Online appendices to accompany "Flourish or Fail? The Risky Reward of Elite High School Admission in Mexico City"

Andrew Dustan, Department of Economics, Vanderbilt University
Alain de Janvry and Elisabeth Sadoulet, Department of Agricultural and Resource Economics, University of California at Berkeley

Appendix A. Comparison of dropout and ENLACE score results with Estrada and Gignoux (2015)

In their paper estimating the effect of IPN admission on the expected returns to higher education, Estrada and Gignoux (2015) also present basic results on dropout and ENLACE scores.[1] 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 (2015) 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.

[1] Estrada, Ricardo, and Jeremie Gignoux. 2015. "Benefits to Elite Schools and the Formation of Expected Returns to Education: Evidence from Mexico City." Paris School of Economics Working Paper 2014-06.

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	3,184	2,928	5,125	10,990	6,914
Adjusted R-squared	0.003	0.004	0.012	0.015	0.018
Mean of dependent variable 1 point below cutoff	0.505	0.493	0.449	0.432	0.416

Panel B. Math ENLACE score

Dependent variable: Math score	Estrada & Gignoux sample (1)	Delete private middle schools (2)	Add State of Mexico middle schools (3)	Add 2006 COMIPEMS, 2010 ENLACE (4)	Selection method used in this paper (5)
Score ≥ cutoff	0.340***	0.331***	0.289***	0.215***	0.205***
	(0.0822)	(0.0853)	(0.0524)	(0.0414)	(0.0502)
Observations	1,570	1,439	2,678	5,978	3,781
Adjusted R-squared	0.097	0.101	0.140	0.132	0.180
Mean of dependent variable 1 point below cutoff	0.341	0.338	0.401	0.387	0.365

Panel C. Spanish ENLACE score

Dependent variable: Spanish score	Estrada & Gignoux sample (1)	Delete private middle schools (2)	Add State of Mexico middle schools (3)	Add 2006 COMIPEMS, 2010 ENLACE (4)	Selection method used in this paper (5)
Score ≥ cutoff	0.072	0.048	-0.017	-0.016	0.043
	(0.0742)	(0.0775)	(0.0616)	(0.0432)	(0.0484)
Observations	1,570	1,439	2,681	5,981	3,784
Adjusted R-squared	0.044	0.042	0.070	0.075	0.124
Mean of dependent variable 1 point below cutoff	0.138	0.133	0.188	0.182	0.131

Notes. 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 headers in columns 2-4 explain the changes made to the sample from the previous column. Standard errors clustered at the admitted school level are in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Appendix B. High school dropout in Mexico City and its relation to ENLACE-taking

This appendix uses school-level school census data and ENLACE registration and taking data, along with the student-level COMIPEMS and ENLACE databases, to evaluate the suitability of the ENLACE exam as a proxy for high school graduation. We show that this proxy is likely to perform well and that the most plausible challenges to its validity are not supported by the data.

I. Most dropout in Mexico City takes place before a student reaches twelfth grade.

Mexico's school census, the Formato 911, is completed by school personnel and submitted in the fall of each year (school begins in the fall in Mexico, as in the United States). It includes information about the number of students in each grade, broken down into students starting the grade and that and those who are repeating, and the number of graduates. Using these data, we can trace out the approximate continuation/dropout pattern separately in the IPN and non-elite subsystems. The results will differ slightly from those that could be obtained from student-level panel data on registration, but these aggregate rates are quite informative of the general pattern of dropout and graduation. Beginning with students enrolled in tenth grade in fall 2005, Table B1 presents the continuation rates.

The tenth and eleventh grade transition probabilities imply that, for the IPN, $0.825 * 0.842 = 69.5$ percent of tenth grade students made it to twelfth grade (and 30.5 percent drop out before reaching twelfth grade). Among all IPN twelfth graders, 83.8 percent graduated in the 2007-8 school year, but another 11.3 percent returned the next year to attempt to complete twelfth grade. If, of these repeaters, 83.8 percent eventually graduate, then $1 - (0.838 + 0.113 * 0.838) = 6.7$ percent of twelfth graders eventually leave high school without graduating. This implies that $0.067 * 0.695 = 4.7$ percent of tenth graders advance to twelfth grade and then drop out. Thus, of students who drop out, only $0.047/(0.305 + 0.047) = 13$ percent do so after advancing to twelfth grade.

