Does School Quality Matter?: Evidence from a Natural Experiment in Rural China

Albert Park, University of Oxford

Xinzheng Shi, Tsinghua University

Chang-tai Hsieh, Chicago Graduate School of Business

Xuehui An, National Center for Education Development Research, China Ministry of Education

Preliminary draft: do not cite without permission of authors

January 2010

Abstract

This paper investigates the impact of high school quality on students' educational attainment using a regression discontinuity research design that compares students just above and below entrance examination score thresholds that strictly determine admission to the best high schools in China's rural counties. Using data from Gansu Province in China, we find that attending the best high school in one's county of residence increases college entrance examination scores by 0.21 standard deviations and increases the probability of qualifying for college admission by 18.7 percentage points conditional on taking the college entrance exam, but decreases the probability of taking a college entrance examination by 9.3 percentage points.

1. Introduction

One of the most fundamental questions in the economics of education is the extent to which school quality affects educational attainment. This is an especially critical question in developing countries, where budgetary resources are scarce and governments must trade off improvements in educational quality with expanding access to education. Most empirical studies of the effects of various measures of school quality (e.g., educational spending, teacher quality) are difficult to interpret because students, teachers, and resources are almost never randomly allocated among schools.[1] Therefore, there are almost always unobserved factors correlated with both school quality and student attributes.

A few studies have attempted to address the endogeneity problem using different approaches, such as comparison with matched control groups (Angrist and Lavy, 2001; Rivkin, Hanushek and Kain, 2005; Rockoff, 2004), conducting randomized trials to examine the impact of specific schooling inputs (Glewwe, Kremer and Moulin, 2007; Banerjee, Cole, Duflo, and Linden, 2006; Dee, 2001), and exploiting natural experiments that create plausibly exogenous variation in class size (Angrist and Lavy, 1999; Hoxby, 2000) or in the quality of schools attended, e.g., lotteries (Gould, Lavy and Paserman, 2004; Hoxby 2007). However, these studies have weaknesses. For example, studies

[1] Although some research has found no significant link between educational attainment, usually measured by test scores, and school quality, measured by teachers' education level, school resources, equipment, class size, etc. (Altonji, Elder and Taber, 2005; Hoxby, 2000), other research has found a significantly positive relationship (Angrist and Lavy, 1999, 2001; Case and Deaton, 1999; Dee, 2001; Eide and Showalter, 1998; Goldhaber and Brewer, 1997; Gould, Lavy and Paserman, 2004; Greenwald, Hedges and Laine, 1996; Krueger, 1999; Krueger and Whitmore, 1999; Lai, Sadoulet and Janvry, 2006; Lavy, 2002; Newhouse and Beegle, 2005; Rivkin, Hanushek and Kain, 2005; Rockoff, 2004; Urquiola, 2006). School quality also has been found to be correlated with students' subsequent economic outcomes (Altonji and Dunn, 1996; Betts, 1995, 2001; Card and Krueger, 1992a).

using matched control groups do not convincingly control for unobservable student differences, and many studies focus narrowly on one dimension of schooling quality even though other dimensions may be more important.

In this paper, we present new evidence from a natural experiment in China. We quantify the impact of school quality on students' educational attainment using a regression discontinuity research design that exploits the fact that in many of China's rural counties admission to the best high schools are strictly determined by entrance examination scores. In addition to informing general debates over the importance of school quality, because many countries in Asia (e.g., Taiwan, Hong Kong, Singapore, and Korea) employ a similar mechanism for allocating students to schools, our results may help such countries evaluate the benefits of maintaining merit-based school assignment systems rather than location- or neighborhood-based systems.

In China, many rural counties operate a magnet school system for high schools. Typically, a rural county has one magnet school located in the county seat and a number of regular high schools spread across the county, and students resident in a given county attend one of the county's public high schools. Graduating middle school students must take county-wide high school entrance examinations, which determine whether they are eligible to attend the magnet high school, a regular high school, or no high school at all. In any given county, the magnet school is widely viewed to have the best quality and has the highest entrance examination score cutoff line.

This paper uses a regression discontinuity design that compares students with nearly identical entrance exams who attend different quality schools because they are just above or just below the cutoff score for admission to the magnet school. Our main

outcome measure is student scores on uniform national college entrance examinations taken at the end of high school. Because these scores strictly determine eligibility to attend college, preparation for the college entrance examination is the focus of teaching effort in China and so is an excellent summary statistic for assessing school quality in the Chinese context. Using information on the cutoff lines for college admission, we can also examine whether attending a magnet school affects the probability of qualifying for college admission. In addition, we can test whether magnet school attendance influences whether students actually take a college entrance examination. Using data from seven counties in Gansu province in China, we find that for students with entrance scores near the cutoff line entering a magnet high school increases students' college entrance exam scores by 0.21 standard deviations, increases students' probability of entering college by 18.7 percentage points conditional on taking the college entrance exam, and decreases the probability of taking the college entrance exam by 9.3 percentage points. For reasons explained below, these are likely to be lower bound estimates of the benefits of attending a magnet school.

