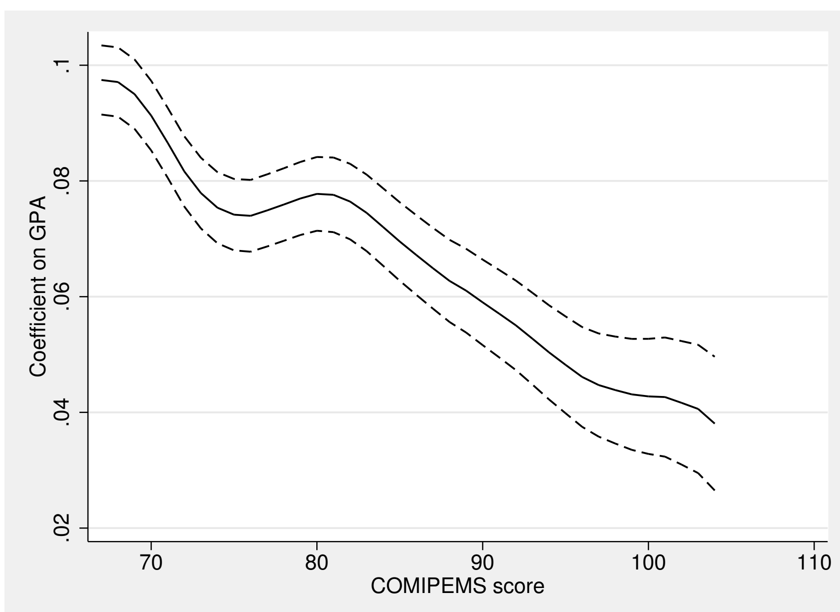
Note. Solid line represents RD estimates of the effect of admission on taking the ENLACE exam and scoring above the score given on the x-axis. Dashed lines give the 95% confidence interval for these estimates.

Figure 9: Density of centered COMIPEMS scores for students in the regression discontinuity sample



Note. Panel (b) is a closer view of the centered score values near the cutoff, presented in order to see more clearly the density of scores close to the cutoff.

Figure 10: Partial correlation of middle school GPA with elite school first-choice preference



Note. Solid line is a smoothed line through the $\hat{\theta}_c$ coefficients from estimating $elite_{imt} = \alpha_{mts} + \beta_c COMIPEMS_i + \theta_c X_i + \varepsilon_{imtc}$, where $elite_{imt}$ is a dummy variable equal to 1 if student i in year t from municipality/delegation m chose an elite school as his first choice, and X_i is middle school GPA. The lines represent the partial correlation between X_i and elite school preference for different COMIPEMS score values. Dashed lines are the 95% confidence intervals for the estimated $\hat{\theta}_c$'s.

Tables

Table 1: Characteristics of students eligible for assignment

	All students	IPN students	UNAM students	Non-elite students	RD sample
	(1)	(2)	(3)	(4)	(5)
Male	0.46	0.65	0.47	0.44	0.62
	(0.50)	(0.48)	(0.50)	(0.50)	(0.49)
Maximum of mother's and father's	10.18	11.38	11.76	9.77	10.71
education	(3.35)	(3.23)	(3.40)	(3.23)	(3.25)
Family income (thousand pesos/month)[a]	4.22	5.21	5.75	3.85	4.62
	(3.35)	(3.64)	(4.08)	(3.07)	(3.38)
Hours studied per week	5.19	6.22	6.58	4.83	5.56
	(3.26)	(3.33)	(3.34)	(3.14)	(3.31)
Index of parental effort[b]	-0.03	0.12	0.18	-0.08	0.04
	(1.00)	(0.95)	(0.97)	(1.00)	(0.97)
Student is employed	0.04	0.03	0.02	0.04	0.04
	(0.19)	(0.18)	(0.15)	(0.20)	(0.20)
Middle school grade point average	8.10	8.54	8.65	7.96	8.29
(of 10)	(0.82)	(0.79)	(0.79)	(0.77)	(0.79)
Distance from assigned school (km)[c]	7.14	10.66	9.40	6.33	9.22
	(6.14)	(7.36)	(6.73)	(5.60)	(7.13)
Number of schools ranked	9.31	9.82	9.45	9.23	9.72
	(3.59)	(3.75)	(3.70)	(3.55)	(3.70)
IPN school as first choice	0.15	0.90	0.03	0.10	1.00
	(0.36)	(0.30)	(0.18)	(0.30)	(0.00)
Number of IPN schools chosen	1.18	4.39	1.24	0.84	3.95
	(1.89)	(2.58)	(1.64)	(1.49)	(2.64)
UNAM school as first choice	0.49	0.10	0.97	0.45	0.00
	(0.50)	(0.30)	(0.18)	(0.50)	(0.00)
Number of UNAM schools chosen	2.53	1.96	4.88	2.20	1.24
	(2.60)	(2.17)	(2.52)	(2.44)	(1.74)
COMIPEMS examination score	63.74	87.96	85.57	57.66	74.63
	(17.95)	(11.06)	(9.90)	(14.29)	(18.49)
Dropped out (did not take ENLACE	0.48	0.38		0.49	0.42
exam; only for non-UNAM students)	(0.50)	(0.49)		(0.50)	(0.49)
ENLACE examination score (for those	-0.03	1.12		-0.18	0.50
who took the exam)[d]	(0.98)	(0.86)		(0.90)	(1.11)
Observations	354581	28551	46265	279765	41075

Note. Standard deviations in parentheses.