The pattern is similar in non-elite COMIPEMS schools, except that a larger percentage of students drop out in the first year than in IPN schools. For these schools, $0.690 * 0.851 = 58.7$ percent of tenth graders make it to twelfth grade, and $1 - (0.829 + 0.094 * 0.829) = 9.3$ percent of twelfth graders eventually leave high school without graduating. Of students who drop out, $0.093/(0.413 + 0.093) = 18.4$ percent do so after advancing to twelfth grade, a rate somewhat higher than in the IPN.[2]

It is striking that so many twelfth graders who do not graduate actually return the next year. The ENLACE-taking microdata give evidence that this is actually true. Of students assigned to a non-UNAM school during the 2005 assignment process and who took the ENLACE during the 2008, 2009, or 2010, administrations, 86 percent did so in 2008, 10 percent in 2009, and 4 percent in 2010. While we cannot decompose the delayed taking, some of it is due to students who repeated twelfth grade one or more times, some is due to students who repeated an earlier grade, and some is due to students repeating multiple grades. Table B2 gives more detail on ENLACE-taking for the 2005 and 2006 COMIPEMS cohorts, separately by IPN and non-elite school assignment. It also sheds light on why attrition in our data is higher in 2006 than in 2005. Combining the IPN and non-elite students, 2.0 percent of students in the 2005 COMIPEMS cohort took the ENLACE

[2] The higher total dropout rate in the non-elite schools does not contradict our finding that marginally-admitted IPN students drop out at a higher rate. Marginally-admitted and marginally-rejected students attend subsets of schools that are different from the average schools in the IPN and non-elite sets (respectively). Marginal students also drop out at rates that are different than their school average dropout rate.

in 2010, five years after taking the COMIPEMS exam. Because the 2006 cohort only matches to ENLACE exams three and four years after taking the COMIPEMS exam, we can expect that about 2 percent of the additional attrition is due to the lack of an additional year of ENLACE results.[3]

II. The average ENLACE registration rate is 99 percent of all twelfth grade students.

The twelfth grade ENLACE is intended to be taken by students in their final "academic period" of high school, which is the semester for the schools being studied. At the time that school administrators register students for the exam, they are unsure of which students will have advanced to their final semester in the spring—that is, which will graduate at the end of the school year. Comparing school-level administrative data on the number of students registered to take the ENLACE with the number of twelfth graders reported in the school census in 2008 and 2009, we find that on average IPN schools register a number of students equal to 99.5 percent of their reported twelfth grade enrollment.[4] A comparison of the IPN mean registration rate with the non-elite mean, weighted by twelfth grade class size, is provided in Column 1 of Table B3. The estimated difference in means is 0.4 percent and is statistically insignificant.[5]

For the purposes of exploring the robustness of the RD results in the paper, we are more interested in whether students at the boundary of the IPN subsystem experience an increase or decrease in this school-level proportion if they are admitted to the IPN. To answer this question, we run the same RD model as in the paper, except that the outcome variable is now the school-level registration rate. Column 1 of Table B4 shows that, for students marginally admitted to the IPN subsystem, there is a statistically insignificant 0.9 percentage point increase in ENLACE registration rate. Given the high registration rate and the similarity of the rates between IPN and non-elite subsystems, there is little chance that the IPN is gaming the system by failing to register its weakest students, thus causing differential attrition in the ENLACE-taking data.

III. The average ENLACE taking rate is 98 percent of graduating twelfth graders.

Some students who have been registered to take the ENLACE do not graduate, and the probability of this outcome is known with much more certainty in the spring, when the exam is administered, than in the fall, when students are registered. The school-level data are consistent with, on average, students who graduate taking the exam and non-graduates not taking the exam.

Column 2 of Table B3 shows that, on average, schools administer the exam to 82 percent of students who were registered to take it, a figure that does not differ between IPN and non-elite schools. Column 3 confirms that, as implied in the first two columns, 81 percent of twelfth graders take the ENLACE. The striking result is in column 4, which shows that the ratio of ENLACE takers to high school graduates is 98 percent, which again is estimated to be nearly identical across school types. The correspondence between registration and twelfth grade enrollment, combined with the similarity between average exam-taking and graduation rates, provides evidence at the aggregate level that administrators register the population of potentially eligible takers in the fall and then allow anticipated graduates to take the exam in the spring.

---

[3] While Table 2 in the paper seems to suggest that attrition was about 3.9 percent higher in 2006, the true difference is 3.2 percent. The gap between these figures is due to small differences in the values of the covariates between years. Thus the unexplained difference in dropout rates between years is just 1.2 percent.

[4] To compute the average rate over the two years, we sum the numerator variable over the years, then sum the denominator over the years, and compute this ratio.

[5] The unweighted difference in means is 0.7 percent (SE = 0.6 percent).