The rest of the paper is organized as follows. Section 2 describes the data and variable definitions. Section 3 presents the methodology for implementing the regression discontinuity design and describes the empirical specification. Section 4 describes students' assignment to different schools in the sample used for analysis. Section 5 presents the empirical results, and Section 6 describes the differences between magnet schools and regular schools that might contribute differences in educational attainment. Section 7 extends the analysis in several directions, and Section 8 concludes.

2. Data and variables

2.1 Data collection

The data used in this paper was collected from high schools in rural counties in Gansu Province in western China during the summer of 2004.[2] Gansu is one of China's poorest provinces, with a GDP per capita of $744 in 2004 which ranked 30th among China's 31 provinces (National Bureau of Statistics, 2005). It had a population of 26 million in 2004, ranking 22nd among Chinese provinces (National Bureau of Statistics, 2005).

The main sample used for analysis includes 11,448 students in the entering classes in the years 1997 to 2001 in 39 high schools located in seven rural counties. Data is missing for some counties in some years, so the main sample covers 13 county-years.[3] China's high schools have three grade levels, so all students in the sample had completed high school and taken college entrance examinations by the time of the survey.

Given our identification strategy, we focus on county-years in which the cutoff lines are used to determine admission to the magnet school. County education bureaus have discretion in determining how students are allocated to schools, and some do not adhere strictly to cutoff lines. For example, in field visits we learned that some magnet schools allow students with poorer scores to pay additional fees in order to enroll, a practice which has become more prevalent over time. It is also possible that personal connections to principals or education officials influence admissions decisions. Some counties lack a clearly identifiable magnet school, because the top two schools are of

[2] In a few counties, data was provided by county educational bureaus.

[3] Data was initially collected from high schools in 19 counties (and 74 county-years). We exclude 37 county-years with missing data that preclude analysis. For 21 county-years we have information on only one school, and for 12 county-years we lack data on entrance score cutoff lines of the school with the highest mean entrance scores. Two county-years have more than 45% of students with missing information on gender; after these students are dropped; information for only one school remains, so we drop these two county-years.

similar quality or because due to large geographic distances most students choose to study at nearby schools.

We define a magnet high school to be the high school in a given county in a given year with the highest average student entrance score and the highest high school entrance score cutoff line for enrollment. We exclude 11 county-years for which the percentage of students entering the magnet high school with entrance scores lower than the school's cutoff line is greater than 30 percent. We also exclude 4 county-years for which the percentage of students attending regular high schools with high school entrance scores greater than the magnet school cutoff line is greater than 70 percent. We also drop 4 county-years for which a magnet school cannot be clearly identified because the school with the highest mean entrance score is different than the school with the highest entrance score cutoff line.

We observe the high school entrance exam score for nearly 100 percent of students in the sample but only for 62 percent of those with high school entrance exam scores do we have data on the student's college entrance examination score. Missing data on college entrance examination scores can be due to multiple reasons: the student could have dropped out or transferred to another school, or decided not to sit for the college entrance examination; or the school could have kept incomplete records. In one school we visited, college entrance examination scores had been kept only for those who had scored high enough to enter college. To avoid bias from the strong selection bias associated with such cases, we exclude 6 county-years for which the percentage of students having college entrance scores is very close to the percentage of students

entering colleges. We are left with a sample of 11,448 students for 13 county-years. Summary statistics for these 13 county-years is reported in Table 1.

The survey collected school administrative data on students' gender, birth year, year of high school entrance, high school entrance exam score, and college entrance examination score. The survey also collected data from schools on the high school entrance exam score cutoff line and school characteristics such as teachers' education levels and the availability and quality of different types of school facilities.

2.2 Variable definitions

Two treatment variables are defined. The variable *magnet* is assigned to equal one if the student actually attended a magnet school as defined above. The other treatment variable *eligible* is assigned to equal one if the student's high school entrance score is higher than the high school entrance exam cutoff line of the magnet school. While *magnet* more accurately reflects whether students actually attended better schools, it is subject to selection bias if eligible students do not attend the magnet schools or ineligible students do attend the magnet school. With such violations of the assignment rule, the coefficient on *eligible* can be viewed as a lower bound estimate of the true treatment effect. Our preferred specification is one in which *eligible* is used as an instrument for *magnet*.