[a] Average 2005-2006 exchange rate was 10.9 pesos/dollar.

[b] The parental effort index is constructed by averaging the scores (1-4 ordinal scale) for 13 questions about parental effort and involvement from the survey filled out at the time of COMIPEMS registration. The survey asked “How often do your parents or adults with whom you live do the following activities?” for activities such as “help you with schoolwork” and “attend school events.” The measure is normalized to have mean zero and standard deviation of 1 in the sample of all students.

[c] Distance is calculated as the straight-line distance between the centroid of the student's postal code and the assigned school.

[d] The normalized ENLACE examination score is constructed by subtracting off the year-specific mean score for all examinees in public high schools within the COMIPEMS geographic area and dividing by the year-specific standard deviation from this same sample.

Table 2: Correlates of high school dropout (not taking ENLACE exam)

Dependent variable: dropout (not taking ENLACE exam)*100	(1)	(2)	(3)	(4)	(5)
COMIPEMS score	-0.27***	-0.26***	-0.29***	-0.34***	-0.28***
	(0.056)	(0.040)	(0.015)	(0.047)	(0.020)
Middle school GPA	-17.44***	-17.00***	-17.54***	-17.61***	-18.29***
	(0.588)	(0.509)	(0.260)	(0.609)	(0.248)
Parental education (years)	-0.30***	-0.49***	-0.51***	-0.32***	-0.41***
	(0.072)	(0.034)	(0.037)	(0.075)	(0.043)
Family income (thousand pesos/mo)	0.05	-0.12***	-0.14***	0.04	-0.06
	(0.072)	(0.033)	(0.038)	(0.075)	(0.043)
Male	-0.47	-0.39	-0.27	-0.85**	-0.38
	(0.502)	(0.251)	(0.270)	(0.429)	(0.323)
Hours studied per week	-0.19***	-0.28***	-0.32***	-0.21***	-0.27***
	(0.050)	(0.039)	(0.034)	(0.051)	(0.035)
Parental effort index	-1.11***	-1.01***	-1.00***	-1.10***	-1.16***
	(0.117)	(0.094)	(0.106)	(0.117)	(0.111)
Employed	7.57***	7.49***	7.45***	7.57***	7.58***
	(0.570)	(0.520)	(0.522)	(0.565)	(0.530)
Exam year 2006	3.74***	3.55***	3.63***	3.90***	3.80***
	(0.494)	(0.461)	(0.500)	(0.502)	(0.559)
Distance from home to school (km)			0.27***		0.42***
			(0.023)		(0.033)
IPN school as first choice				-1.38***	-1.62***
				(0.489)	(0.446)
Admitted to IPN school				9.94***	8.26***
				(1.571)	(1.623)
Admitted high school fixed effects	NO	YES	YES	NO	NO
Observations	253506	253506	218870	253506	218870
Adjusted R^2	0.111	0.149	0.148	0.113	0.118
Mean of dependent variable	48.7	48.7	46.9	48.7	46.9

Note. Sample excludes students admitted to an UNAM high school, since these schools do not proctor the ENLACE exam used as the proxy for graduation.

Standard errors, clustered at high school level, in parentheses.

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

Table 3: Regression discontinuity estimates of effect of IPN admission on school characteristics experienced by student

Dependent variable	Mean COMIPEMS score	Mean middle school GPA	Mean parental education (yrs.)	Distance from home to school (km)
	(1)	(2)	(3)	(4)
Score ≥ cutoff	20.319***	0.519***	1.166***	4.485***
	(0.2253)	(0.0056)	(0.0180)	(0.1990)
	[0.00]	[0.00]	[0.00]	[0.00]
Observations	18997	21532	20281	22175
Adjusted R-squared	0.780	0.765	0.627	0.101
Mean of dependent variable	74.565	8.144	10.727	9.305
Bandwidth	13.9	15.9	14.6	17.5

Note. Estimates are from local linear regressions, including separate linear terms for each of the 16 IPN cutoff schools and cutoff school fixed effects. The edge kernel is used in each regression and in computation of the corresponding optimal Imbens-Kalyanaraman bandwidth. Dependent variables in columns 1-3 are mean characteristics of all students admitted to the student's admitted school in his admission year. Huber-White robust standard errors are in parentheses. Wild-cluster bootstrapped p-values, clustered on the discrete COMIPEMS score values, are in brackets.

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

Table 4: Regression discontinuity estimates of effect of IPN admission on dropout

Dependent variable	Dropout (not taking ENLACE exam)			Dropout or late ENLACE (4+ years)		
	(1)	(2)	(3)	(4)	(5)	(6)
Score ≥ cutoff	0.095***	0.092***	0.110***	0.125***	0.120***	0.144***
	(0.0150)	(0.0212)	(0.0110)	(0.0160)	(0.0225)	(0.0116)
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Observations	20281	11122	35475	17748	9783	32658
Adjusted R-squared	0.013	0.017	0.015	0.021	0.026	0.021
Mean of dependent variable	0.436	0.443	0.424	0.528	0.539	0.514
Bandwidth	15.3	7.6	30.6	13.5	6.7	26.9

Note. Estimates are from local linear regressions, including separate linear terms for each of the 16 IPN cutoff schools and cutoff school fixed effects. The edge kernel is used in each regression and in computation of the corresponding optimal Imbens-Kalyanaraman bandwidth. Huber-White robust standard errors are in parentheses. Wild-cluster bootstrapped p-values, clustered on the discrete COMIPEMS score values, are in brackets.