Again, we are more interested in whether marginal admission to the IPN subsystem causes students to attend schools where these rates are significantly different. Table B4, columns 2 through 4, give RD estimates of IPN admission on ENLACE taking rates. The point estimates for each are small and insignificant, although we note that the point estimate for ENLACE takers as a proportion of graduates suggests that marginal IPN admits attend schools were this rate is 2.9 percentage points higher.

IV. Only 0.25 percent of ENLACE-taking students re-take the exam

While we cannot ascertain with these aggregated data how many non-graduates took the exam and how many graduates did not, the student-level ENLACE microdata show that very few students take the exam twice. Matching students in COMIPEMS across years 2008-2010 of the ENLACE database using their national ID number (CURP), we find that only 0.25 percent of observations correspond to students who took the ENLACE in multiple years. This rate is low both for students in IPN schools (0.06 percent) and in non-elite schools (0.27 percent). The effect that retaking can have on the paper's results are small. For example, it cannot be that some schools retain many twelfth graders, requiring them to take the exam in both years and thus increasing the ENLACE taking rate (and decreasing the implied dropout rate).

V. Because we exclude private middle school students from the sample, differential enrollment in private high schools has little effect on the results

Students residing inside the COMIPEMS boundaries attend three types of high schools:

1. COMIPEMS-participating schools, accounting for 72 percent of high school graduates in school years 2007-8 and 2008-9).[6] Recall that the UNAM schools do not administer the ENLACE.
2. A small number of public schools run by the Federal District government, accounting for only 1 percent of graduates. They do not administer the ENLACE.
3. Private schools, accounting for 27 percent of graduates. Some of these schools administer the ENLACE.

While 27 percent of graduates is not a small share for private schools, there are multiple reasons that transfers to private schools cannot explain the results found in the paper. First, note that this 27 percent figure is not the share of COMIPEMS participants who go to private school. Students planning to enroll in a private high school can do so without participating in COMIPEMS. Second, according to the school census, in the 2007-8 and 2008-9 school years, 66.4 percent of private school graduates in the COMIPEMS area graduated from schools that administered the ENLACE exam to their students. This implies that only $0.27 * (1 - 0.664) = 9.1$ percent of graduates in the COMIPEMS area do so from a private school that does not administer the ENLACE. Because we match COMIPEMS-takers to their ENLACE records using individual-specific information rather than school-specific records, we observe ENLACE scores for students graduating from participating private schools. Hence, in most cases, graduation from a private school does not result in attrition from our sample.

Furthermore, since we can observe the school at which COMIPEMS-takers took the ENLACE, it is possible to obtain a rough estimate of how many students did so at private high schools. Among public middle school COMIPEMS-takers assigned to an IPN or non-elite high school, only 2.1 percent took the ENLACE at a private school. Assuming that these 2.1 percent

[6] The statistics in this discussion come from the Formato 911 school census, already discussed in detail.

represent 66.4 percent of the private school graduates, then only 0.021 ∗ 0.664 = 3.2 percent of COMIPEMS-takers from public middle schools graduate from a private high school (1.1 percent from schools not participating in the ENLACE). On the other hand, 14.2 percent of COMIPEMS-takers from private middle schools who are assigned to an IPN or non-elite school graduate from a private high school. This implies that 0.142 ∗ 0.664 = 21 percent of this population takes the ENLACE at a private high school. Such a large proportion is the primary reason that private middle school students are excluded from the sample: many of them do not enter a public school.

Anecdotally, many private middle school students take the COMIPEMS exam hoping to enter a particular set of elite schools, and continue in private education if they fail to do so. This is another reason to exclude private middle school students: in the RD design, private middle school students marginally rejected from the IPN are much more likely to move to the private high school system. Table B5 shows the results of estimating the standard RD equation with a dummy for "Took the ENLACE in a private high school" as the dependent variable (where 0 can represent either taking in a public high school or not taking at all). The effect of admission for private middle school students is -0.086, implying that marginally rejected students in this group are much more likely to take the ENLACE in a private high school. Including these students results in an understatement of the IPN admission effect on ENLACE-taking because some of students must end up in private high schools that do not administer the ENLACE, meaning that they are counted as dropouts. On the other hand, the coefficient for public middle school students is just -0.012, implying a very small substitution toward private high schools upon marginal rejection. Again, this means that transfers to private schools are more likely to bias the dropout effect toward zero than away from it, although this implied bias in the sample of public middle school students is small.