The high school entrance examination scores and college entrance examination scores are the key variables used in the analysis. Scores are normalized to be equal to the number of standard deviations from mean scores of students taking the same examination. Specifically, define the normalized high school entrance examination score as follows:

$$\widehat{HS}_{ijt} = \frac{HS_{ijt} - \overline{HS_{jt}}}{HSSD_{jt}} \quad (1)$$

Here, HS_{ijt} is the high school entrance exam score for student *i* in county *j* who entered high school in year *t*, $\overline{HS_{jt}}$ is the mean high school entrance exam score for students in county *j* who entered high school in year *t*, and $HSSD_{jt}$ is the standard deviation of high school entrance exam scores among students in county *j* who began high school in year *t*.

$$\widehat{CS}_{ipt} = \frac{CS_{ipt} - \overline{CS}_{pt}}{CSSD_{pt}} \quad (2)$$

College entrance examination scores are similarly defined in Equation 2. The only difference is that the subscript *j* is replaced with subscript *p* for the college exam type, which in China can be liberal arts, science, physical education, musical education, or arts education. College entrance examinations differ from high school exams in two respects: they are nationally standardized and they are specialized by subject area. Thus, CS_{ipt} is the college entrance examination score for student *i* entering high school in year *t* taking test in subject *p*.

Another educational attainment variable is an indicator variable for whether the student is qualified to attend college. We compare the student's college entrance exam score with the lowest college admission cutoff lines in Gansu Province in the year the student took the college entrance exam in order to determine whether the student qualified for college.[3] This measure is similar to the eligibility variable in that it is not subject to selection biases associated with the student's decision to actually attend college conditional on his or her entrance exam score being above the cutoff line. Such decisions

[3] The cutoff lines from different provinces come from http://www.eol.cn/include/cer.net/gaokao/zhuanti/2006_fenshuxian.shtml#2000. There are different lowest cutoff lines for different types of college entrance exams.

could be influenced by credit constraints, family income and wealth, parental expectations, and other factors that could be correlated with learning outcomes.

In this paper, student age at the time of entering high school is first calculated by subtracting students' birth year from the year in which they entered high school. Unfortunately, there are a number of missing values for birth year. In such cases, we impute student age by setting missing values to be equal to the average student age in the same entering class in the same school in the same year.[4]

Table 2 gives summary statistics for the variables used in the analysis. The upper panel is for all students used in the analysis. Thirty-seven percent of students are female, and the average age is 15.79. Forty-four percent of students in the sample actually attend magnet schools, while 50 percent are eligible to attend magnet schools. The bottom panel shows the statistics for students whose college entrance scores are not missing. Most values are similar to the full sample. Thirty-five percent of students are female, mean age is 15.80, 49 percent of students attend magnet schools, and 56 percent are eligible to attend magnet schools. About 50 percent of students with college examination scores qualify for college entrance. Thus, the percentages of students entering magnet schools or being eligible to attend magnet schools are 6 percentage points higher in the smaller sample, suggesting that better students are more likely to take the college entrance examination.

3. Methodology

[4] In the few cases in which age is missing for all students in the school, age is imputed to be equal to the mean student age of all students entering high school in the same county in the same year. And if data is missing for all students in the county-year, then age is imputed to equal the average student age for all students in the sample entering high school in the same year.

3.1. Regression discontinuity design

We employ a regression discontinuity design (RDD) to quantify the impact of school quality on educational attainment. First developed by Thistlethwaite and Campbell (1960), in recent years there has been renewed interest in RDD identification issues (Hahn, Todd and van der Klaauw, 2001; Lee, 2004), estimation issues (Porter, 2003), and applications (Angrist and Lavy, 1999; van der Klaauw, 2002).

The basic regression model we employ is the following:

$$Y_i = \beta + \alpha T_i + u_i \text{ , where } T_i = 1\{S_i \geq \bar{S}\} \quad (3)$$

Here, Y_i is the outcome variable, and T_i is the treatment variable, which equals one if treated and zero otherwise. Those whose high school entrance exam score S_i falls below some distinct cutoff point $\bar{S}$ are placed in the control group (T_i=0), whereas those on or above the point are placed in the treatment group (T_i=1).

In regression discontinuity design, under the assumption that the conditional mean function E[u|S] is continuous at $\bar{S}$, the treatment effect α can be identified as follows:

$$\alpha = \lim_{S \downarrow \bar{S}} E[Y|S] - \lim_{S \uparrow \bar{S}} E[Y|S] \quad (4)$$

Intuitively, the treatment effects are identified by the sample of individuals within a very small interval around the cutoff point. Since these individuals have essentially the same S_i value, we can expect individuals just below the cutoff line on average to be very similar to individuals just above the cutoff line and thus to have similar average outcomes regardless of treatment status, thus providing credible estimates of the true treatment effect.

In the model, if S_i is correlated with the outcome, Y_i, then T_i would be related to Y_i as well. This could cause OLS estimates of α to be inconsistent. One approach proposed

by Heckman and Robb (1985) is to specify a conditional mean function E[u| T, S] as a "control function" in the outcome equation. Therefore, in practice, the following equation is estimated:

$$Y_i = \beta + \alpha T_i + k(S_i) + \omega_i \quad (5)$$

As long as $k(S_i)$ is continuous in S_i, identification is achieved because of the discontinuity in the function $T(S_i)$. In this paper, $k(S_i)$ is approximated by a 4th order polynomial function of the high school entrance exam score.

