Have Elite Schools Earned their Reputation?:
High School Quality and Student Tracking in Mexico City

Andrew Dustan*
Second Year Paper
April 21, 2010

Mexico City uses a competitive admissions process to allocate students among its public high schools. Among these are school systems with a reputation as elite institutions with perceived quality far surpassing the other systems. This paper exploits the allocation mechanism to estimate the academic quality of elite schools in comparison to other public schools. Admission to the subset of elite schools examined increases end-of-high school test scores by an average of 0.19 standard deviations. The benefits of elite schools are found to accrue disproportionately to students of higher socioeconomic status, due both to lower chances of elite school admission for low-SES students as a result of lower test scores and also the fact that low-SES students request elite schools less frequently.

1. Introduction

A salient feature of public education systems in many countries, both developing and developed, is the existence of large disparities in quality between schools. Sizable amounts of money are often spent on a small set of elite or flagship schools that benefit a relatively small number of students, while the rest of the population is served by schools of lower quality and less funding. While this phenomenon is most visible in higher education, it exists at the secondary level as well. Such an allocation of resources and attention may have not only distributional consequences—recipients of an elite education gain an advantage over those without access—but also efficiency costs due to good students receiving low-quality instruction.

Mexico City's public high school system is a good example of perceived disparities in quality. High schools affiliated with two of the country's most prestigious universities, the Instituto Politécnico Nacional (IPN) and Universidad Nacional Autónoma de México (UNAM), are regarded as providing a vastly better education than the other public schools. Mexico City's high school system is notable in another area as well: it uses a standardized exam, along with students' stated preferences for schools, to allocate all students among the different schools. The result is an intense competition for a limited

* I thank my advisors, Alain de Janvry and Elisabeth Sadoulet. Thanks also go to Rafael de Hoyos and the advisors at the SEMS for their assistance with the data and understanding of institutions, as well as Lydia Ashton, Fang Lai, Ethan Ligon, Ayako Matsuda, Sofia Villas-Boas, and participants at the University of California, Berkeley ARE development workshop for their comments. All errors are mine.

number of seats in the IPN and UNAM schools, leaving many ambitious students outside of their preferred schools because their exam scores are too low and many talented students outside of competitive schools because they did not put them on their list of preferences. The desire for admission to an IPN or UNAM school is so prevalent that the myriad of entrance exam preparation companies often advertise themselves by warning students "no te quedes fuera" (don't be left out) of these schools.

This paper exploits the school allocation mechanism to confront the perception of the elite high schools as genuinely high-quality institutions by estimating their academic quality relative to other public schools. A simple regression discontinuity design is used to discover whether students experience a gain in test scores from being admitted to an elite school, using their next most-preferred school that would accept them as the counterfactual. This approach is useful because it does not require assumptions about the nature of student preferences over schools. It also estimates a quantity relevant to students by showing how high the stakes are in the allocation competition and how much a student loses academically by barely missing his chance at a top school. The resulting estimates indicate that the stakes are indeed high for students targeting elite schools: students who score barely high enough to attend an IPN school experience an average gain of 0.19 standard deviations on their exit exams, compared to a benefit of admission to other schools that is nearly zero on average.

While a structural treatment of student preferences is not the subject of this paper, I also present reduced form evidence showing that students of lower socioeconomic status are much less likely to attend an elite high school. This is not simply due to the obvious point that low-SES students score lower on the entrance exam and thus cannot enter top schools, but also because they choose elite schools less often even when comparing to nearby high-SES students with the same entrance exam score. Thus the allocation of already-advantaged students to elite schools, both in terms of academic achievement and family background, appears to exacerbate inequality in educational outcomes.

The rest of the paper is organized as follows. Section 2 reviews the school quality and academic outcomes literature. Section 3 gives a detailed overview of the Mexico City high school admissions system. Section 4 sets forth the methodology for identifying school quality. Section 5 describes the data and Section 6 gives the empirical results from estimating quality. Section 7 concludes.

2. Review of literature

Identifying the causal effect of attending a particular school on a student's academic outcomes is difficult because students are rarely randomly assigned to schools without regard for their unobservable characteristics. Some studies, for example Dearden, Ferri, and Meghir (2002) and Newhouse and Beegle (2006) take the view that a selection on observables approach is sufficient to identify some aspects of school quality. Their key assumption is that the school attended is as good as randomly assigned after conditioning on family and student characteristics, in the latter case including a family fixed effect.

An ideal research design might randomly assign a subset of students to schools and evaluate their academic outcomes as indicators of the effect of attendance to the assigned schools. While the random assignment of students to schools is rare, such circumstances have been exploited by some authors. Gould, Lavy, and Paserman (2004) use the quasi-random assignment of Ethiopian immigrants to elementary schools in Israel to estimate the causal effect of school quality on academic performance, where school quality is measured by a school's test scores prior to the assignment of immigrants.

Many school systems in the United States and throughout the world have introduced some form of lottery system for school assignment. Typically, students and their families express their preferences over the set of available schools, after which the lottery allocates students based both on preferences and luck. Hastings, Kane, and Staiger (2006) and Hastings and Weinstein (2008) examine such a system in Charlotte, North Carolina and conclude that allocation to a high-performing school due to winning the assignment lottery improves student test scores. Conversely, Cullen, Jacob, and Levitt (2005 and 2006) find that even when students win a spot in a high-performing Chicago school, academic outcomes do not necessarily improve.

Recent work has used merit-based school assignment mechanisms similar to the one used in Mexico City to measure aspects of school quality. Jackson (2010) uses the high school assignment mechanism in Trinidad and Tobago to measure the effect of peer quality on test scores. He utilizes the assignment rules to create an instrument for predicted school assignment, which is used in the first stage to predict the mean entrance exam score of a student's peers. Predicted mean peer score is then used as the regressor of interest in the second stage to explain student end-of-high school exam performance.

The use of peer ability to measure school quality is not new, but the innovation of Jackson's paper is to use the assignment rule to generate exogenous variation in peer ability. This is a peculiar exercise in the context of an assignment system in which students explicitly sort on ability. Suppose that peer ability has no causal effect on own performance but that most students demand schools that have the highest quality in terms of teachers, resources, etc. Then the highest-scoring students are usually sorted into schools with the most resources, and regressing exit exam scores on predicted peer entrance exam scores will attribute performance differences to peer quality. Preference- and merit-based assignment systems do not allow for credible disentanglement of the roles of school resource vs. peer effects. This paper will focus instead on the reduced form impact of admission.

Clark's (2007) paper is most similar in spirit to the present study. He employs a regression discontinuity design using entrance exam assignment rules for attending elite high schools in the United Kingdom, finding little effect of admission on exit exam scores four years later. This paper uses related methods to see if a similar result is found in the context of a developing country where the line between "elite" and "non-elite" high schools is stark and perceived differences in quality between them are extremely strong.

3. Mexico City public high school system and student enrollment mechanism

Beginning in 1996, the nine public high school systems 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 (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 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 matter 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 curriculum.

A potentially important aspect of the examination is that there are actually two different exams given, with the exam taken depending on which school the student selected as her first choice. If a student chose a high school affiliated with the UNAM, then she must take the exam prepared by the UNAM. Otherwise she takes the exam prepared by the Centro Nacional de Evaluación (Ceneval).[3] Both exams are administered on the same weekend. Informational materials distributed to students and their families claim that the exams are equivalent. Each exam is scored by its respective preparer.

After the scoring process, assignment of students to schools is carried out in July by Ceneval, under the observation of representatives from each school system and independent auditors. The assignment process is as follows. First, each school system 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 less than 31 points or failed to complete middle school is disqualified from participating. Next, a computer program proceeds in descending order through the students, assigning each student to her highest-ranked school with seats remaining when her turn arrives. In some cases, multiple students with the same score have requested the final seats available in a particular school, such that the number of students outnumbers the number of seats. When this happens, the representatives in

[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 I will use the term "school" to refer to a program throughout. Programs in the same school are *never* combined in the analysis. No elite school has multiple programs at the same school.

[3] The reason for this separation of examinations is historical and political rather than practical.

attendance from the respective school system must choose to either admit *all* of the tied applicants, exceeding the initial quota, or reject *all* of them, taking fewer students than the quota. If by the time a student's turn arrives, all of her selected schools are full, she must wait until after the selection process is complete and choose from the schools with open spots 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.[4]

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 decisions offer a powerful avenue for strategically shaping a school's student body, other than the obvious case of drastically underreporting available seats at a school to reduce enrollment. Setting an artificially low seat quota and planning to accept students up to a level close to "true" capacity in the event of a tie either results in the school being under-enrolled (if there are too many tied students to accept) or enrolled near the level that would prevail with the true quota reported and all ties rejected.