VI. Marginal admission to non-elite schools does not increase dropout probability

One candidate explanation for the finding that IPN admission increases dropout probability is that schools discourage their worst students from taking the ENLACE exam, either intentionally or unintentionally. While the school-level ENLACE registration and taking rates are sufficiently high that this is unlikely, the student-level microdata provide further evidence on this point. If discouraging the worst students from taking the ENLACE is widespread, then we should see marginal admission increasing the dropout rate not only at the IPN/non-elite boundary, but also at the admissions cutoff of each oversubscribed school.[7] This is because marginal admission to any oversubscribed school implies being tied for the worst COMIPEMS score there, while marginal rejection may result in the student being assigned to a school where he is substantially higher in the COMIPEMS distribution. Admission may also affect ENLACE-taking probability through a true effect on dropout, of course, but by focusing on non-elite schools we can examine a context where admission does not result in exposure to a presumably much more difficult academic environment.

Table B6 shows the results from estimating the RD equation for all students who had a COMIPEMS score close to a non-elite school's admissions cutoff score, and for whom marginal rejection would result in assignment to a non-UNAM school (since UNAM schools do not administer the ENLACE).[8] On average, admission results in almost 1/2 of a standard deviation in

[7] Table 9 gives small, positive, but statistically insignificant effects of admission to a higher-cutoff IPN school on dropout probability.

[8] Note that some students are close to multiple cutoffs. In this case, there is one observation per student-cutoff pair.

the assigned school's mean COMIPEMS score, and column 2 shows that as a consequence admitted students fall from the median of the school-level score distribution to about the 7th percentile. Despite this fact, and the finding in column 5 that admitted students only travel on average 0.47 kilometers less to their assigned school, admission actually results in a slightly higher probability of taking the ENLACE. This is contrary to what we would expect to find in the case of marginally admitted students being discouraged from taking the ENLACE.

Table B1
Grade transition and graduation rates for selected years

School year	Transition	Rate (IPN schools)	Rate (Non-elite COMIPEMS schools)
2005-2006	10^{th} → begin 11^{th}	0.794	0.642
	10^{th} → repeat 10^{th}	0.031	0.048
	Total 10^{th} grade continuation	*0.825*	*0.690*
2006-2007	11^{th} → begin 12^{th}	0.726	0.774
	11^{th} → repeat 11^{th}	0.115	0.077
	Total 11^{th} grade continuation	*0.842*	*0.851*
2007-2008	12^{th} → graduate	0.838	0.829
	12^{th} → repeat 12^{th}	0.113	0.094

Notes. Rates are computed from the Formato 911 school census using the raw sum, over the given set of schools (IPN or non-elite COMIPEMS), of students in each initial category during the indicated school year, divided by the raw sum of students appearing in the transition categories in the following school year.

Table B2
Proportion of COMIPEMS-takers taking the ENLACE in each year

	Students assigned to IPN schools		Students assigned to non-elite schools	
	COMIPEMS 2005 cohort	COMIPEMS 2006 cohort	COMIPEMS 2005 cohort	COMIPEMS 2006 cohort
Proportion taking in 2008	0.542	0.000	0.450	0.000
Proportion taking in 2009	0.067	0.523	0.054	0.436
Proportion taking in 2010	0.033	0.075	0.018	0.056
Total	0.642	0.598	0.523	0.493

Notes. Proportions are computed from the matched student-level COMIPEMS-ENLACE data.

Table B3

Comparison of ENLACE registration and taking rates between IPN and non-elite schools

Dependent variable	ENLACE registrants/12th grade students	ENLACE takers/ ENLACE registrants	ENLACE takers/12th grade students	ENLACE takers/HS graduates
	(1)	(2)	(3)	(4)
IPN school	0.004	0.002	0.007	0.008
	(0.0085)	(0.0176)	(0.0166)	(0.0235)
Constant (Non-elite school mean)	0.991***	0.820***	0.810***	0.977***
	(0.0075)	(0.0108)	(0.0106)	(0.0139)
Adjusted R-squared	-0.003	-0.003	-0.003	-0.003
Observations	287	287	287	287

Notes. Estimates are from weighted linear regressions, where weights are the number of 12th grade students in the school. Huber-White robust standard errors are in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table B4
Regression discontinuity estimates of effect of IPN admission on ENLACE-taking proportion at assigned school

Dependent variable	ENLACE registrants/ 12th grade students at assigned school	ENLACE takers/ ENLACE registrants at assigned school	ENLACE takers/12th grade students at assigned school	ENLACE takers/HS graduates at assigned school
	(1)	(2)	(3)	(4)
Score ≥ cutoff	0.009	-0.003	0.005	0.029
	(0.0083)	(0.0108)	(0.0103)	(0.0180)
Observations	13,518	12,229	16,090	13,521
Adjusted R-squared	0.057	0.181	0.118	0.137
Mean of dependent variable 1 point below cutoff	0.990	0.811	0.802	0.955
Bandwidth	10.0	9.0	11.5	9.9