Lee (2003) proposes a direct test of the continuity assumption by checking whether there are discontinuities in the relationship between the treatment effect and any predetermined characteristics. That is, the following equation can be estimated as a pre-assumption test:

$$X_i = \delta + \varphi T_i + k(S_i) + \varepsilon_i \quad (6)$$

If φ is not statistically significant, then the continuity assumption is valid. In this paper, student age and gender are the only two predetermined characteristics that can be tested.

3.2. Fuzzy regression discontinuity design and IV estimation

In the regression discontinuity design, treatment depends on the selection variable S in a deterministic way. However, in reality, it is likely for treatment assignment to depend on S in a stochastic manner, which is referred to in the literature as fuzzy regression discontinuity design. In this case, OLS estimates of equation (3) may be biased.

In our context, even after we apply an exclusion criteria to ensure that violations of the eligibility rule are less than 30 percent in the magnet school and exclude county-years

for which a high percentage of regular students have high school entrance scores above the cutoff line, there are still students in magnet schools whose high school entrance exam scores are below the cutoff line and students in regular schools whose scores are above the cutoff line. In this case, the OLS estimate of α in equation (3) using the variable *magnet* could be subject to selection bias. In our main sample, 18.5 percent of students not in magnet schools have high school entrance examination scores above the cutoff line, and 10.2 percent of students in magnet schools have high school entrance examination scores below the cutoff.

This is where the second treatment variable *eligible* can help avoid the problems associated with bias caused by fuzzy discontinuity design. The variable *eligible* itself does not suffer from fuzziness and so can be used to cleanly estimate a treatment effect. However, the impact of eligibility is not of primary interest; we would like to estimate the impact of actually attending better schools. To obtain an unbiased estimate of the impact of attending magnet schools rather than regular schools, we thus use *eligible* as an instrument for *magnet*.

4. Student assignment

Students are assigned to magnet schools and regular schools according to their high school entrance examination score. Figure 1 shows the distribution of students with different high school entrance scores in magnet and regular schools in the seven counties. In each panel, on the y-axis, a "one" indicates that the student entered a magnet school; while a "zero" indicates that the student attended a regular school. In order to pool data from different years for each county, we create a variable that indicates each student's

score relative to the entrance cutoff score in each county-year, which is shown on the x-axis. Each graph is a nonparametric prediction using a local polynomial smoother (with degree equal to zero and bandwidth equal to 0.25). These figures highlight the fact that there are sharp changes in the probability of treatment close to the cutoff. We also conduct regressions of *magnet* on *eligible* after controlling for a female dummy, age, county-year fixed effects, and middle school fixed effects. A student whose score makes him just eligible to attend a magnet school is 59.7% more likely to actually be enrolled in a magnet school (Table 3, column 1).[5]

However, the second thing we can see from these graphs is that, in practice, the strictness of implementation of cutoff line is different in different counties. From the figures, we know that the cutoff line should be implemented the most strictly in counties 16, 20, 26, and 29. Columns 9 and 10 in Table 1 show the percentage of magnet school students having entrance score lower than cutoff line and the percentage of regular students having entrance scores higher than cutoff line in each county-year. We can see considerable variation in these two percentages. For example, 28.9 percent of magnet students having high school entrance scores lower than the cutoff line in county 28 in the year 2000; but in county 16 in 1997, county 16 in 1999, and county 29 in 1998, there are no magnet school students having high school entrance score lower than cutoff line. We can also see that 39.5 percent of regular high school students in county 19 in the year 2000 have high school entrance examination scores greater than the cutoff line; this percentage is the highest among all 13 county-years (see column 10 in Table 1); however, in county 16 in 2001, only 2.64 percent of regular students have high school entrance

[5] The impact of eligibility on enrolment is similar for the subsample of students with college entrance exam scores (57.7%, see column 5 of Table 3).

scores higher than the cutoff line. Because of this fuzziness in the implementation of cutoff lines, in the following analysis, our preferred results come from the regressions using *eligible* as an instrument for *magnet*. In section 7, we also examine the robustness of the results to excluding county-years with greatest fuzziness in assignment to magnet schools.

5. Results

5.1 Pre-assumption tests

We first report results for pre-assumption tests that examine whether gender composition or mean age of students jumps in a discontinuous fashion at the entrance exam cutoff line. In Figure 2, the x-axis measures the difference between the high school entrance exam scores and the cutoff line for each county-year; the y-axis in Panel A measures the proportion of students that are female, while in Panel B it measures mean student age. The curve in each graph is a nonparametric prediction from a local polynomial smoother with degree 0 and bandwidth 0.25. It is evident that there is no jump at x=0, the point at which the high school entrance examination score is equal to the cutoff line. This provides support for the validity of the regression discontinuity design.