It is not clear whether some students can improve placement outcomes by strategically (not) choosing an UNAM high school as first choice in order to take a particular version of the exam. Even if one exam is easier than the other, this is not a prevalent perception. If exam version is not a consideration, then listing all desired schools in the order of true preference is a dominant strategy. Choosing a school that fills up quickly as a first choice does not penalize students whose scores are too low for admission, because they are allowed to compete for seats in their second or lower options. Leaving aside the issues of ties and of students remaining unassigned if all of their choices fill up, the system can be thought of as lining up all students in order of score, then asking them to choose their favorite school from the list of remaining options when they reach the front of the line.

At the end of the final year of high school (grade twelve), students who are currently enrolled

[4] To illustrate the assignment process, consider an example. A student chooses schools A, B, C, and D, in that order. He has the 100th highest Comipems exam score, so he is assigned after the 99 students who scored higher than him. When his turn arrives, all seats at schools A and C have been filled, but there are available seats at schools B and D. The student is assigned to school B since it is his highest-ranked available school.

take a national examination called the Evaluación Nacional de Logro Académico en Centros Escolares (Enlace), which tests students in Spanish and mathematics. Even students who are not expected to graduate on time take the exam. This examination has no bearing on graduation or university admissions. It is a benchmark of student and school achievement and progress.

4. Methodology

The principal goal of this paper is to determine whether elite public high schools actually have high quality *relative* to other public schools. Put another way, the econometric challenge is to measure the effect on academic outcome from admission to a school in an elite system versus the next-best option. Here I define outcome as 12^{th} grade Enlace score so that the effect of interest is the change in Enlace score due to elite school admission instead of the student's next choice, holding Comipems score and all student characteristics constant. This effect results from all aspects of a school that a student experiences as a result of being admitted, including the characteristics of the school itself (teachers, resources, etc.) as well as those of the student body that has been selected into each school.

Because students have different preferences over schools, assignment is not random conditional on Comipems score. If preferences are correlated with unobserved student characteristics that affect score gains between the Comipems and Enlace exams, then naïve estimates of school quality such as those from equation (1) are biased:

$$Enlace_{ij} = f(Comipems_i) + \mu_j + \epsilon_{ij}, \tag{1}$$

where $Enlace_{ij}$ is the Enlace score for student i who attended school j, f is a function of Comipems score, μ_j is the fixed effect for school j, and ϵ_{ij} is an i.i.d. error term. (I suppress other student observables for simplicity.) If school assignment were random conditional on $Comipems_i$, then $\hat{\mu}_j$ would be an unbiased estimator of school j's quality and $\hat{\mu}_A - \hat{\mu}_B$ would give the effect on Enlace score from attending School A instead of School B. But $E\left[\epsilon_{ij} \middle| Comipems_i, school = j\right] = 0$, the necessary assumption for unbiased coefficient estimates, is clearly tenuous and is unlikely to hold.

The Comipems assignment mechanism allows for a more rigorous strategy for identifying relative school quality, namely a sharp regression discontinuity (RD) design. Each school j that is oversubscribed (i.e. it has more demand than available seats) accepts all applicants at or above some cutoff Comipems exam score c_j, and rejects all applicants below c_j. Whether or not a student who wants to attend a particular school is actually admitted is determined entirely by whether or not he is above or below the cutoff score, giving a sharp discontinuity in the probability of admission when the student reaches the cutoff. Considering one school at a time, the RD specification for school j is:

$$Enlace_{ij} = g_j(Comipems_i) + \delta_j\ admit_{ij} + \epsilon_{ij} \quad (2)$$

where g is a function of Comipems score and $admit_i$ is equal to 1 if the student was admitted to school j and zero otherwise. The sample consists only of students who would have liked to attend school j when their turn for assignment arrived. That is, they listed school j as a preference and when their turn for assignment arrived, all schools listed above school j had already been filled. Hence $admit_{ij} = 1$ if and only if $Comipems_i \geq c_j$. We see that $\mathbb{I}(Comipems_i \geq c_j)$ is an instrument that perfectly predicts $admit_{ij}$ for the selected sample. Provided that the control function g is specified correctly and is continuous at $Comipems_i = c_j$, $\hat{\delta}_j$ gives the estimated local average treatment effect (LATE) of admission to school j compared to admission to those schools attended by rejected students (Imbens and Lemieux 2008).

An advantage of this estimation strategy 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 accepted and rejected students near a school's cutoff have the same expected characteristics, including school preferences. Even if students are choosing strategically or making mistakes in their selections, this behavior should 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.

Because (2) uses the set of schools attended by students rejected from school j as the comparison group for measuring school j's relative school quality, and the set of schools attended by rejected students varies depending on the choice of j, the LATEs are not comparable across schools. Rather, $\hat{\delta}_j$ measures how much the marginally accepted students gained by being admitted to school j, compared to the counterfactual of admission to their next most-preferred alternative that would have accepted them. These are useful estimates for evaluating relative quality of elite schools. If the elite schools' LATEs are large and positive, then missing out on admission has a large academic cost. This would provide credible evidence that elite schools are of significantly higher academic quality than other public schools and hence that the stakes are high in competing for a spot at an elite school.

5. Data description

5.1. Comipems and Enlace exam results

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 Comipems entrance examination process, and the scores from the 2008 12th grade Enlace exam. 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, the version of the exam taken (UNAM or Ceneval), exam score out of 128 points, and the school to which the student was assigned as a result of the assignment process. Assignment data includes the school chosen after the initial allocation for students who missed the cutoff for all of their listed schools. It also includes student responses to a multiple choice survey turned in at the time of registration for the exam, but many students filled out only parts of the survey and non-response is high for some questions. Response rate is high for parental education and family income, at 88% for each.

The Enlace dataset consists of exam scores for all students who took the test in Spring 2008, the first year that the 12th grade Enlace was given. The scores are reported as a continuous variable, reflecting the weighting of raw scores by question difficulty and other factors. I normalize the scores by subtracting off the sample mean and dividing by the sample standard deviation. The Enlace scores are matched with the 2005 Comipems-takers by using the Clave Única de Registro de Población (CURP), a unique identifier assigned to all Mexicans. Matching is performed by name if no CURP match is found.

There are two issues with the Comipems and Enlace datasets that need to be addressed carefully. First is that of all Comipems examinees who are assigned to a school (either during initial allocation or afterward), 12% are admitted to high schools affiliated with the UNAM, but these schools chose not to have their students take the Enlace. This is unfortunate because the UNAM is one of the two school systems in Mexico City that are considered elite. Without Enlace data for the UNAM system, the focus is shifted to the IPN high school system, which is the other group of elite schools.

The second issue is that there is a high attrition rate between the Comipems exam and the Enlace three years later. The 2005 Comipems assignment process allocated 195,802 students to schools outside of the UNAM system. But only 79,588 (41%) have a valid Enlace score reported in 2008.[5] This feature of the data is discussed at length in Section 6.

The concept of a student being located at a specific school's admission threshold is key to the empirical analysis, so it is explained in detail here. Consider a particular school, called school A. Determining the set of students above A's threshold is simple. Any student who was admitted to A but was less than 5 Comipems exam points above A's cutoff score is considered above A's threshold. This is the set of all students admitted to A who would be rejected if they lost 5 points or less. The set of students below A's threshold is more subjective. Any student attending a different school is considered below A's threshold if he would have been admitted to A upon gaining a small number of points. Specifically, if the student would attend A by gaining one point, he is below A's threshold. If he would attend school B upon gaining one point and A upon gaining two points, he is at both A and B's thresholds. The maximum number of points that a student must gain to be admitted to a school and still be considered at its threshold is set to 5.

Summary statistics for the full sample of Comipems-takers who were assigned to a school during initial allocation or afterward, the Comipems-Enlace matched sample, and the subsample consisting only of students located at one or more school thresholds are in Table 1. Students chose nine schools on average, but ten percent of students did not get into any of their listed choices during initial allocation and had to choose from the remaining seats. Compared to the full sample of assigned students, those who took the Enlace have slightly lower family income and parental education, have

[5] The sample size is 79,231 because 284 Enlace takers attended a school that did not report complete Enlace results and 73 had other data errors.

higher grade point averages, and are more likely to be female. The parity in Comipems scores between these samples masks the fact that many low-Comipems students dropped out prior to the Enlace. No UNAM student could take the Enlace, which lowers the Comipems mean in the Enlace sample as well as income and parents' education. The major difference between all Enlace takers and the subsample of those located at a threshold is in exam scores. This is because the highest scoring students in the sample were likely to have Comipems scores high above the cutoff for their first choice school, meaning they appear less frequently in the threshold sample.