Notes. Estimates are from local linear regressions, including separate linear terms for each of the 16 IPN cutoff schools and cutoff school-year fixed effects. The edge kernel is used in each regression and in computation of the corresponding optimal Imbens-Kalyanaraman bandwidth. Standard errors clustered at the admitted high school level are in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table B5
Regression discontinuity estimates of effect of IPN admission on taking ENLACE at a private high school

Sample	Public middle school students	Private middle school students
Dependent variable	Took ENLACE at a private high school	Took ENLACE at a private high school
	(1)	(2)
Score ≥ cutoff	-0.012***	-0.086**
	(0.0045)	(0.0363)
	[0.00]	[0.02]
Observations	21,095	1,695
Adjusted R-squared	0.005	0.042
Mean of dependent variable 1 point below cutoff	0.027	0.114
Bandwidth	16.3	15.8

Notes. Estimates are from local linear regressions, including separate linear terms for each of the 16 IPN cutoff schools and cutoff school-year fixed effects. The edge kernel is used in each regression and in computation of the corresponding optimal Imbens-Kalyanaraman bandwidth. Standard errors clustered at the admitted high school level are in parentheses. Wild cluster bootstrapped p-values, clustered at the centered COMIPEMS score level, are in brackets. * p<0.10, ** p<0.05, *** p<0.01.

Table B6

Regression discontinuity estimates of effect of admission on school characteristics and student outcomes, all oversubscribed non-elite schools

Dependent variable	Mean COMIPEMS score	Student location in school's COMIPEMS distribution (quantile)	Mean middle school GPA	Mean parental education (years)	Distance from home to school (km)	Dropout (not taking ENLACE exam)	ENLACE math score	ENLACE Spanish score
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Score ≥ cutoff	7.633***	-0.414***	0.093***	0.332***	-0.473***	-0.022**	-0.002	-0.001
	(0.2208)	(0.0083)	(0.0068)	(0.0262)	(0.0931)	(0.0089)	(0.0087)	(0.0086)
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.84]	[0.94]
Observations	114,483	46,792	136,249	114,483	206,739	253,495	157,577	157,699
Adjusted R-squared	0.810	0.660	0.705	0.674	0.055	0.049	0.215	0.219
Mean of dependent variable 1 point below cutoff	57.608	0.480	7.842	9.691	6.882	0.531	-0.367	-0.242
Bandwidth	5.2	1.8	5.7	5.1	9.7	10.7	15.0	15.1

Notes. Estimates are from local linear regressions, including piecewise-linear terms in COMIPEMS score and cutoff school-year 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, and 4 are mean characteristics of all students admitted to the student's admitted school in his admission year. Standard errors clustered at the admitted high school level are in parentheses. Wild cluster bootstrapped p-values, clustered at centered COMIPEMS score level, are in brackets. * p<0.10, ** p<0.05, *** p<0.01.