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

Table 5: Regression discontinuity estimates of heterogeneous effects of IPN admission on dropout

Dependent variable: dropout (not taking ENLACE exam)	(1)	(2)	(3)	(4)
Score ≥ cutoff	0.052**	0.086***	0.058*	0.087***
	(0.0210)	(0.0208)	(0.0341)	(0.0234)
	[0.05]	[0.00]	[0.02]	[0.00]
(Score ≥ cutoff) * (Low middle school GPA)	0.069**			
	(0.0286)			
	[0.01]			
(Score ≥ cutoff) * (No parent w/HS degree)		0.020		
		(0.0315)		
		[0.59]		
(Score ≥ cutoff) * (Admission increases commute)			0.070*	
			(0.0403)	
			[0.00]	
(Score ≥ cutoff) * (Δ mean HS peer COMIPEMS exam score > median)				0.022
				(0.0292)
				[0.19]
Observations	21216	19009	13880	23735
Adjusted R-squared	0.080	0.015	0.020	0.015
Mean of dependent variable	0.416	0.432	0.434	0.429

Note. Estimates are from local linear regressions, including separate linear terms for each of the 16 IPN cutoff schools and cutoff school fixed effects. Linear terms terms and fixed effects are interacted with the corresponding covariate in each column, so that point estimates are equivalent to separately estimating the regression for each value of the covariate. The edge kernel is used in each regression and in computation of the corresponding optimal Imbens-Kalyanaraman bandwidths, which are computed separately for each value of the covariate. Huber-White robust standard errors are in parentheses. Wild-cluster bootstrapped p-values, clustered on the discrete COMIPEMS score values, are in brackets.

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

Table 6: Tests for balance of baseline covariates with respect to IPN assignment

Panel (a) At time of assignment

Dependent variable	Middle school GPA	Parental education	Family income (thousand pesos/mo)	Male	Hours studied per week	Parental effort index	Employed
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Score ≥ cutoff	-0.019	0.099	0.106	-0.001	0.097	-0.034	-0.000
	(0.0188)	(0.1025)	(0.1006)	(0.0127)	(0.1015)	(0.0322)	(0.0061)
	[0.10]	[0.27]	[0.25]	[0.90]	[0.29]	[0.24]	[0.92]
Observations	27136	18414	18188	25007	19519	17351	21007
Adjusted R-squared	0.068	0.016	0.010	0.096	0.012	0.003	0.001
Mean of dependent variable	8.26	10.70	4.57	0.63	5.48	0.05	0.04
S.D. of dependent variable	0.76	3.17	3.26	0.48	3.27	0.97	0.20
Bandwidth	21.2	15.2	15.2	18.8	15.9	13.7	17.6
p-value, joint significance of all admission coefficients	0.58						

Panel (b) After dropout

Dependent variable	Middle school GPA	Parental education	Family income (thousand pesos/mo)	Male	Hours studied per week	Parental effort index	Employed
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Score ≥ cutoff	0.072***	0.192*	0.123	-0.010	0.272**	0.045	-0.001
	(0.0228)	(0.1103)	(0.1206)	(0.0149)	(0.1115)	(0.0392)	(0.0064)
	[0.00]	[0.16]	[0.38]	[0.44]	[0.04]	[0.19]	[0.82]
Observations	17752	15782	12815	18289	16242	11160	16253
Adjusted R-squared	0.103	0.027	0.014	0.102	0.022	0.004	-0.000
Mean of dependent variable	8.48	10.84	4.62	0.59	5.74	0.11	0.03
S.D. of dependent variable	0.75	3.19	3.27	0.49	3.32	0.96	0.18
Bandwidth	25.1	23.6	18.9	26.2	24.7	16.4	26.1
p-value, joint significance of all admission coefficients	0.01						

Note. Estimates are from local linear regressions of the specified order, including separate linear terms for each of the 16 IPN cutoff schools and cutoff school fixed effects. The edge kernel is used in each regression and in computation of the corresponding optimal Imbens-Kalyanaraman bandwidths. The p-values for joint significance are from chi-square tests that the admission coefficients are all equal to zero, estimated using seemingly unrelated regression. “At time of assignment" refers to all students in the RD sample, while “after dropout" is restricted to students who took the ENLACE exam. Huber-White robust standard errors are in parentheses. Wild-cluster bootstrapped p-values, clustered on the discrete COMIPEMS score values, are in brackets.