Figure 1 gives the distributions of the Comipems and Enlace exam scores for each sample, where applicable. These suggest that there is no test score ceiling for either exam. Score ceilings present a problem for assessing school quality because there is no way for students with the highest score to demonstrate academic progress. The Comipems exam intentionally avoids a ceiling in order to sort students during assignment. The Comipems score distribution is truncated below at 31 points because scoring lower than 31 disqualifies students from assignment to any school. While it is uncertain from the data whether any student attained the maximum Enlace score, the right tail has very little mass. As expected, the full sample has more students at the upper end of the Comipems distribution because it includes UNAM admittees while the Enlace subsamples do not.

5.2. School system characteristics

There are nine public school systems in the Comipems consortium. The appendix gives more information about the six largest systems, which account for 98.5% of all students admitted. Table 2 presents system-level summary statistics for these six systems. Only the UNAM and IPN filled all of their seats, admitting 53,667 of the 230,427 students. Their mean Comipems scores are far higher than the other systems and the IPN's mean Enlace score surpasses the next best system by 1.04 standard deviations.

Figure 2 summarizes exam score distributions at the school level. Panel A shows the distribution of cutoff scores for schools that were oversubscribed. Of the 634 schools, 378 (62%) were oversubscribed. Most oversubscribed schools had cutoff scores below the sample mean exam score.

That is, relatively few schools would reject an above-average student who applied. Almost all of the high-cutoff schools belong to the UNAM or IPN systems. Panels B and C show the mean Comipems and Enlace exam scores of all schools, respectively. The dispersion in both scores is quite large. Both distributions have substantial mass in their right tail, indicating that there are sizable subsets of schools with above-average students prior to and/or upon completion of high school. The majority of these are UNAM or IPN schools.

6. Empirical analysis

6.1. Correlates of UNAM and IPN admission

The previous discussion showed that UNAM and IPN high schools generally have the highest admission cutoff scores and highest average exam scores among Mexico City public schools. Here I provide evidence that these school systems are the most demanded by students and that the competition allocates students of high socioeconomic status to these schools. The objective is to show that if we find large benefits from admission to the elite schools, these benefits accrue disproportionately to children of wealthier and better-educated parents.

First-choice demand for the UNAM and IPN schools far surpasses that for the other school systems, as illustrated in Table 3. For every available seat in the UNAM system, there are about 3 applicants selecting an UNAM school as their first choice. For IPN schools, the ratio is nearly 2. All other systems have more supply of seats than first-choice demand (although demand is at the school level and some of these schools do fill up, often with people rejected from the UNAM or IPN schools).

Who wins a spot in the highly-demanded UNAM and IPN schools, given that the assignment system allocates on the basis of score? Figure 3 presents simple graphical evidence that admitted students are from more advantaged socioeconomic backgrounds. The solid line shows that students with higher-educated parents obtain higher Comipems exam scores, thus increasing the chance of being admitted to a competitive school if requested.[6] The dashed line shows that among students who end up

[6] This and the following results are qualitatively similar using family income instead of parental education. I focus on parental education because measurement error of income is very high at the individual level. Income is only reported in 1000-peso ranges.

obtaining a Comipems score high enough to attend the least-competitive school in the UNAM or IPN system, stated preference for UNAM and IPN schools rises in parental education. Thus SES is positively correlated with the chance of attending an elite school through both exam score and choice of schools. The dotted line bears out this correlation further, showing that actual assignment to an elite school rises with parental education. For example, only 13% of students whose parents have an elementary education attend an elite school, as opposed to 48% of students with parents holding a bachelor's degree.

To confirm that *conditional on Comipems score*, children of more-educated parents are more likely to attend an elite school, the following equation is estimated for the sample of all students with a high enough score (65 points) to attend at least one UNAM or IPN school: $Admit_{ims} = \alpha_m + \sum_{c=65}^{128} \mathbb{I}(s = c) \times (\theta_c + \beta_c educ_i) + \epsilon_{imc}$, where $Admit_{ims}$ is a dummy variable equal to 1 if student *i* from municipality/delegation[7] m with score s was admitted to an UNAM or IPN school and $educ_i$ is years of parental education. The parameters of interest are the β_c's, which measure the marginal effect (though not a causal relationship) of parental education on elite school admission only for students with $s = c$. Figure 4 graphs these coefficients and shows that for all values of Comipems score, higher education is correlated with higher rates of elite school admission. For example, at a score of 80 points, moving from elementary school graduate to bachelor's degree increases the probability of elite school assignment by 22% (over a base rate of 54% for elementary graduates).[8] This is a large effect, 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 of lower SES are much less likely to attend one.

While the underlying preferences that generate observed school choices are beyond the scope of this paper, it is worth noting that the results in Figures 3 and 4 do not necessarily imply that parents and students from poorer backgrounds value elite schools (or school quality) less. Suppose that elite schools are perceived to have high quality by all families and that families select schools based only on quality

[7] There are 41 municipalities and 16 delegations comprising the Mexico City metropolitan area, covering 7,815 km^2.
[8] The estimated education effect is lower for scores near 65 because few students with those scores attend an elite school. Similarly for scores over 100 because almost all students with those scores do attend.

and distance from home. Then poorer families may select fewer elite schools because 1) they care less about quality, 2) they incur a higher utility cost from school distance because transportation costs are more burdensome or children need to work, or 3) they live farther away from elite schools. In any case, the estimation of relative school quality does not require an understanding of families' preferences and constraints, and the simple evidence presented here shows that children of low SES are less likely to become the beneficiaries of any school quality advantages offered by elite schools.

6.2. Regression discontinuity results

This section focuses on the RD estimates of relative school quality for the IPN high schools, since students at the UNAM schools did not take the 12th grade Enlace exam. First, however, the RD estimates for all oversubscribed schools are summarized. The empirical specification of equation (2) that is estimated is:

$$\begin{aligned} Enlace_{ij} = \alpha_j + \beta_j Comipems_{ij} + \beta_j'\left(Comipems_{ij} \times admit_{ij}\right) + \delta_j\, admit_{ij} \\ + \gamma_j UNAM_i + \epsilon_{ij}. \end{aligned} \tag{3}$$

The sample for school j's regression is limited to students who wanted to attend the school when their turn for assignment arrived and who were within five points of the cutoff score, either above or below. The 439 students who would have attended an UNAM school upon rejection are excluded, as their rejected counterparts did not take the Enlace. The control function takes a piecewise-linear functional form. Higher-order polynomials are not desirable because Comipems score is discrete, with only ten values in the domain. A dummy variable for whether the student took the UNAM version of the entrance exam is included to increase precision. To account for possible correlation of the errors between students from the same middle school, all oversubscribed schools' regressions are run simultaneously with errors clustered at the middle school level. Any students appearing at more than one school's admissions threshold have one observation per threshold. Schools with fewer than five admitted or five rejected students inside the five-point window are excluded. This limits the estimation of admission LATEs to 304 of the 378 oversubscribed schools.

Reporting full regression results for each school is not feasible. Instead, we can examine the distribution of the LATEs to get an idea of the extent to which marginal admission to or rejection from a school might matter for exam scores. To obtain the mean of all of the LATEs, weighted by the number of students at each school's admission thresholds, I estimate a regression similar to (3):

$$(4) \quad \begin{aligned} Enlace_{ij} = \alpha_j + \beta_j Comipems_{ij} + \beta_j^{'}\left(Comipems_{ij} \times admit_{ij}\right) + \bar{\delta} admit_{ij} \\ + \gamma_j UNAM_i + \epsilon_{ij}, \end{aligned}$$

where $Comipems_{ij}$ has been normalized to be distance from school *j*'s cutoff score. The difference between (4) and (3) is that $\bar{\delta}$ measures, for *all* students in the sample, the mean effect of marginal admission versus rejection and subsequent attendance at the next-best school. The point estimate of $\bar{\delta}$ is nearly identical to the weighted sum of $\hat{\delta}_j$'s in (3), but (4) has the advantage of allowing easy estimation of the clustered standard error for the average LATE. Table 4, Panel A reports the results: choosing a marginally accepted or rejected student at random, the expected gain in Enlace score due to marginal admission to the preferred school is only 0.06 standard deviations. For students not at an IPN admission threshold, the average stakes are even lower, with admission yielding 0.03 standard deviations on the Enlace.

The admission LATEs from (3) are noisy due to small sample size. The median sample size per admission threshold is 52 students. The median standard error for the LATEs is 0.31, which is large considering the standard deviation of the LATE estimates themselves is 0.44. Figure 5 shows the distribution of the estimated LATEs, while Figure 6 plots them against school cutoff score. There is little discernible relationship between LATE and cutoff except that LATEs appear to increase for the highest cutoff schools. Dispersion of the estimated LATEs is high for most cutoff score values.