Appendix C. Additional figures and tables

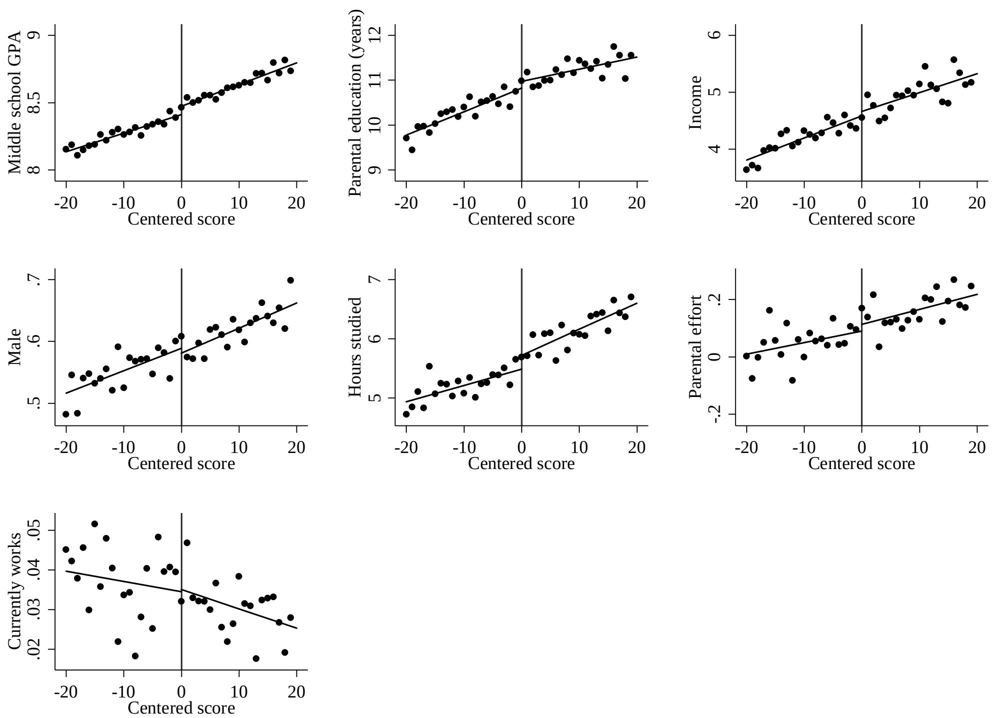

Figure C1
Balance of covariates with respect to IPN admission, after dropout
Notes. Dependent variables indicated on the vertical axes. Plots are for students belonging to the regression discontinuity sample defined in the text.

Table C1

Regression discontinuity estimates of effect of IPN admission on dropout and ENLACE scores, allowing students with feasible UNAM assignments

Dependent variable	Dropout (not taking ENLACE exam)			ENLACE math score			ENLACE Spanish score		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Score ≥ cutoff	0.074***	0.090***	0.095***	0.223***	0.208***	0.226***	0.017	0.025	0.037
	(0.0234)	(0.0185)	(0.0163)	(0.0480)	(0.0403)	(0.0368)	(0.0499)	(0.0383)	(0.0349)
	[0.02]	[0.00]	[0.00]	[0.01]	[0.00]	[0.00]	[0.54]	[0.22]	[0.01]
Observations	9,246	15,260	21,340	5,044	8,449	12,025	5,047	8,454	12,033
Adjusted R-squared	0.017	0.016	0.014	0.125	0.184	0.236	0.085	0.126	0.151
Mean of dependent variable 1 point below cutoff	0.423	0.417	0.415	0.299	0.295	0.292	0.177	0.148	0.131
Bandwidth	5.0	10.0	15.0	5.0	10.0	15.0	5.0	10.0	15.0

Notes. Estimates are from local linear regressions, including separate linear terms for each of the 16 IPN cutoff schools and cutoff school-year fixed effects. The edge kernel is used in each regression. Sample selection allows students who could be assigned to UNAM high schools for centered scores outside of the fixed bandwidth, as well as assignments to non-elite high schools for centered scores higher than the upper bound given by the bandwidth. Standard errors clustered at the admitted high school level are in parentheses. Wild cluster bootstrapped p-values, clustered at the centered COMIPEMS score level, are in brackets. * p<0.10, ** p<0.05, *** p<0.01.

Table C2
Regression discontinuity estimates of effect of IPN admission on dropout and ENLACE scores, rectangular kernel

Dependent variable	Dropout (not taking ENLACE exam)	ENLACE math score	ENLACE Spanish score	Dropout (not taking ENLACE exam)	ENLACE math score	ENLACE Spanish score	Dropout (not taking ENLACE exam)	ENLACE math score	ENLACE Spanish score
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Score ≥ cutoff	0.091***	0.237***	0.045	0.088***	0.227***	0.042	0.077***	0.205***	0.050
	(0.0177)	(0.0372)	(0.0370)	(0.0199)	(0.0409)	(0.0404)	(0.0247)	(0.0579)	(0.0532)
	[0.00]	[0.00]	[0.03]	[0.00]	[0.00]	[0.25]	[0.07]	[0.10]	[0.44]
Observations	16,476	9,227	8,508	11,122	6,183	5,417	5,550	3,055	3,058
Adjusted R-squared	0.013	0.268	0.160	0.013	0.218	0.135	0.020	0.176	0.124
Mean of dependent variable 1 point below cutoff	0.416	0.284	0.117	0.416	0.284	0.117	0.416	0.284	0.117
Bandwidth	12.0	12.3	11.1	8.0	8.2	7.4	4.0	4.1	3.7
Lee bound (upper)		0.357***	0.158***		0.341***	0.150***		0.308***	0.138*
		(0.0402)	(0.0421)		(0.0504)	(0.0530)		(0.0651)	(0.0717)
Lee bound (lower)		0.115***	-0.117***		0.105**	-0.120***		0.091	-0.084
		(0.0387)	(0.0413)		(0.0433)	(0.0444)		(0.0668)	(0.0820)

Notes. Estimates are from local linear regressions, including separate linear terms for each of the 16 IPN cutoff schools and cutoff school-year fixed effects. The rectangular kernel is used in each regression and in computation of the corresponding optimal Imbens-Kalyanaraman bandwidth. Columns 1-3 use the optimal bandwidth, while columns 4-6 and 7-9 use 2/3 and 1/3 of the optimal bandwidth, respectively. Standard errors clustered at the admitted high school level are in parentheses. Wild cluster bootstrapped p-values, clustered at the centered COMIPEMS score level, are in brackets. * p<0.10, ** p<0.05, *** p<0.01.