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

Table 7: Regression discontinuity estimates of effect of IPN admission on ENLACE score

Panel (a) Composite score (math & Spanish)

	(1)	(2)	(3)
Score ≥ cutoff	0.168***	0.160***	0.171***
	(0.0290)	(0.0409)	(0.0210)
	[0.00]	[0.00]	[0.00]
Observations	10685	5413	19141
Adjusted R-squared	0.269	0.209	0.402
Mean of dependent variable	0.322	0.292	0.409
Bandwidth	13.9	6.9	27.8

Panel (b) Math score

	(1)	(2)	(3)
Score ≥ cutoff	0.245***	0.223***	0.251***
	(0.0297)	(0.0420)	(0.0217)
	[0.00]	[0.00]	[0.00]
Observations	12115	6183	20386
Adjusted R-squared	0.250	0.183	0.397
Mean of dependent variable	0.409	0.378	0.516
Bandwidth	15.6	7.8	31.3

Panel (c) Spanish score

	(1)	(2)	(3)
Score ≥ cutoff	0.049	0.052	0.039
	(0.0339)	(0.0480)	(0.0245)
	[0.03]	[0.27]	[0.00]
Observations	10693	5417	19155
Adjusted R-squared	0.154	0.129	0.237
Mean of dependent variable	0.158	0.131	0.217
Bandwidth	14.1	7.1	28.2

Note. Estimates are from local linear regressions of the specified order, including separate linear terms for each of the 16 IPN cutoff schools and cutoff school fixed effects. The edge kernel is used in each regression and in computation of the corresponding optimal Imbens-Kalyanaraman bandwidth. Huber-White robust standard errors are in parentheses. Wild-cluster bootstrapped p-values, clustered on the discrete COMIPEMS score values, are in brackets.

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

Table 8: Regression discontinuity estimates of heterogeneous effects of IPN admission on ENLACE score

Dependent variable: Composite score (math & Spanish)	(1)	(2)	(3)	(4)
Score ≥ cutoff	0.128***	0.163***	0.285***	0.146***
	(0.0359)	(0.0357)	(0.0629)	(0.0310)
	[0.00]	[0.00]	[0.00]	[0.00]
(Score ≥ cutoff) * (Low middle school GPA)	0.080			
	(0.0509)			
	[0.10]			
(Score ≥ cutoff) * (No parent w/HS degree)		0.019		
		(0.0565)		
		[0.80]		
(Score ≥ cutoff) * (Admission increases commute)			-0.170**	
			(0.0760)	
			[0.22]	
(Score ≥ cutoff) * (Δ mean HS peer COMIPEMS exam score > median)				0.054
				(0.0502)
				[0.34]
Observations	13767	11790	7155	14636
Adjusted R-squared	0.322	0.302	0.258	0.339
Mean of dependent variable	0.367	0.360	0.325	0.350

Note. Estimates are from local linear regressions of the specified order, including separate linear terms for each of the 16 IPN cutoff schools and cutoff school fixed effects. Linear terms and fixed effects are interacted with the corresponding covariate in each column, so that point estimates are equivalent to separately estimating the regression for each value of the covariate. The edge kernel is used in each regression and in computation of the corresponding optimal Imbens-Kalyanaraman bandwidths, which are computed separately for each value of the covariate. Huber-White robust standard errors are in parentheses. Wild-cluster bootstrapped p-values, clustered on the discrete COMIPEMS score values, are in brackets.

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

Table 9: Regression discontinuity estimates of effect of admission to a higher-cutoff IPN school

Dependent variable	Mean COMIPEMS score	Mean middle school GPA	Mean parental education (yrs.)	Distance from home to school (km)	Dropout (not taking ENLACE exam)	ENLACE composite score (math & Spanish)	ENLACE math score	ENLACE Spanish score
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Score ≥ cutoff	4.747***	0.115***	0.305***	-2.267***	0.020	0.075**	0.074**	0.058*
	(0.2109)	(0.0058)	(0.0182)	(0.3300)	(0.0188)	(0.0293)	(0.0329)	(0.0333)
	[0.00]	[0.00]	[0.00]	[0.01]	[0.24]	[0.07]	[0.06]	[0.18]
Observations	7391	7391	7391	9820	13215	10072	9237	10084
Adjusted R-squared	0.740	0.749	0.684	0.050	0.013	0.352	0.314	0.204
Mean of dependent variable	86.090	8.419	11.481	11.609	0.425	0.676	0.776	0.402
Bandwidth	4.7	4.7	4.8	7.5	8.7	12.1	11.5	12.1

Note. Estimates are from local linear regressions, including separate linear terms for each of the 15 IPN schools (excludes lowest-cutoff IPN school) and cutoff school fixed effects. The edge kernel is used in each regression and in computation of the corresponding optimal Imbens-Kalyanaraman bandwidth. Dependent variables in columns 1-3 are mean characteristics of all students admitted to the student's admitted school in his admission year. Standard errors accounting for clustering at the student level are in parentheses. Wild-cluster bootstrapped p-values, clustered on the discrete COMIPEMS score values, are in brackets.

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

Appendix A. Comparison of dropout and ENLACE score results with Estrada and Gignoux (2014)

In their paper estimating the effect of IPN admission on the expected returns to higher education, Estrada and Gignoux (2014) also present basic results on dropout and ENLACE scores. They estimate smaller effects of admission on dropout than us, as well as larger effects on ENLACE math scores. We show in Table A1 that most of the difference between our results is due to the different samples used. Panel (a), column 1 reproduces EG's dropout result using their sample description, a five-point bandwidth, and rectangular kernel, as in their paper. The sample size (3,184 vs. 3,206) and estimated effect on dropout (0.031 vs. 0.036) are nearly identical between their results and our replication. Dropping private middle schools from their sample increases the point estimate to 5.6 percentage points. Adding State of Mexico students (column 3) increases the point estimate on admission further. Adding the 2006 COMIPEMS exam and the 2010 ENLACE results to the sample (column 4), the estimated effect declines to 7.6 percentage points. Excluding students who, given their stated preferences, could be assigned to an UNAM school for some point values higher than the IPN admission cutoff score (column 5), the point estimate increases slightly to 8.5 percentage points, which is close to the result we obtain from using nonparametric regression in the body of this paper.