Table 4 presents the regression results for the pre-assumption tests. Columns 1-4 are for the whole sample, and columns 5-8 are for the sub-sample of students with college entrance scores. In the regression specifications, we control for a 4^{th}-order polynomial function of the high school entrance exam score, county-year fixed effects and middle school fixed effects. Columns 1, 3, 5 and 7 present results for OLS regressions of female and age on *eligible*; while columns 2, 4, 6 and 8 present results for regressions of female

and age on *magnet* using *eligible* as an IV. From the table, we can see that, the coefficients for female and age are never statistically significant, whether they are estimated using OLS or IV regressions. These results confirm that the students' pre-treatment characteristics are continuous at the cutoff line. This is equally true if we exclude all imputed values.

5.2 Treatment effects

Figure 3 plots the three outcome variables as a function of the high school entrance exam score relative to cutoff lines: an indicator variable for having college entrance scores, the college entrance exam score, and an indicator variable for college eligibility conditional on taking the college entrance examination. Each graph plots two nonparametric predictions, one for students attending magnet schools and one for students attending regular schools.[6]

Panel A in Figure 3 presents the estimates for having a college entrance exam score. One can see that there is a gap between the two curves at x=0 (the entrance score cutoff line) which has a magnitude of roughly 0.05; and the curve representing untreated students is above the curve representing treated students. This suggests that entering a magnet school reduces the probability of taking a college entrance examination. Panel B in Figure 3 presents a similar graph for the college entrance examination score. The curve for magnet school students is above the curve for regular school students at x=0; the magnitude of the gap is about 0.3, which implies a 0.3 standard deviation increase in college entrance exam score if one attends a magnet school. Finally, in Panel C which plots the results for college eligibility, we can see that the curve of treated students is

[6] Each nonparametric prediction is from a local polynomial smoother with degree 0 and bandwidth 0.5.

above the curve of untreated students at x=0. The gap magnitude is about 0.2, which implies a 20 percent increase in the probability of qualifying for college. These pictures provide a visual estimate of the size of the treatment effect using RDD, but do not control for individual characteristics of students.

5.2.0 First stage results

Table 3 presents the first stage regression results. Columns 1 and 2 present results using the full sample; and columns 3 and 4 present results for the sub-sample of students having college entrance examination scores. All specifications control for county-year fixed effects and a 4^{th}-order polynomial function of the high school entrance exam score. In columns 1 and 3, we do not control for other variables; in columns 2 and 4, we control for gender, age and middle school fixed effects. The last row of Table 3 presents F-values for the null hypothesis that the coefficient of *eligible* is equal to zero.

For the full sample, both coefficients for *eligible* are statistically significant at the 1% significance level, with magnitudes of 0.490 and 0.397. The F-values are 990.4 and 682.4. For the sub-sample, both coefficients for *eligible* are statistically significant at the 1% level, with magnitudes of 0.478 and 0.397, which are almost identical to the results for the full sample. The corresponding F-values are 633.2 and 444.9. These results suggest that whether one takes the college entrance exam is not systematically related to selectivity associated with enrolling in a magnet school conditional on being eligible. They also show that *eligible* is an extremely strong predictor of actually enrolling in a magnet school, justifying its use as an instrument.

5.2.1 Effects on students' having a college entrance examination score

Table 5 presents estimates of the effect of entering a magnet school on whether a college examination score is observed. As for subsequent outcomes, three types of specifications are estimated. The first one uses *magnet* as the independent variable of interest. The second uses *eligible* as the independent variable. The third uses *eligible* as an instrument for *magnet*, which is our preferred specification for the reasons described earlier. In all specifications, we control for county-year fixed effects and a 4^{th}-order polynomial function of the high school entrance exam score. In columns 1 to 3, we do not control for gender, age, or middle school fixed effects. All three coefficients are negative and statistically significantly different from zero, with magnitudes of -0.054, -0.048, and -0.098. Columns 4 to 6 present results adding controls for gender, age, and the middle school the student attended. The impact of *magnet* increases in magnitude to -0.064 (column 4), the impact of *eligible* falls to -0.037 (column 5); and the IV estimate changes little, becoming -0.093 (column 6).

One interpretation for the negative effect is related to the fact that the RDD compares students just above the cutoff line, who tend to be the worst students in the best school, with students with scores just below the cutoff line, who tend to be the best students in regular schools. It is possible that the worst students in the best school have less confidence and decide not to take the test. Or they may have greater ambitions (via peer effects) and so postpone taking the college entrance examination for a year to allow for more preparation (the scores of such students are not observed in the data). It is also possible that teachers in magnet schools, who feel tremendous pressure to have their classes attain a high college pass rate (a highly visible statistic), discourage less capable

students from taking the college entrance examination. In contrast, the best students in regular schools have greater confidence and are expected by their teachers to take the college entrance examination.

One concern that arises with missing college examination scores is that our estimates of the impact of attending a magnet school on college examination scores could suffer from bias caused by differences in the selectivity of who takes examinations in magnet schools and in regular schools. For example, if magnet schools only allowed high ability students to take the college entrance exam, while taking the college entrance examination is fairly random in regular schools, we will overestimate the impact of attending a magnet school on college entrance examination scores. If selectivity associated with ability were greater in regular schools, then the bias would be in the opposite direction.