In order to know how much marginal admission matters, we need to go beyond the mean of the admission LATEs and estimate their true variance, which is overstated due to measurement error. Koedel and Betts (2007) propose a method for estimating the true variance of fixed effects in the context of teacher value added, which can be generalized to the variance of the LATEs since they are jointly estimated in (3). Summarizing their method: first, they decompose each estimated fixed effect into the true effect and unobserved estimation error, $\hat{\delta}_j = \delta_j + \lambda_j$. Assuming that $Cov(\delta, \lambda) = 0$ gives

$Var(\hat{\delta}) = Var(\delta) + Var(\lambda)$. Koedel and Betts show that by scaling the $\hat{\delta}_j$ parameters using their variance-covariance matrix, they can approximate $Var(\hat{\delta})/Var(\lambda)$, which in turn allows them to recover an estimate of $Var(\delta)$.[9]

This process is carried out for the sample size-weighted LATEs (with proper weighting of the covariance matrix), giving $\widehat{SD(\delta)} = 0.13$ for all schools and $\widehat{SD(\delta_{-IPN})} = 0.08$ for non-IPN schools. These estimates are much smaller than those not accounting for measurement error, and imply that outside of the IPN (and UNAM), the mean effect of marginal admission to a particular school is very unlikely to approach 0.2 standard deviations on the Enlace. The estimates also suggest that marginal admission often has a negative effect. Thus schools outside the IPN that offer large benefits to admission versus the next-best option are rare. This does not imply that all schools are of similar quality. Rather, there may exist a continuum of quality so that rejection from one school merely causes a student to fall slightly in the quality distribution to a similar school.

6.3. Regression results for IPN schools

In contrast with the other school systems, the IPN schools yield substantial benefits for students who are admitted at the margin. Panel B of Table 4 shows the results from estimating (4) for IPN schools only. The average LATE is 0.19 standard deviations on the Enlace, much larger than the 0.03 for non-IPN schools. The effect is largest for students who would be forced out of the IPN system upon rejection—there, the average LATE is 0.24. For students who would attend a different IPN school upon rejection, the effect is still substantial at 0.17.

These results suggest that elite schools (at least the IPN) deserve their reputation for high quality relative to other public schools. Even when students are able to attend their next-best option outside the IPN, rejection costs them 0.24 standard deviations on the Enlace. There do not appear to be options outside the elite schools that are of sufficient quality that students can "step down" the quality

---

[9] More precisely, note $Var(\hat{\delta}) = \left(\frac{1}{J-1}\right)\sum_{j=1}^{J}[\hat{\delta}_j - \left(\frac{1}{J}\right)\sum_{j=1}^{J}(\hat{\delta}_j)]^2$ where J is the number of LATEs estimated. Define $\tilde{\delta}$ as a Jx1 vector containing the sample average of the LATEs in each entry, and $\hat{V}_J^{-1}$ as their variance-covariance matrix. Then $\widehat{Var(\delta)} = Var(\hat{\delta}) - \frac{Var(\hat{\delta})}{\frac{1}{J-1}[(\hat{\delta}-\tilde{\delta})\hat{V}_J^{-1}(\hat{\delta}-\tilde{\delta})]}$.

distribution gradually, as appears to be occurring between the non-IPN schools. The results give another interesting insight: even within the IPN system, the school to which a student is admitted matters. Figure 7 gives a graphical representation of each RD regression, while Table 5 gives the full school-by-school RD regression results for all 16 schools as well as various school characteristics. Notably, the mean Comipems score of admitted students is quite different between schools. While the coefficients on admission are not comparable across the different regressions because they measure admission versus different counterfactual outcomes, it is clear that rejection from different schools can have very different consequences. Rejection from the top IPN school, campus 9, costs 0.41 standard deviations on the Enlace while rejection from Campus 10 has no discernible effect. To summarize, acceptance to an IPN school on average has large positive benefits both for students who are at the margin of the system and those who would remain in the system upon rejection, but this effect does not appear to be uniform across schools. This is strong evidence that the IPN schools are indeed of high relative quality.

Having shown that those of higher SES are more likely to be the beneficiaries of the elite schools' quality advantage, we may wonder if those of lower SES would benefit equally from access to this quality. Table 6 answers this question by using equation (4) to estimate the IPN admission effects separately for the children of high- and low-educated parents. For each group, the observations are weighted so that the contribution of each IPN school to the mean admission effect is the same as for the full-sample estimates. This is done so that any differences in the estimated effects are due to differences in how the groups are affected by admission or rejection to the same composition of schools. Columns (1) through (3) reproduce the LATEs from Panel B of Table 4, for reference. Columns (4) through (6) give analogous results for students whose parents have a middle school education or less. Columns (7) through (9) do the same for those with higher than middle school education. The average effect of admission is slightly smaller for the low-education group: 0.16 standard deviations on the Enlace compared to 0.21. The difference is more pronounced when only considering students who would leave the IPN system upon rejection: 0.18 and 0.27 standard deviations, respectively. Still, the benefits of IPN admission appear to be large regardless of parental education.

Results from regressions using 2 through 10 point window sizes, not reported here, are very similar to those obtained with the 5-point window. Adding a cubic control function for the 10 point window regressions does not change the results.

6.3. Attrition and robustness

The high attrition rate between Comipems and Enlace exams is a potential challenge to the validity of the empirical results. I address this issue with several arguments, concluding that while attrition is nonrandom and leads to an overstatement of elite school effects, it is unlikely that they are driving the qualitative nature of the results. First, I show that the attrition rate observed in the matched Comipems-Enlace data is similar to that in aggregate administrative data, suggesting that observed attrition reflects actual schooling behavior rather than a data problem. Second, few students transfer between schools, so assignment is a good proxy for attendance (i.e. little within-sample, between-school attrition). Third, although students admitted to the IPN experience higher rates of attrition and this attrition is arguably among weaker students, accounting for this fact does not change the results importantly.

There are several reasons for the high attrition between the Comipems and Enlace exams. The importance of each reason can be measured using aggregate school-level data from a school census and statistics released by the Enlace administrators. First, some students participated in the Comipems competition but never enrolled in a public high school. The number of new 10^{th} graders in the school census is about 90% of the total number of students assigned through the Comipems process. Second, comparing new 10^{th} grade enrollment in 2005 and new 12^{th} grade enrollment in 2007, about 38% of enrolling students drop out or fail a grade prior to 12^{th} grade. Failing a grade would mean that the student does not take the 2008 Enlace. Third, the school-level Enlace statistics indicate about 14% of registered 12^{th} grade students did not take the Enlace due to dropping out during the year or failing to appear for the exam. Schools have little incentive to prevent certain students from taking the Enlace because the exam scores are not used to evaluate schools. The implied rate of attrition is $1 - (0.9) \times$

$(1 - 0.38) \times (1 - 0.14) = 0.52$, which matches closely the 59% rate observed in the microdata.[10]

Of the students who take the Enlace, 93% do so at the school to which they were assigned. Since the LATEs are for assignment rather than attendance, they can be regarded as reflecting an intention to treat conditional upon remaining in the sample. But since the compliance rate is high, and because some of the 7% of non-compliers attended their assigned school for part of their high school careers, the issue of transferring schools does not appear to substantively challenge the interpretation of the admission effects as local average treatment effects of attendance.

Focusing only on the IPN admission LATEs because they form the key argument of the paper, the first step in assessing attrition's impact on the estimates is to see if admission to an IPN school changes the probability of attrition. Performing the RD regression from equation (4) with attrition as the dependent variable instead of Enlace score, column 1 of Table 7 shows that marginal admission to an IPN school raises the chance of attrition by 4.9% over a base rate of 55% among the marginally rejected students. To explore the possibility that attrition occurs among different kinds of students in the admitted and rejected schools, (4) is run with different observable characteristics as dependent variables both for the pre-attrition sample (where characteristics should be balanced) and the post-attrition sample (where attrition may have led to unbalanced observables). Columns 2 through 4 show that grade point average, income, and parental education are balanced across admitted and rejected groups pre-attrition.

After attrition occurs, however, these characteristics are no longer balanced. Columns 5 through 7 show that the admitted group has parents with a quarter of a year more education (statistically insignificant), 329 pesos/month (about $30) higher family income (marginally significant), and 0.12 points higher middle school grade point average out of 10 possible (highly significant). This is consistent with IPN schools failing the weakest of their marginally admitted students at a higher rate than the marginally rejected students experience at their schools. With better students taking the Enlace exam at the IPN schools, we expect that the effects of admission on scores are overstated. The question is to what extent this affects the results.

[10] The matching process between the Enlace and Comipems exam data is imperfect, especially when the student does not provide her CURP. This accounts for some of the small disparity between the two figures.

If the unbalanced observables, particularly grade point, account for much of the difference between attritted and non-attritted students (i.e. there is little differential attrition on unobservables after conditioning on observables), then controlling for them in the RD regression will yield valid estimates even in the presence of differential attrition. In particular, middle school grade point average may be a good proxy for student academic skills and motivation beyond what is captured in the Comipems exam score. In regressions whose results are not reported here, all RD specifications for the IPN schools were augmented with parental education, income, and middle school GPA, where the coefficients were allowed to vary by the school marginally accepting or rejecting the student. The estimated LATEs change very little in every case, almost always within 0.02 Enlace standard deviations of the original estimates. Thus differential attrition with respect to observables does not seem to have influenced the results importantly even though these characteristics appear intuitively correlated to unobserved differences in students that affect dropout or failing a grade.