Our replication of Estrada and Gignoux's (2014) ENLACE math score effects are in column 1 of Panel (b). The sample size in our replication is larger than theirs (1,570 vs. 1,115) because they limit the sample to students who were included in the random sample for a supplementary survey given to ENLACE-takers. The point estimates are almost identical to each other (0.34 standard deviations), but they are significantly higher than the result found in this paper. This estimate declines to 0.33 when we exclude private school students, falls further to 0.29 when adding State of Mexico students, and decreases to 0.22 when adding the 2006 COMIPEMS and 2010 ENLACE data. The Spanish score effects in Panel (c) are statistically insignificant for all samples.

Table A1: Dropout and ENLACE regression discontinuity results for different sample selection criteria

Panel (a) Dropout (not taking ENLACE exam)

	Estrada & Gignoux sample	Delete private middle schools	Add State of Mexico middle schools	Add 2006 COMIPEMS, 2010 ENLACE	Selection method used in this paper
	(1)	(2)	(3)	(4)	(5)
Score ≥ cutoff	0.031	0.056*	0.090***	0.076***	0.085***
	(0.0316)	(0.0327)	(0.0280)	(0.0215)	(0.0236)
Observations	3184	2928	5125	10990	6914
Adjusted R-squared	0.003	0.004	0.012	0.015	0.018
Mean of dependent variable	0.506	0.508	0.476	0.455	0.452

Panel (b) Math ENLACE score

	Estrada & Gignoux sample	Delete private middle schools	Add State of Mexico middle schools	Add 2006 COMIPEMS, 2010 ENLACE	Selection method used in this paper
Dependent variable: Math score	(1)	(2)	(3)	(4)	(5)
Score ≥ cutoff	0.340***	0.331***	0.289***	0.215***	0.205***
	(0.0822)	(0.0853)	(0.0524)	(0.0414)	(0.0502)
Observations	1570	1439	2678	5978	3781
Adjusted R-squared	0.097	0.101	0.140	0.132	0.180
Mean of dependent variable	0.341	0.338	0.401	0.387	0.365

Panel (c) Spanish ENLACE score

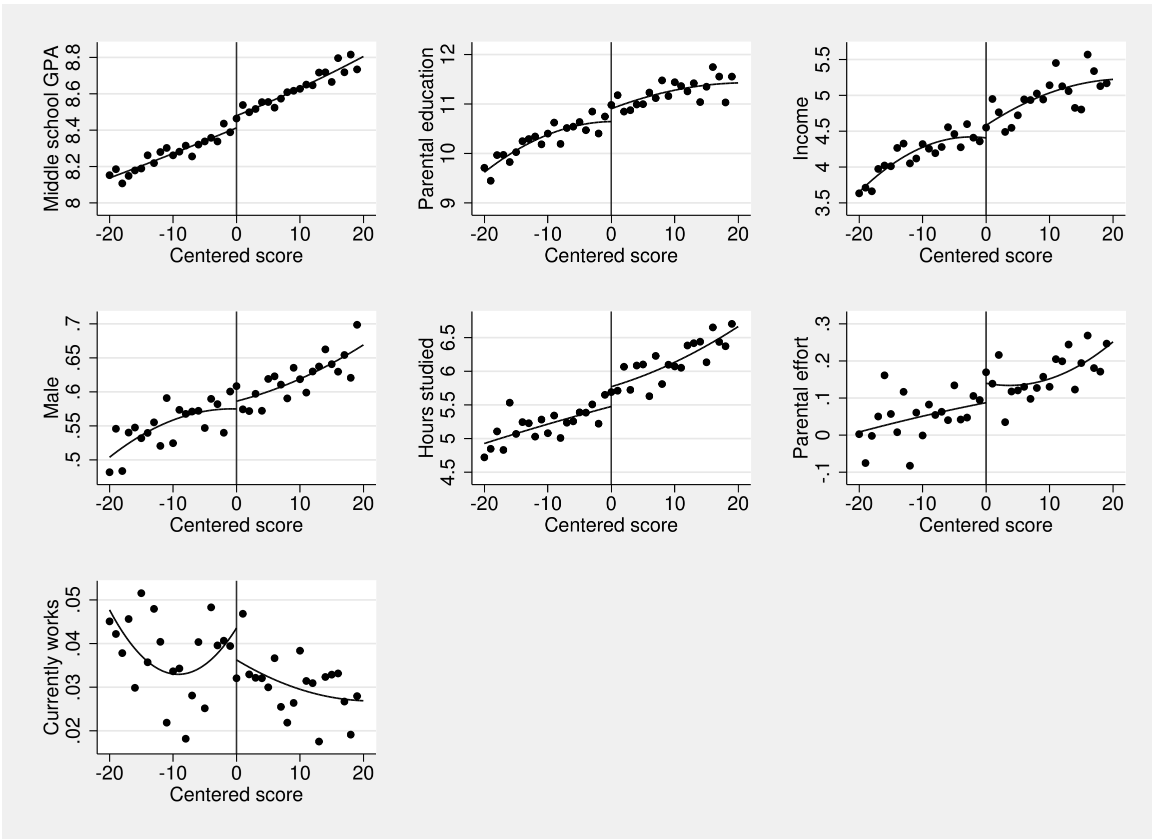
	Estrada & Gignoux sample	Delete private middle schools	Add State of Mexico middle schools	Add 2006 COMIPEMS, 2010 ENLACE	Selection method used in this paper
Dependent variable: Spanish score	(1)	(2)	(3)	(4)	(5)
Score ≥ cutoff	0.072	0.048	-0.017	-0.016	0.043
	(0.0742)	(0.0775)	(0.0616)	(0.0432)	(0.0484)
Observations	1570	1439	2681	5981	3784
Adjusted R-squared	0.044	0.042	0.070	0.075	0.124
Mean of dependent variable	0.138	0.133	0.188	0.182	0.131