While we cannot say anything about unobservables, we can gain insight into the likelihood of selectivity bias by examining how the likelihood of taking the college entrance examination is related to high school entrance examination scores. In Table 6, columns 1 and 2 presents results of regressions of observing a college entrance examination score on high school entrance exam scores for the samples of students in magnet and regular schools, respectively. A one standard deviation higher high school entrance exam score increases the probability of observing a college examination score by 6.1 percent in magnet schools and by 6.7 percent in regular schools. Both are statistically significant at the one percent level. The results show that selectivity associated with observed ability is slightly greater in regular schools than in magnitudes,

which suggests that if anything our estimates of the impact of attending a magnet school on college entrance examination scores is downward biased.

5.2.2 Effects on students' college entrance score and college eligibility

Table 7 presents the estimates of the effects of entering a magnet school on a student's college entrance exam score and a student's probability of qualifying to enter college. For each outcome, we estimate the same three specifications as before, where the treatment variables are *magnet*, *eligible*, and *magnet* instrumented by *eligible*. The sample includes only students for whom we observe a college entrance examination score, so the results should be interpreted as the effects of school quality conditional on students' taking the college entrance exam. In columns 1 to 3, we control only for the 4th order polynomial function of high school entrance exam scores and county-year fixed effects, and do not control for gender, age, or middle school fixed effects. All three coefficients for the treatment variables are positive and statistically significantly different from zero, with magnitudes of 0.157, 0.119, and 0.248. As expected, the preferred IV estimate is larger than both OLS estimates. Columns 4 to 6 present results adding controls for gender, age, and which middle school the student attended. The impact of *magnet* decreases to 0.118 (column 4), the impact of *eligible* falls to 0.083 (column 5); and the IV estimate of the impact of *magnet* falls to 0.210 (column 6). The latter is our preferred estimate for the impact of entering a magnet school on students' college entrance examination scores.

The bottom panel of Table 7 shows similar results for the impact of school quality on students' probability of qualifying to enter college conditional on taking the college entrance examination. In the simplest specification without individual controls (columns

1 to 3), the effect of *magnet* is 0.056, the effect of *eligible* is 0.087, and the IV estimate for *magnet* is 0.183. Again, as expected the largest estimate is the IV estimate. After controlling for gender, age, and middle school fixed effects, the effect of *magnet* decreases to 0.041, the effect of *eligible* decreases to 0.074, and the IV *magnet* effect estimate increases to 0.187 (columns 4 to 6). Thus, according to our preferred estimate, attending a better high school increases the probability of qualifying for college entrance by 18.7 percentage points conditional on taking the college entrance examination.

The unconditional effects of school quality on students' probability of being admitted to college might be of even greater interest to students and policy-makers than the conditional effects. In Table 8, we repeat the estimates presented in Table 7 of the determinants of entering college, but this time we include the full sample, so that students not being admitted to college include both those who did not take the college entrance examination and those who took the exam but earned scores too low to be admitted to college. Columns 1 to 3 show the results from the simplest specification; the effect of *magnet* is 0.010 (column 1), the effect of *eligible* is 0.047 (column 2), and the IV estimate of the effect of *magnet* is 0.096. Controlling for gender, age, and middle school fixed effects, we find that the effect of *magnet* decreases to -0.006, the effect of *eligible* decreases to 0.037, and the IV estimate of the effect of *magnet* increases to 0.094 (colums 4 to 6). Thus, entering a better high school increases the probability of qualifying for college entrance by 9.4 percentage points, or by nearly 20 percent (since average college pass rates are about 50 percent).

6. Magnet schools versus regular schools

In this section, we explore the ways in which magnet schools actually differ from regular schools, which could account for the better performance of students in magnet schools. The school questionnaire provides annual information on a variety of school characteristics. We measure school quality for an entering class by the 4-year average values of school indicators that span the years that they attended the school. For example, for students starting high school in September 1997 and graduating in June 2000, we take mean values for the years 1997 to 2000.

Table 9 presents simple regression-based comparisons between magnet and regular schools. Columns 1 to 4 show the difference in learning environments between magnet and regular schools. From columns 1 and 2, we can see that the share of teachers with the highest quality ranks in magnet schools is 0.132 higher than in regular schools, and the share of teachers with four-year college education is 0.410 higher in magnet schools than in regular schools. Both of these two differences are statistically significant at the 1% level. Teacher quality ranks are cumulative, reflecting teachers' credentials, experience, and past teaching performance. Columns 3 and 4 compare peer students' quality (measured by average high school entrance exam scores of students entering the same high school in the same year) and class size. As expected, magnet schools have higher high school entrance exam scores; the difference is 1.53 standard deviations (column 3) which is significant at the 1% level. From column 4, we can see that class size in magnet schools is greater (by 9 students, or 17 percent) than in regular schools and the difference is statistically significant at the 5 prcent level. Columns 5 to 8 examine the number of students and school facilities. Magnet schools have 1098 more students (or 144 percent) than regular schools. In addition, compared to regular schools magnet schools are larger

in area by 53 thousand square meters (212 percent), have 66,090 (702 percent) more library books, and the share of magnet schools meeting national criteria for facilities is 0.284 (74 percent) greater than of regular schools. All differences are significantly different from zero at the one percent level except for equipment quality which is insignificant.