While the above results provide some reassurance that attrition is not driving the RD results, it would be preferable to obtain a likely upper bound on the bias induced by attrition. Unfortunately, the extremely high attrition rate in this sample makes it difficult to decide what constitutes a reasonable bound. I propose one possibility: suppose that, conditioning on observables, the schools on both sides of the IPN schools' cutoffs fail or drop out the same kind of students, except that the additional 4.9% attrition among marginally admitted students is from students who would have scored in the 10^{th} percentile of all students in their school with the same Comipems score who did take the Enlace. This attributes all excess attrition among the marginally admitted group to the worst students. Even under this assumption, the LATE for admission to an IPN school from equation (4) is estimated to be 0.13 (SE=0.04), lower than the 0.19 estimated earlier but far larger than the 0.03 average LATE for non-IPN schools.

Finally, (4) is re-estimated only for admission to the half of the IPN schools with the lowest attrition of marginal students, where admission is only estimated to increase probability of attrition by 1%. The objective is to see if it is the high-attrition schools that drive the positive admission results. On the contrary, the average LATE for low-attrition schools is 0.20 (SE=0.06), which is slightly larger than the 0.19 estimated for all IPN schools. Similar results are obtained with this sample restriction for the other regressions run in this paper. Thus while differential attrition is certainly a concern for identification of admission effects, the evidence presented here suggests that the results are quite robust.

7. Conclusion

This paper has used Mexico City's high school allocation mechanism to identify the quality of a subset of its elite public schools relative to their non-elite counterparts. Admission to an elite school is found to positively affect student test scores, yielding benefits that accrue in large part to students of relatively high socioeconomic status. Needless to say, the policy implication is *not* to randomly allocate students to elite schools, since part of the value of these schools surely lies in peer quality. Rather, these findings provide clear evidence that the polarization of public school quality coupled with a competitive entry process does have important consequences for students. The fact that the entrance exam relies heavily on curriculum-based questions covering material that should have been taught in middle school implies that bright students in low-quality middle schools may have their bad luck persist through high school and beyond. Given that many school systems feature elite schools and competitive allocation, both in developing and developed countries, the findings here suggest that further work in this area may be useful in informing important policy decisions.

Appendix. School system details

The six largest Comipems member systems are named and described here. The Universidad Nacional Autónoma de México is the country's flagship public university and also administers several of its own high schools in Mexico City. These programs focus on traditional rather than technical curricula and offer successful students a chance to pass directly to the university after graduation. All but one of its campuses are inside the Federal District rather than the surrounding State of Mexico.

The Instituto Politécnico Nacional is also a prestigious public research university with several affiliated high schools. In contrast to the UNAM, the academic offerings include a vocational component that students can choose upon admission. Examples include electronics and industrial mechanics. All but one of the campuses are in the Federal District.

The Colegio Nacional de Educación Profesional Técnica (Conalep) schools provide technical and professional education. Examples of commonly offered educational programs are accounting and automotive mechanics. Campuses are located throughout Mexico City.

The Dirección General de Educación Tecnológica Industrial (DGETI) offers hybrid academic-technical education programs. Examples of its diverse offerings are industrial design and tourism. Campuses are spread throughout the City.

The Colegio de Bachilleres (Colbach) provides traditional academic curricula with its campuses in the Federal District.

The Secretaría de Educación (SE) of the State of Mexico administers most of the schools in Mexico City outside of the Federal District. These include a wide range of programs, from strictly academic curricula to technical degrees similar to those offered by Conalep and DGETI.

References

Clark, Damon. 2007. "Selective Schools and Academic Achievement." IZA Discussion Paper No. 3182.

Cullen, Julie, Brian Jacob, and Steven Levitt. 2005. "The Impact of School Choice on Student Outcomes: An Analysis of the Chicago Public Schools." *Journal of Public Economics*. 89(5-6): 729-760.

________. 2006. "The Effect of School Choice on Participants: Evidence from Randomized Lotteries." *Econometrica*, 74(5): 1191-1230.

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

Gould, Eric, Victor Lavy, and Daniele Paserman. 2004. "Immigrating to Opportunity: Estimating the Effect of School Quality Using a Natural Experiment on Ethiopians in Israel." *Quarterly Journal of Economics*, 119(2): 489-526.

Hastings, Justine, Thomas Kane, and Douglas Staiger. 2006. "Preferences and Heterogeneous Treatment Effects in a Public School Choice Lottery." NBER Working Paper No. 12145.

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

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

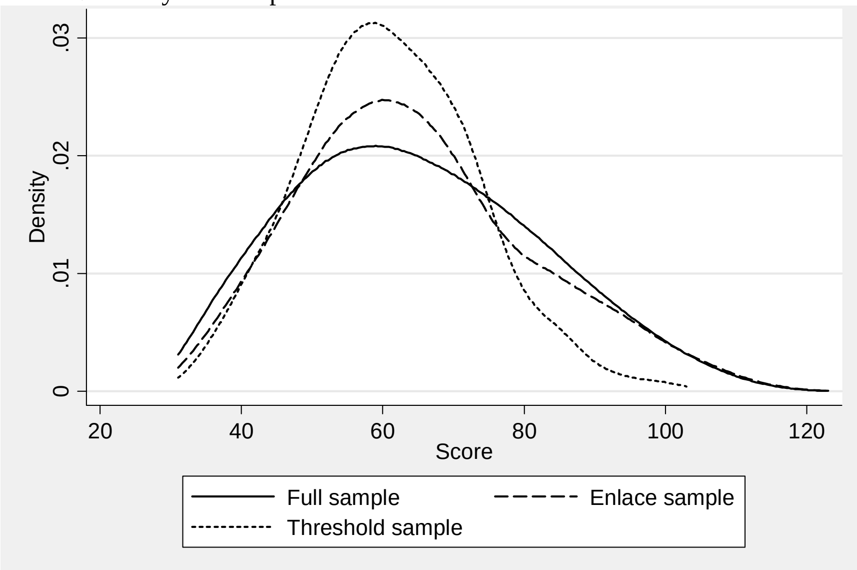
Jackson, C. Kirabo. 2010. "Ability-Grouping and Academic Inequality: Evidence from Rule-Based Student Assignments." *The Economic Journal*, forthcoming.

Koedel, Cory, and Julian Betts. 2007. "Re-Examining the Role of Teacher Quality in the Educational Production Function." University of California San Diego.

Newhouse, David, and Kathleen Beegle. 2006. "The Effect of School Type on Academic Achievement: Evidence from Indonesia." *Journal of Human Resources*, 41(3): 529-557.

Figure 1. Density of exam scores students

Panel A. Density of Comipems exam score



Panel B. Density of Enlace exam score

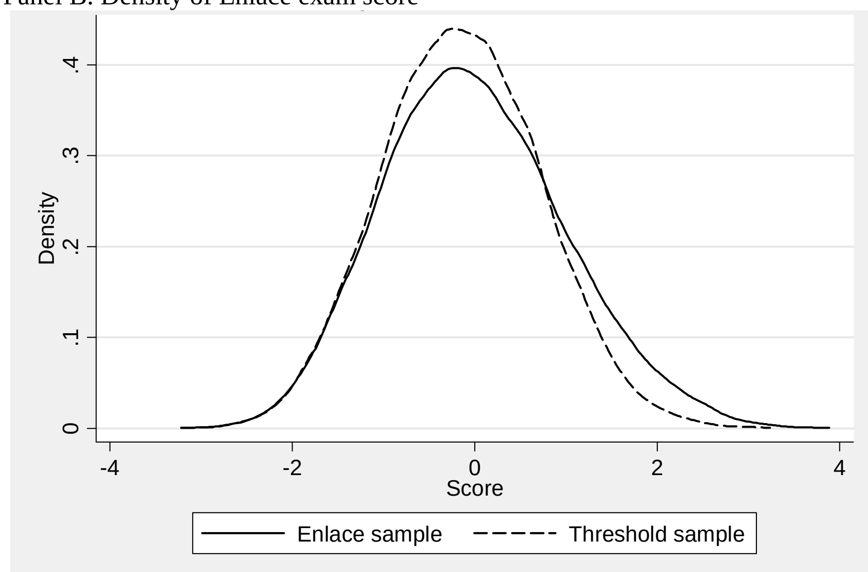
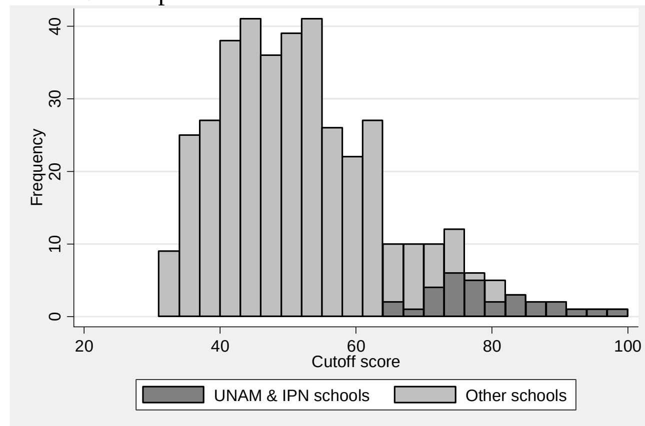
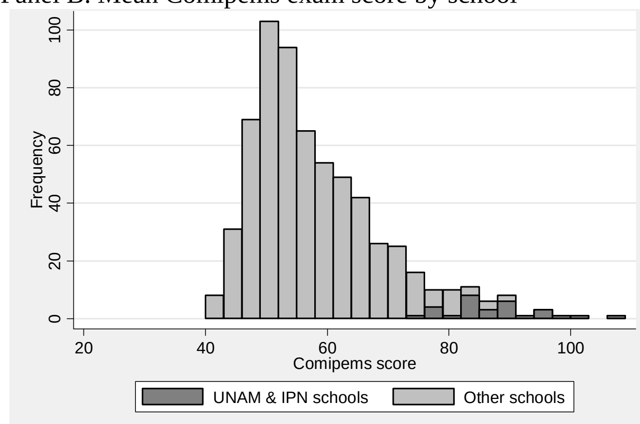