Note. Estimates are from local linear regressions of the specified order, including separate linear terms for each of the 16 IPN cutoff schools and cutoff school fixed effects. The rectangular kernel is used in each regression and the bandwidth is fixed to 5 in order to compare to the sample selection in Estrada and Gignoux (2014). The header in columns 2-4 explain the changes made to the sample from the previous column. Standard errors clustered at the assigned school level are in parentheses.

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

Appendix B. Additional figure and table

Figure B1: Balance of covariates with respect to IPN admission, after dropout



Note. Dependent variables indicated on the vertical axes. Plots are for students belonging to the regression discontinuity sample defined in the text.

Table B1: Regression discontinuity estimates of heterogeneous effects of IPN admission on ENLACE subscores

Panel (a) Math score				
	(1)	(2)	(3)	(4)
Score ≥ cutoff	0.222***	0.206***	0.356***	0.223***
	(0.0378)	(0.0396)	(0.0687)	(0.0346)
	[0.00]	[0.00]	[0.00]	[0.00]
(Score ≥ cutoff) * (Low middle school GPA)	0.033			
	(0.0505)			
	[0.36]			
(Score ≥ cutoff) * (No parent w/HS degree)		0.096		
		(0.0604)		
		[0.24]		
(Score ≥ cutoff) * (Admission increases commute)			-0.170**	
			(0.0818)	
			[0.23]	
(Score ≥ cutoff) * (Δ mean HS peer COMIPEMS exam score > median)				0.045
				(0.0553)
				[0.45]
Observations	16566	11779	7553	14037
Adjusted R-squared	0.322	0.266	0.235	0.291
Mean of dependent variable	0.442	0.429	0.403	0.427

Panel (b) Spanish score				
	(1)	(2)	(3)	(4)
Score ≥ cutoff	-0.005	0.077*	0.124**	0.010
	(0.0415)	(0.0432)	(0.0627)	(0.0329)
	[0.86]	[0.02]	[0.02]	[0.62]
(Score ≥ cutoff) * (Low middle school GPA)	0.109*			
	(0.0626)			
	[0.09]			
(Score ≥ cutoff) * (No parent w/HS degree)		-0.055		
		(0.0655)		
		[0.51]		
(Score ≥ cutoff) * (Admission increases commute)			-0.105	
			(0.0791)	
			[0.26]	
(Score ≥ cutoff) * (Δ mean HS peer COMIPEMS exam score > median)				0.068
				(0.0546)
				[0.18]
Observations	12680	11789	8396	16798
Adjusted R-squared	0.180	0.167	0.163	0.224
Mean of dependent variable	0.209	0.176	0.163	0.194

Note. Estimates are from local linear regressions of the specified order, including separate linear terms for each of the 16 IPN cutoff schools and cutoff school fixed effects. Linear terms and fixed effects are interacted with the corresponding covariate in each column, so that point estimates are equivalent to separately estimating the regression for each value of the covariate. The edge kernel is used in each regression and in computation of the corresponding optimal Imbens-Kalyanaraman bandwidths, which are computed separately for each value of the covariate. Huber-White robust standard errors are in parentheses. Wild-cluster bootstrapped p-values, clustered on the discrete COMIPEMS score values, are in brackets.

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

Appendix C. Local quadratic regression results

Table C1: Regression discontinuity estimates of effect of IPN admission on school characteristics experienced by student

Dependent variable	Mean COMIPEMS score	Mean middle school GPA	Mean parental education (yrs.)	Distance from home to school (km)
	(1)	(2)	(3)	(4)
Score ≥ cutoff	20.256***	0.517***	1.169***	4.561***
	(0.2302)	(0.0083)	(0.0223)	(0.2780)
	[0.00]	[0.00]	[0.00]	[0.00]
Observations	34860	23856	28233	23246
Adjusted R-squared	0.825	0.771	0.652	0.101
Mean of dependent variable	74.423	8.145	10.726	9.310
Bandwidth	30.1	17.7	22.3	19.1

Note. Estimates are from local quadratic regressions, including separate quadratic terms for each of the 16 IPN cutoff schools and cutoff school fixed effects. The edge kernel is used in each regression and in computation of the corresponding optimal Imbens-Kalyanaraman bandwidth. Dependent variables in columns 1-3 are mean characteristics of all students admitted to the student's admitted school in his admission year. Huber-White robust standard errors are in parentheses. Wild-cluster bootstrapped p-values, clustered on the discrete COMIPEMS score values, are in brackets.