Unfortunately, because we only are able to estimate treatment effects for 7 counties, we are unable to credibly identify the relative contribution of different school characteristics in a multivariate framework. Thus, we are unable to distinguish whether teacher quality, school facilities, peer effects, or some other characteristic explain the difference in student performance. In the context studied, there are clear differences in all of these domains between magnet and regular high schools.

7. Extensions

7.1 Robustness to excluding fuzzy implementation of cutoff lines

In our sample, counties differ in the strictness with which high school entrance cutoff lines determine whether students attend magnet schools. As a robustness test, we estimate results restricting the sample to those county-years in which the percentage of students in magnet schools having high school entrance exam scores lower than the cutoff line is less than 10 percent rather than the 30 percent criteria used in the previous analysis. The results are presented in Table 10.

The estimated specifications are the same as those used to produce the results in Tables 5 and 7. The upper panel shows the impacts on the probability of observing a college entrance exam score. In columns 1 to 3, we report results without individual

control variables. The coefficient on *magnet* is 0.076, which is statistically significant at the one percent level. The coefficient on *eligible* is -0.010, and the IV estimate is -0.014, neither of which are statistically significantly different from zero. In columns 4 to 6, we add controls for gender, age, and middles school fixed effects, which increases the three estimates to 0.109, 0.011, and 0.020. The latter two estimates remain statistically insignificant. Thus, with a more restricted sample, the negative impact of attending a magnet school on taking the college entrance exam disappears.

The middle panel shows the impacts on college entrance exam scores. Without individual controls, the three coefficients of interest are 0.392, 0.368 and 0.488 (columns 1 to 3), all statistically significant at the 1% level. Controlling for gender, age, and middle school fixed effects, the coefficient on *magnet* increases to 0.430, the coefficient on *eligible* decreases to 0.309, and the IV estimate of the impact of *magnet* increases to 0.511 (columns 4 to 6). All of these estimates are statistically significant at the one percent level, and are nearly double the size of the estimated impacts using the main sample (Table 7).

The impacts of school quality on the probability of students qualifying for college admission are presented in the bottom panel. Without controls, the three coefficients of interest are 0.179, 0.194, and 0.257 (columns 1 to 3). Controlling for gender, age, and middle school fixed effects, the coefficient on *magnet* increases to 0.197, the coefficient on *eligible* decreases to 0.172, and the IV estimate for the impact of *magnet* increases to 0.284 (colums 4 to 6). All the coefficients in this panel are statistically significant at the one percent level, and the preferred IV estimate is substantially larger (by 52 percent) than the measured impact using the main sample (Table 7).

Differences in the estimates using the smaller sample are consistent with magnet teachers discouraging more students from taking the college entrance examination when there are more lower quality students who are admitted to the magnet school despite having scores below the cutoff line. By diluting the quality of students in magnet schools relative to regular schools, more violations of the assignment rule based on cutoff lines also could weaken peer effects, thus reducing the impact of being in a magnet school on college entrance exam scores. Having more low quality students could also distract the attention of teachers in magnet schools, hurting their overall performance.

7.2 Impacts of entering good schools on the types of exams students take

In China, high school students can choose different courses of study, focusing on liberal arts (wenke), natural sciences (like), physical education, or musical education. The vast majority (97% in our sample) choose either liberal arts or natural sciences. If entering magnet schools increases the probability of students choosing liberal arts (viewed as less competitive than natural sciences) and it is easier for students studying liberal arts to earn a higher entrance score, then our estimates of the impact of attending a magnet school on college entrance scores and college admissions probability could be upward biased.

To test whether this concern is likely to be important, we estimate the impact of entering a magnet school on the probability of choosing different courses of study. The results are presented in Table 11. In the upper panel, we restrict the sample to students studying liberal arts or natural sciences, and define the dependent variable to equal one if the student chooses natural sciences, and zero otherwise. We estimate the same

specifications as before. With no individual control variables, the coefficient on *magnet* is 0.029, statistically significant at 5 percent level, the coefficient on *eligible* is 0.003, and the IV estimate for *magnet* is 0.006, neither of the latter two estimates being statistically significantly different from zero. When we add controls for gender, age, and middle school fixed effects, none of the coefficients are statistically significant, with the magnitudes of the three coefficients being 0.023, 0.006, and 0.015 (columns 4 to 6). These results are robust to including students taking all tests other than natural sciences in the category coded as zero (see bottom panel). Thus, overall there is no evidence that attending a magnet school significantly impacts students' courses of study.