Figure 2. Exam score distributions by school

Panel A. Comipems exam cutoff scores for oversubscribed schools



Panel B. Mean Comipems exam score by school



Panel C. Mean Enlace exam score by school

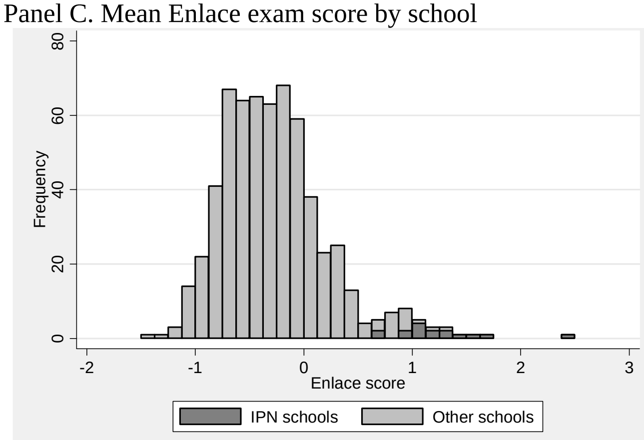


Figure 3. Comipems score, preference for and admission to elite schools by parental education

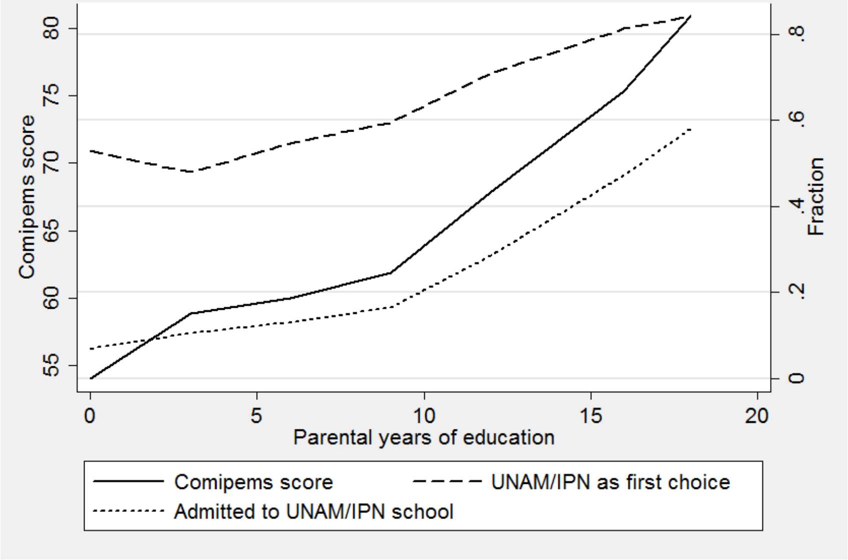
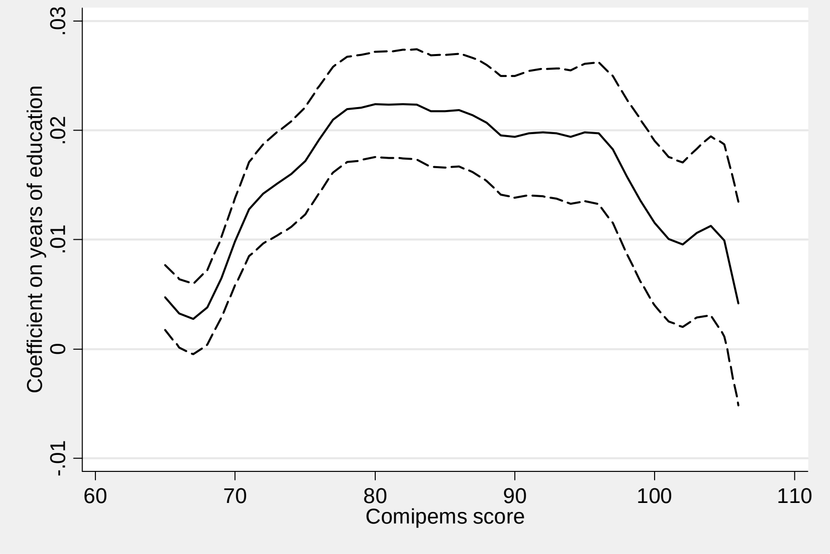


Figure 4. Correlation between elite school admission and parental years of education for different values of Comipems score



Note. Solid line is local polynomial fit through coefficients on years of education for each value of Comipems score. Dashed lines fit the 95% confidence intervals for each coefficient.