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

Table C2: Regression discontinuity estimates of effect of IPN admission on dropout

Dependent variable	Dropout (not taking ENLACE exam)			Dropout or late ENLACE (4+ years)		
	(1)	(2)	(3)	(4)	(5)	(6)
Score ≥ cutoff	0.087***	0.078**	0.095***	0.116***	0.115***	0.130***
	(0.0215)	(0.0326)	(0.0155)	(0.0216)	(0.0329)	(0.0156)
	[0.00]	[0.11]	[0.00]	[0.00]	[0.16]	[0.00]
Observations	21532	11122	36641	21532	11122	36065
Adjusted R-squared	0.013	0.016	0.016	0.020	0.024	0.023
Mean of dependent variable	0.436	0.443	0.422	0.527	0.537	0.509
Bandwidth	16.4	8.2	32.9	16.1	8.1	32.2

Note. Estimates are from local quadratic regressions, including separate quadratic terms for each of the 16 IPN cutoff schools and cutoff school fixed effects. The edge kernel is used in each regression and in computation of the corresponding optimal Imbens-Kalyanaraman bandwidth. Huber-White robust standard errors are in parentheses. Wild-cluster bootstrapped p-values, clustered on the discrete COMIPEMS score values, are in brackets.

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

Table C3: Regression discontinuity estimates of heterogeneous effects of IPN admission on dropout

Dependent variable: dropout (not taking ENLACE exam)	(1)	(2)	(3)	(4)
Score ≥ cutoff	0.042	0.078***	0.049	0.075***
	(0.0322)	(0.0268)	(0.0374)	(0.0288)
	[0.37]	[0.10]	[0.06]	[0.01]
(Score ≥ cutoff) * (Low middle school GPA)	0.073*			
	(0.0395)			
	[0.09]			
(Score ≥ cutoff) * (No parent w/HS degree)		0.020		
		(0.0406)		
		[0.62]		
(Score ≥ cutoff) * (Admission increases commute)			0.075	
			(0.0469)	
			[0.01]	
(Score ≥ cutoff) * (Δ mean HS peer COMIPEMS exam score > median)				0.021
				(0.0401)
				[0.36]
Observations	27093	24748	18823	24936
Adjusted R-squared	0.067	0.015	0.019	0.014
Mean of dependent variable	0.444	0.426	0.439	0.433

Note. Estimates are from local quadratic regressions, including separate quadratic terms for each of the 16 IPN cutoff schools and cutoff school fixed effects. Quadratic terms terms and fixed effects are interacted with the corresponding covariate in each column, so that point estimates are equivalent to separately estimating the regression for each value of the covariate. The edge kernel is used in each regression and in computation of the corresponding optimal Imbens-Kalyanaraman bandwidths, which are computed separately for each value of the covariate. Huber-White robust standard errors are in parentheses. Wild-cluster bootstrapped p-values, clustered on the discrete COMIPEMS score values, are in brackets.

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

Table C4: Regression discontinuity estimates of effect of IPN admission on ENLACE score

Panel (a) Composite score (math & Spanish)			
	(1)	(2)	(3)
Score ≥ cutoff	0.153***	0.153***	0.181***
	(0.0396)	(0.0590)	(0.0287)
	[0.00]	[0.05]	[0.00]
Observations	12823	6183	21110
Adjusted R-squared	0.291	0.219	0.450
Mean of dependent variable	0.333	0.297	0.449
Bandwidth	16.6	8.3	33.2

Panel (b) Math score			
	(1)	(2)	(3)
Score ≥ cutoff	0.221***	0.200***	0.242***
	(0.0420)	(0.0625)	(0.0304)
	[0.00]	[0.03]	[0.00]
Observations	12823	6951	21679
Adjusted R-squared	0.266	0.192	0.426
Mean of dependent variable	0.414	0.380	0.540
Bandwidth	17.4	8.7	34.8

Panel (c) Spanish score			
	(1)	(2)	(3)
Score ≥ cutoff	0.050	0.057	0.069**
	(0.0447)	(0.0658)	(0.0326)
	[0.20]	[0.29]	[0.00]
Observations	13483	6955	21937
Adjusted R-squared	0.172	0.134	0.290
Mean of dependent variable	0.165	0.139	0.260
Bandwidth	18.1	9.0	36.2

Note. Estimates are from local quadratic regressions of the specified order, including separate quadratic terms for each of the 16 IPN cutoff schools and cutoff school fixed effects. The edge kernel is used in each regression and in computation of the corresponding optimal Imbens-Kalyanaraman bandwidth. Huber-White robust standard errors are in parentheses. Wild-cluster bootstrapped p-values, clustered on the discrete COMIPEMS score values, are in brackets.