Table C3
Regression discontinuity estimates of effect of IPN admission on dropout and ENLACE scores, students with technical high school school below cutoff

	Dropout (not taking ENLACE exam)	ENLACE math score	ENLACE Spanish score
	(1)	(2)	(3)
Score ≥ cutoff	0.132***	0.327***	0.153***
	(0.0204)	(0.0411)	(0.0498)
	[0.00]	[0.00]	[0.00]
Observations	10,847	6,032	4,382
Adjusted R-squared	0.016	0.336	0.188
Mean of dependent variable 1 point below cutoff	0.402	0.319	0.039
Bandwidth	21.2	20.4	13.5

Notes. Sample is limited to students whose assigned school for a score one point below the IPN admission cutoff is to a "bachillerato tecnológico" (technical high school), the same category of school as the IPN campuses. Estimates are from local linear regressions, including separate linear terms for each of the 16 IPN cutoff schools and cutoff school-year fixed effects. The edge kernel is used in each regression and in computation of the corresponding optimal Imbens-Kalyanaraman bandwidth. Standard errors clustered at the admitted high school level are in parentheses. Wild cluster bootstrapped p-values, clustered at the centered COMIPEMS score level, are in brackets. * p<0.10, ** p<0.05, *** p<0.01.

Table C4
Regression discontinuity estimates of heterogeneous effects of admission on dropout, all oversubscribed non-elite schools

Dependent variable	Dropout (not taking ENLACE exam)	ENLACE math score	ENLACE Spanish score
	(1)	(2)	(3)
Score ≥ cutoff	-0.012*	0.005	0.005
	(0.0063)	(0.0103)	(0.0114)
	[0.03]	[0.54]	[0.58]
(Score ≥ cutoff) * (Change in commute due to admission)	0.004***	0.000	-0.001
	(0.0011)	(0.0019)	(0.0023)
	[0.03]	[0.97]	[0.66]
Observations	161,964	117,132	111,624
Adjusted R-squared	0.053	0.219	0.223
Mean of dependent variable 1 point below cutoff	0.513	-0.327	-0.208
Bandwidth	10.2	15.7	14.8

Notes. Estimates are from local linear regressions, including piecewise-linear terms in COMIPEMS score and cutoff school-year 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, and 4 are mean characteristics of all students admitted to the student's admitted school in his admission year. Standard errors clustered at the admitted high school level are in parentheses. Wild cluster bootstrapped p-values, clustered at the centered COMIPEMS score level, are in brackets. * p<0.10, ** p<0.05, *** p<0.01.

Table C5
Regression discontinuity estimates of effect of IPN admission on dropout and ENLACE scores, restricted to students near core of Federal District

Dependent variable	Dropout (not taking ENLACE exam)	ENLACE math score	ENLACE Spanish score
	(1)	(2)	(3)
Score ≥ cutoff	0.088***	0.197***	0.034
	(0.0244)	(0.0399)	(0.0471)
	[0.01]	[0.00]	[0.34]
Observations	8,429	7,086	5,250
Adjusted R-squared	0.009	0.373	0.189
Mean of dependent variable 1 point below cutoff	0.492	0.117	0.021
Bandwidth	17.2	30.7	21.2

Notes. Sample is limited to students residing in the boroughs constituting the "core" of the Federal District, where 14 of the 16 IPN schools are located (Álvaro Obregón, Azcapotzalco, Benito Juarez, Coyoacán, Cuauhtémoc, Gustavo A. Madero, Iztacalco, Iztapalapa, Miguel Hidalgo, and Venustiano Carranza). Estimates are from local linear regressions, including separate linear terms for each of the 16 IPN cutoff schools and cutoff school-year fixed effects. The edge kernel is used in each regression and in computation of the corresponding optimal Imbens-Kalyanaraman bandwidth. Standard errors clustered at the admitted high school level are in parentheses. Wild cluster bootstrapped p-values, clustered at the centered COMIPEMS score level, are in brackets. * p<0.10, ** p<0.05, *** p<0.01.