Figure 5. Distribution of school admission LATEs

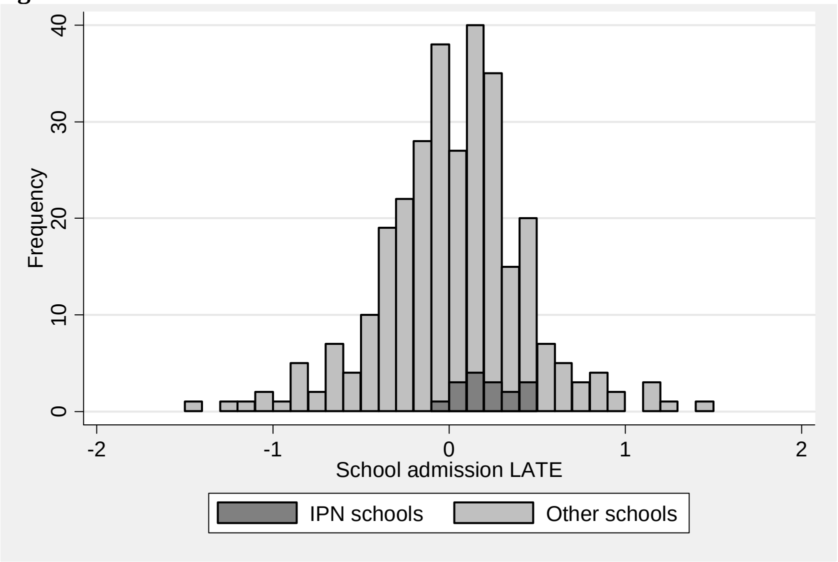
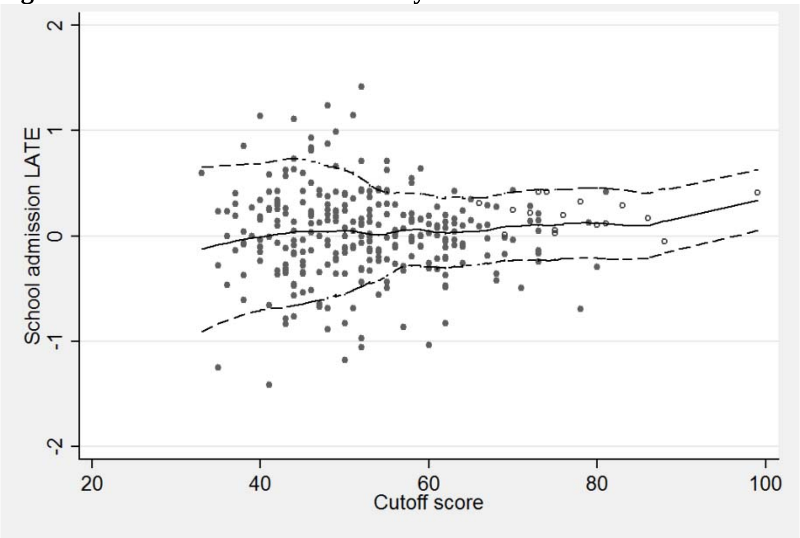
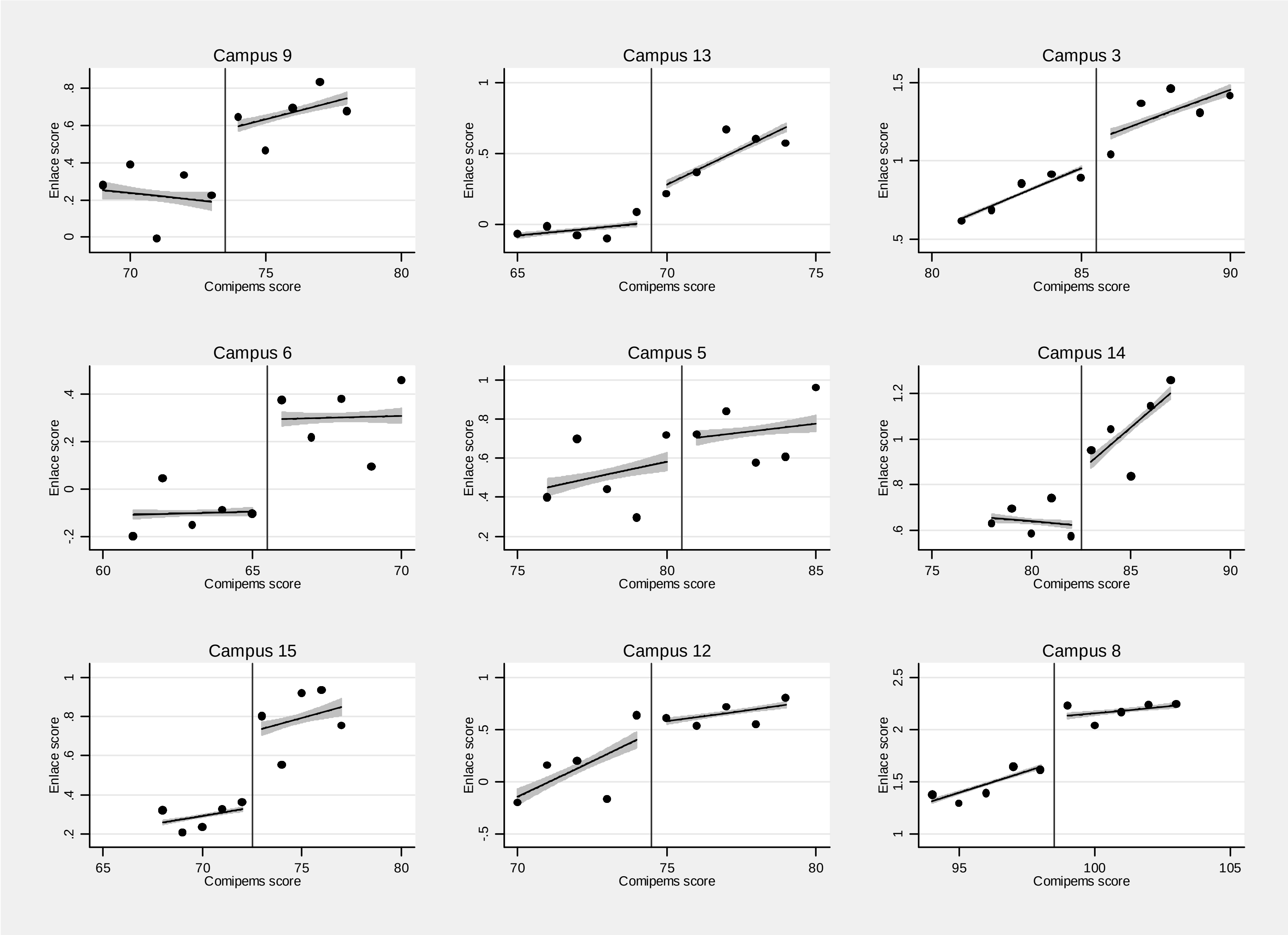


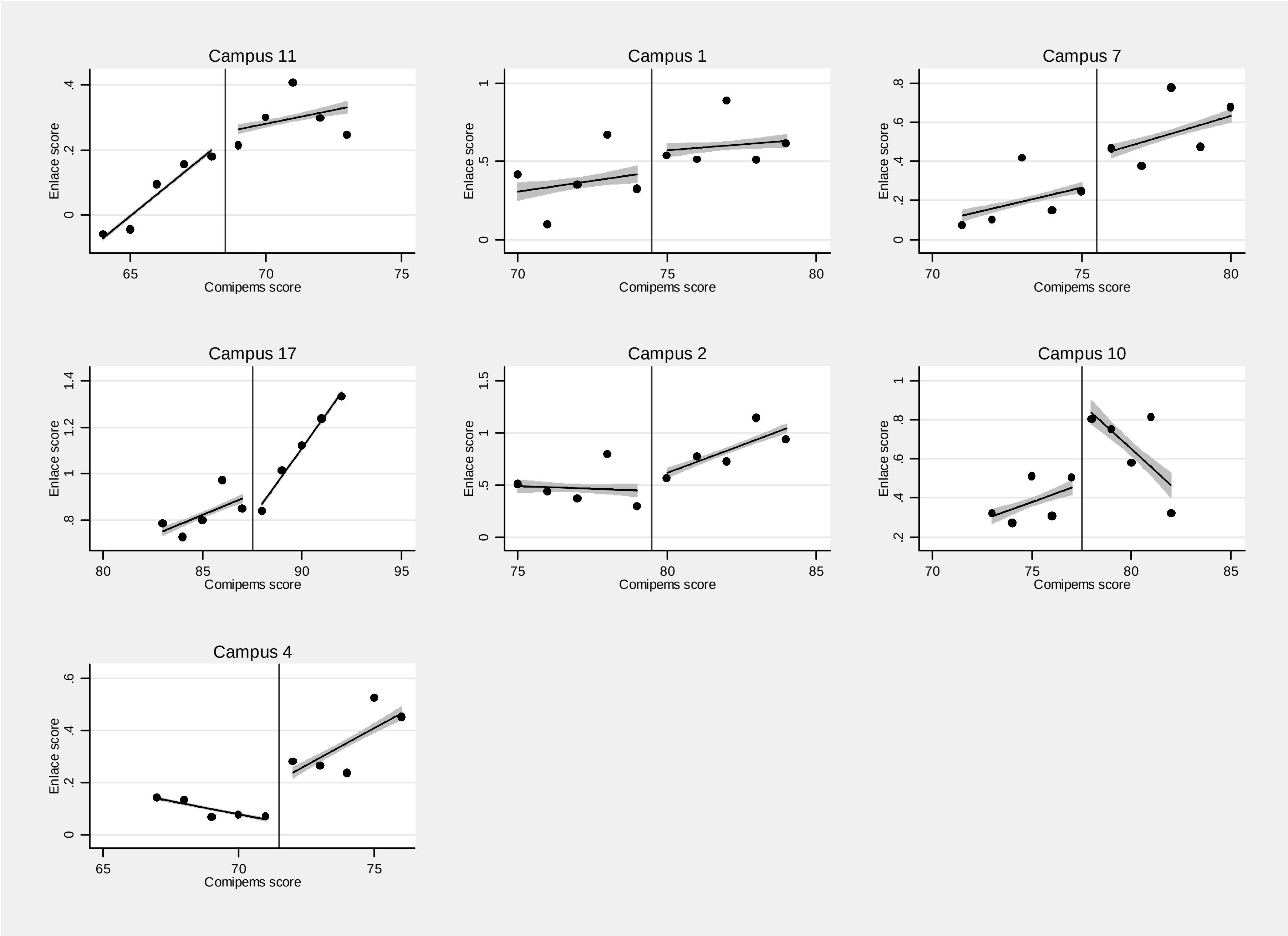
Figure 6. School admission LATEs by cutoff score



Note. Hollow circles correspond to IPN schools. Solid line is local polynomial fit through LATEs. Dashed lines fit average 95% confidence interval for each LATE.

Figure 7. Regression discontinuity plots for IPN schools





Note. Schools are ordered by cutoff score. Gray bands represent 95% confidence intervals for linear fit.