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

Table C5: Regression discontinuity estimates of heterogeneous effects of IPN admission on ENLACE score

Dependent variable: Composite score (math & Spanish)	(1)	(2)	(3)	(4)
Score ≥ cutoff	0.051	0.157***	0.309***	0.122**
	(0.0669)	(0.0452)	(0.0810)	(0.0560)
	[0.35]	[0.00]	[0.00]	[0.02]
(Score ≥ cutoff) * (Low middle school GPA)	0.171**			
	(0.0832)			
	[0.01]			
(Score ≥ cutoff) * (No parent w/HS degree)		-0.018		
		(0.0783)		
		[0.88]		
(Score ≥ cutoff) * (Admission increases commute)			-0.181**	
			(0.0907)	
			[0.22]	
(Score ≥ cutoff) * (Δ mean HS peer COMIPEMS exam score > median)				0.068
				(0.0769)
				[0.23]
Observations	12442	14177	13383	13443
Adjusted R-squared	0.297	0.362	0.389	0.305
Mean of dependent variable	0.294	0.433	0.367	0.343

Note. Estimates are from local quadratic regressions of the specified order, including separate quadratic terms for each of the 16 IPN cutoff schools and cutoff school fixed effects. Quadratic terms and fixed effects are interacted with the corresponding covariate in each column, so that point estimates are equivalent to separately estimating the regression for each value of the covariate. The edge kernel is used in each regression and in computation of the corresponding optimal Imbens-Kalyanaraman bandwidths, which are computed separately for each value of the covariate. Huber-White robust standard errors are in parentheses. Wild-cluster bootstrapped p-values, clustered on the discrete COMIPEMS score values, are in brackets.

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

Table C6: Regression discontinuity estimates of effect of admission to a higher-cutoff IPN school

Dependent variable	Mean COMIPEMS score	Mean middle school GPA	Mean parental education (yrs.)	Distance from home to school (km)	Dropout (not taking ENLACE exam)	ENLACE composite score (math & Spanish)	ENLACE math score	ENLACE Spanish score
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Score ≥ cutoff	4.966***	0.128***	0.311***	-2.210***	0.026*	0.055**	0.073**	0.023
	(0.1780)	(0.0054)	(0.0142)	(0.2690)	(0.0150)	(0.0262)	(0.0298)	(0.0273)
	[0.00]	[0.00]	[0.00]	[0.00]	[0.08]	[0.11]	[0.05]	[0.59]
Observations	13215	9458	14598	16374	22273	13843	13095	17751
Adjusted R-squared	0.765	0.763	0.681	0.054	0.015	0.392	0.344	0.261
Mean of dependent variable	84.572	8.376	11.385	11.196	0.410	0.687	0.782	0.435
Bandwidth	8.7	6.8	9.8	12.0	15.7	17.4	15.5	22.9

Note. Estimates are from local quadratic regressions, including separate quadratic terms for each of the 15 IPN schools (excludes lowest-cutoff IPN school) and cutoff school fixed effects. The edge kernel is used in each regression and in computation of the corresponding optimal Imbens-Kalyanaraman bandwidth. Dependent variables in columns 1-3 are mean characteristics of all students admitted to the student's admitted school in his admission year. Standard errors accounting for clustering at the student level are in parentheses. Wild-cluster bootstrapped p-values, clustered on the discrete COMIPEMS score values, are in brackets.

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

Table C7: Regression discontinuity estimates of heterogeneous effects of IPN admission on ENLACE subscores

Panel (a) Math score				
	(1)	(2)	(3)	(4)
Score ≥ cutoff	0.135*	0.189***	0.363***	0.179***
	(0.0698)	(0.0527)	(0.0745)	(0.0577)
	[0.11]	[0.00]	[0.00]	[0.01]
(Score ≥ cutoff) * (Low middle school GPA)	0.130			
	(0.0874)			
	[0.01]			
(Score ≥ cutoff) * (No parent w/HS degree)		0.092		
		(0.0811)		
		[0.51]		
(Score ≥ cutoff) * (Admission increases commute)			-0.181**	
			(0.0874)	
			[0.20]	
(Score ≥ cutoff) * (Δ mean HS peer COMIPEMS exam score > median)				0.075
				(0.0824)
				[0.14]
Observations	13136	13909	13932	13470
Adjusted R-squared	0.272	0.305	0.357	0.274
Mean of dependent variable	0.382	0.457	0.442	0.422

Panel (b) Spanish score				
	(1)	(2)	(3)	(4)
Score ≥ cutoff	-0.076	0.089*	0.149	0.045
	(0.0705)	(0.0535)	(0.0973)	(0.0640)
	[0.04]	[0.03]	[0.03]	[0.31]
(Score ≥ cutoff) * (Low middle school GPA)	0.195**			
	(0.0906)			
	[0.02]			
(Score ≥ cutoff) * (No parent w/HS degree)		-0.120		
		(0.0912)		
		[0.38]		
(Score ≥ cutoff) * (Admission increases commute)			-0.130	
			(0.1114)	
			[0.38]	
(Score ≥ cutoff) * (Δ mean HS peer COMIPEMS exam score > median)				0.030
				(0.0853)
				[0.65]
Observations	14123	13950	11309	15034
Adjusted R-squared	0.180	0.204	0.187	0.189
Mean of dependent variable	0.146	0.235	0.175	0.187

Note. Estimates are from local quadratic regressions of the specified order, including separate quadratic terms for each of the 16 IPN cutoff schools and cutoff school fixed effects. Quadratic terms and fixed effects are interacted with the corresponding covariate in each column, so that point estimates are equivalent to separately estimating the regression for each value of the covariate. The edge kernel is used in each regression and in computation of the corresponding optimal Imbens-Kalyanaraman bandwidths, which are computed separately for each value of the covariate. Huber-White robust standard errors are in parentheses. Wild-cluster bootstrapped p-values, clustered on the discrete COMIPEMS score values, are in brackets.

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