Table C6

Regression discontinuity estimates of effect of IPN admission on dropout and ENLACE scores, further heterogeneity results

Dependent variable	Dropout (not taking ENLACE exam)				ENLACE math score				ENLACE Spanish score			
Heterogeneity variable	Family income	Hours studied per week	Mean MS GPA	Mean MS COMIPEMS score	Family income	Hours studied per week	MS mean GPA	MS mean COMIPEMS score	Family income	Hours studied per week	Mean MS GPA	Mean MS COMIPEMS score
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Score ≥ cutoff	0.100***	0.093***	0.094***	0.094***	0.257***	0.259***	0.244***	0.247***	0.043	0.044	0.044	0.046
	(0.0164)	(0.0170)	(0.0166)	(0.0168)	(0.0345)	(0.0320)	(0.0355)	(0.0356)	(0.0375)	(0.0364)	(0.0343)	(0.0339)
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.02]	[0.03]	[0.04]	[0.04]
	-0.001	-0.006	-0.026	-0.001	0.003	-0.010	0.035	-0.002	0.023*	0.002	0.159	0.006
(Score ≥ cutoff) * (Heterogeneity variable)	(0.0049)	(0.0046)	(0.0653)	(0.0023)	(0.0084)	(0.0077)	(0.1316)	(0.0049)	(0.0134)	(0.0103)	(0.1372)	(0.0058)
	[0.84]	[0.35]	[0.76]	[0.73]	[0.81]	[0.21]	[0.80]	[0.58]	[0.06]	[0.84]	[0.07]	[0.15]
Observations	19,302	18,400	18,400	20,281	20,281	11,582	11,582	13,483	12,115	12,115	12,115	9,645
Adjusted R-squared	0.014	0.018	0.018	0.017	0.016	0.267	0.267	0.291	0.252	0.252	0.252	0.156
Mean of dependent variable 1 point below cutoff	0.401	0.409	0.409	0.416	0.416	0.262	0.262	0.271	0.284	0.284	0.284	0.125
Bandwidth	16.3	14.6	14.6	15.3	15.3	17.5	17.5	19.7	15.6	15.6	15.6	13.8

Notes. Estimates are from local linear regressions, including separate linear terms for each of the 16 IPN cutoff schools and cutoff school-year fixed effects. The edge kernel is used in each regression. Sample selection allows students who could be assigned to UNAM high schools for centered scores outside of the fixed bandwidth, as well as assignments to non-elite high schools for centered scores higher than the upper bound given by the bandwidth. Standard errors clustered at the admitted high school level are in parentheses. Wild cluster bootstrapped p-values, clustered at the centered COMIPEMS score level, are in brackets. * p<0.10, ** p<0.05, *** p<0.01.

Table C7

Regression discontinuity estimates of effects of IPN admission on dropout and ENLACE scores, accounting for student preferences

Dependent variable	Dropout (not taking ENLACE exam)	Dropout (not taking ENLACE exam)	ENLACE math score	ENLACE math score	ENLACE Spanish score	ENLACE Spanish score
	(1)	(2)	(3)	(4)	(5)	(6)
Score ≥ cutoff	0.088***	(One per cutoff)	0.250***	(One per cutoff)	0.062	(One per cutoff)
	(0.0171)		(0.0357)		(0.0400)	
	[0.00]		[0.00]		[0.06]	
		0.044		0.103*		0.059
(Score ≥ cutoff) * (Cutoff IPN school was student's first choice)		(0.0416)		(0.0587)		(0.0774)
		[0.50]		[0.37]		[0.47]
Fixed effects, all combinations of first two school choices	YES	YES	YES	YES	YES	YES
Controls for cutoff scores of first five choices	YES	YES	YES	YES	YES	YES
Observations	20,281	20,281	12,115	12,115	10,693	10,693
Adjusted R-squared	0.036	0.018	0.279	0.255	0.174	0.156
Mean of dependent variable 1 point below cutoff	0.416	0.416	0.284	0.284	0.117	0.117
Bandwidth	15.3	15.3	15.6	15.6	14.1	14.1

Notes. Estimates are from local linear regressions, including separate linear terms for each of the 16 IPN cutoff schools and cutoff school-year fixed effects. Columns 2, 4, and 6 estimate one admission effect per cutoff school and include piecewise-linear terms in centered COMIPEMS score interacted with the first "cutoff IPN school was student's first choice" dummy in each column, as well as the dummy itself. The edge kernel is used in each regression and in computation of the corresponding optimal Imbens-Kalyanaraman bandwidths. Standard errors clustered at the admitted high school level are in parentheses. Wild cluster bootstrapped p-values, clustered at the centered COMIPEMS score level, are in brackets. * p<0.10, ** p<0.05, *** p<0.01.