Table 1. Student characteristics for 2005 Comipems/2008 Enlace cohort

	(1)	(2)	(3)		
	Assigned students (N=230,427)	Enlace takers (N=79,231)	Students at a school threshold[a] (N=30,607)	p-value for equality of (1) and (2)	p-value for equality of (2) and (3)
Male	0.50	0.44	0.39	0.00	0.00
	(0.50)	(0.00)	(1.00)		
Maximum of mother's and father's education	10.05	9.80	9.86	0.00	0.01
	(3.43)	(3.31)	(3.24)		
Family income (2005 pesos/month)[b]	4,322	4,004	4,038	0.00	0.13
	(3,510)	(3,152)	(3,154)		
Middle school grade point average (of 10)	8.00	8.23	8.23	0.00	0.50
	(0.85)	(0.82)	(0.77)		
Number of schools ranked	8.90	8.93	9.36	0.02	0.00
	(3.62)	(3.56)	(3.56)		
UNAM school as first choice	0.45	0.34	0.47	0.00	0.00
	(0.50)	(0.47)	(0.50)		
IPN school as first choice	0.15	0.19	0.18	0.00	0.01
	(0.35)	(0.39)	(0.39)		
Allocated to any choice during initial assignment	0.90	0.91	0.92	0.00	0.00
	(0.30)	(0.29)	(0.28)		
Rank of allocated school from initial assignment, if applicable	3.05	3.08	4.36	0.00	0.00
	(2.77)	(2.81)	(2.99)		
Comipems examination score	64.89	64.94	60.94	0.53	0.00
	(17.61)	(16.79)	(12.30)		
Enlace examination score		0.00	-0.14		0.00
		(1.00)	(0.87)		
Proportion taking Enlace at same school as assigned		0.93	0.92		

Note. Standard deviations in parentheses.
[a] A student is at a threshold if he either attended that school and would not attend it upon losing 5 or fewer points, or does not attend that school and would attend it upon gaining 5 or fewer points.
[b] Average 2005 exchange rate was 10.9 pesos/dollar.

Table 2. School system characteristics

	UNAM	**IPN**	**SE**	**DGETI**	**Colbach**	**Conalep**
Campuses (programs)	14 (14)	16 (16)	133 (193)	51 (188)	20 (40)	57 (176)
% of programs oversubscribed	100	100	79.3	68.0	62.5	31.3
Educational offering(s)	Academic	Tech-Academic	Academic, Technical	Tech-Academic	Academic	Technical
Admitted students	34,625	19,042	48,817	42,303	40,831	41,399
Mean Comipems score of admittees	84.4	87.2	61.6	58.2	62.2	51.1
Enlace examinees	0	9,622	26,442	15,082	12,751	13,842
Enlace examinees/admittees		0.51	0.54	0.36	0.31	0.33
Mean Enlace score		1.20	0.06	-0.37	0.03	-0.62
Mean family income (2005 pesos/month)[a]	6,175	5,432	3,732	3,714	4,330	3,382
Mean parents' years of education	11.9	11.3	9.4	9.5	10.0	9.0

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

Table 3. Supply and demand for schools, by system

	Students selecting as first choice	Total students assigned	First choice/total assigned
UNAM	103,013	34,625	2.98
IPN	34,056	19,042	1.79
SE, Estado de Mexico	38,778	48,817	0.79
DGETI	20,057	42,303	0.47
Colbach	17,033	41,399	0.41
Conalep	14,033	40,831	0.34
Others	3,457	3,410	1.01

Note. The first column is the total of all assigned students who chose a school from that system as their first choice. The second column is the number of students assigned to schools in that system, which is a lower bound for seats available in all systems except UNAM and IPN, since some schools may not have filled.

Table 4. Summary of regression discontinuity results

Panel A. All Schools

	Including IPNs	Excluding IPNs	
Average LATE	0.06***	0.03**	
Standard error	(0.01)	(0.01)	
Total observations	31,872	26,695	
R^2	0.47	0.39	

Panel B. IPN Schools

	Full sample	Rejected leave IPN	Rejected stay in IPN
Average LATE	0.19***	0.24***	0.17***
Standard error	(0.04)	(0.05)	(0.06)
Total observations	5,177	2,918	2,259
R^2	0.33	0.23	0.33

Note. Estimates are each from a single regression with a dummy variable for admission to the threshold school and threshold fixed effects interacted with the other covariates. Standard errors (in parentheses) are clustered by middle school attended.

Table 5. Regression discontinuity estimates for IPN schools

Panel A. School-specific characteristics and regression results

	Campus 9	Campus 13	Campus 3	Campus 6	Campus 5	Campus 14	Campus 15	Campus 12
School characteristics								
Cutoff score	99	88	86	83	81	80	78	76
Mean Comipems[a]	106.3	94.5	95.2	93.4	90.2	89.3	87.6	84.1
Enlace takers	330	789	636	896	524	347	280	621
Regression coefficients								
Admitted	0.41**	-0.06	0.17	0.29**	0.12	0.10	0.32*	0.19
	(0.16)	(0.15)	(0.14)	(0.14)	(0.17)	(0.24)	(0.18)	(0.13)
Comipems score	0.08**	0.04	0.08**	-0.01	0.03	0.01	0.04	0.03
	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)	(0.05)	(0.05)	(0.04)
Score x admitted	-0.06	0.08	-0.01	0.09*	-0.01	0.10	-0.10	0.00
	(0.06)	(0.05)	(0.05)	(0.05)	(0.06)	(0.07)	(0.07)	(0.05)
UNAM	—	0.25	0.29	0.22	0.26	0.66***	0.45***	0.45***
		(0.28)	(0.20)	(0.16)	(0.17)	(0.13)	(0.12)	(0.10)

	Campus 8	Campus 11	Campus 1	Campus 7	Campus 17	Campus 2	Campus 10	Campus 4
School characteristics								
Cutoff score	75	75	74	73	72	70	69	66
Mean Comipems	87.5	88.5	85.5	84.5	83.8	82.6	78	76.6
Enlace takers	676	853	518	633	558	739	584	638
Regression coefficients								
Admitted	0.06	0.02	0.42**	0.41***	0.22	0.24*	0.01	0.31**
	(0.21)	(0.17)	(0.19)	(0.13)	(0.17)	(0.15)	(0.10)	(0.13)
Comipems score	0.13***	0.05	-0.02	0.01	-0.02	0.03	0.06***	0.01
	(0.04)	(0.04)	(0.05)	(0.03)	(0.04)	(0.03)	(0.02)	(0.03)
Score x admitted	-0.09	-0.03	0.06	0.04	0.09*	0.08	-0.03	0.02
	(0.06)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)
UNAM	0.35	0.70***	0.36***	0.27***	0.43***	0.25**	0.38***	0.41***
	(0.23)	(0.15)	(0.10)	(0.07)	(0.11)	(0.10)	(0.05)	(0.06)

Note. Estimates are from a single regression with threshold fixed effects fully interacted with the covariates. Standard errors (in parentheses) are clustered by middle school attended. N = 5,177. $R^2 = 0.33$.

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

[a] Mean Comipems score is over all admitted students.

[b] "Admitted" is the estimated coefficient for being admitted to the specified school.

Table 6. IPN regression discontinuity results by parental education

	All students at IPN admission threshold			Students with 9 or fewer years of schooling			Students with more than 9 years of schooling		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Full sample	Rejected leave	Rejected stay	Full sample	Rejected leave	Rejected stay	Full sample	Rejected leave	Rejected stay
Average LATE	0.19***	0.24***	0.17***	0.16**	0.18**	0.17	0.21***	0.27***	0.19**
Standard error	(0.04)	(0.05)	(0.06)	(0.07)	(0.08)	(0.12)	(0.05)	(0.07)	(0.08)
Total observations	5,177	2,918	2,259	1,936	1,218	718	3,206	1,721	1,485
R^2	0.33	0.23	0.33	0.34	0.23	0.37	0.34	0.26	0.32

Note. Estimates are each from a single regression with a dummy variable for admission to the threshold school and threshold fixed effects interacted with the other covariates. Columns (4) through (9) weight observations so that each IPN school has the same weight as in the corresponding columns (1) through (3). Standard errors (in parentheses) are clustered by middle school attended.

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

Table 7. Attrition analysis for IPN schools

		Balance of observables at assignment			Balance of observables for Enlace-takers		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Attrition	Middle school GPA	Parents' education	Family income	Middle school GPA	Parents' education	Family income
Average LATE	0.05***	-0.02	0.09	0.09	0.12***	0.29	0.33*
Standard error	(0.02)	(0.03)	(0.12)	(0.14)	(0.04)	(0.19)	(0.19)
Total observations	11,850	11,802	10,773	10,671	5,103	4,739	4,704
R^2	0.03	0.07	0.04	0.03	0.11	0.05	0.04

Note. Dependent variables are below column numbers. Attrition is defined as being assigned to the school during the Comipems assignment process but not taking the Enlace. Estimates are each from a single regression with a dummy variable for admission to the threshold school and threshold fixed effects interacted with the other covariates. Standard errors (in parentheses) are clustered by middle school attended.
*** p<0.01, ** p<0.05, * p<0.1