7.3 Heterogeneity tests

The results presented in Section 5 are average impacts of school quality on students' educational attainment. However, the impacts might be different for different groups of students. In this section, we present the results of heterogeneity tests with respect to gender and age, with results presented in Table 12.

In all specifications, we control for female, age, county-year fixed effects, middle school fixed effects and a 4^{th} order polynomial function of high school entrance exam scores. In Table 12, we only present the preferred IV estimates. Columns 1, 2 and 3 report results when the dependent variables are whether a college entrance exam score is observed, college entrance exam scores, and the probability of being admitted to college. The coefficients on the interaction of *magnet* and the female dummy are not statistically significant for any of the outcomes, equal to -0.038, -0.044 and 0.039 respectively. Columns 4 to 6 report results when we investigate whether there are heterogeneous

impacts with respect to age. From the table, we can see that all the coefficients on the interaction of *magnet* and age are not statistically significant (they are -0.002, 0.090, and -0.037, respectively). These results provide no evidence that entering a magnet high school has different impacts on girls versus boys, or on students of different ages.

8. Conclusions

Whether school quality can improve students' educational attainment is an important and highly debated question in the economics of education. This paper uses China's magnet school system and a regression discontinuity design approach to credibly estimate the effects of higher high school quality on students' educational attainment.

Using data from seven counties in Gansu province in China, according to our preferred IV estimates, we find that for students whose high school entrance scores are near the cutoff line for being admitted to magnet schools, attending a magnet high school decreases the probability of taking the college entrance examination by 9.4 percent, increases students' college entrance score by 0.210 standard deviations, and increases the probability of qualifying for college by 18.7 percent conditional on taking the college entrance exam.

References

Ahn, Hyungtaik, and James L. Powell (1993). *Semiparametric estimation of censored selection models with a nonparametric selection mechanism,* Journal of Econometrics 58 3-29.

Altonji, Joseph G. and Thomas A. Dunn (1996). *Using siblings to estimate the effect of school quality on wages,* The Review of Economics and Statistics, vol. 78, No. 4, 665-671.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber (2005). *Selection on observed and unobserved variables: assessing the effectiveness of Catholic schools*, Journal of Political Economy, vol. 113, No. 1, 151-184.

Angrist, Joshua D. and Victor Lavy (1999). *Using Maimonides' rule to estimate the effect of class size on scholastic achievement,* The Quarterly Journal of Economics, 533- 575.

Angrist, Joshua D. and Victor Lavy (2001). *Does teacher training affect pupil learning? Evidence from matched comparisons in Jerusalem public schools*, Journal of Labor Economics, vol. 19, No. 2, 343-369.

Benerjee, A., S. Cole, E. Duflo, and L. Linden (2006) Remedying education: evidence from two randomized experiments in India. MIT working paper.

Betts, Julian R. (1995). *Does school quality matter? Evidence from the national longitudinal surveys of youth* The Review of Economics and Statistics, vol. 77, No. 2, 231-250.

Betts, Julian R. (2001). *The impact of school resources on women's earnings and educational attainment: Findings from the national longitudinal survey of young women* Journal of Labor Economics, vol. 19, No. 3.

Card, David, and Alan B. Krueger (1992a). *Does school quality matter? Returns to education and the characteristics of public schools in the United States*, Journal of Political Economy, Vol. 100, No. 1. 1-40.

Card, David and Alan B. Krueger (1992b), *School quality and black-white relative earnings: A direct assessment,* The Quarterly Journal of Economics, vol. 107, No. 1, 151-200.

Case, Anne and Angus Deaton (1999). *School inputs and educational outcomes in South Aftica,* The Quarterly Journal of Economics, vol. 114, No. 3. 1047-1084.

Dee, Thomas S. (2001). *Teachers, race and student achievement in a randomized experiment,* NBER Working Paper 8432.

Eide, Eric and Mark H. Showalter (1998). *The effect of school quality on student performance: a quantile regression approach,* Economics Letters, 58, 345-350.

Greenwald, Rob, Larry V. Hedges, and Richard D. Laine (1996). *The effect of school resources on student achievement*, Review of Educational Research, vol. 66, No. 3, 361-396.

Goldhaber, D. Dan and Dominic J. Brewer (1997). *Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity,* The Journal of Human Resources, vol. 32, No. 3, 505-523.

Gould, Eric D., Victor Lavy, and M. Daniele Paserman (2004). *Immigrating to opportunity: Estimating the effect of school quality using a natural experiment on Ethiopians in Israel*, The Quarterly Journal of Economics, 489-526.

Hahn, Jingyong, Petra Todd, and Wilbert van der Klaauw (2001). *Identification and estimation of treatment effects with a regression-discontinuity design*, Econometrica, 69, 201-209.

Heckman, James J. (2005). *China's human capital investment,* China Economic Review 16 50-70.

Heckman, James J. and R. Robb (1985). *Alternative methods for evaluating the impact of interventions* Longitudinal Analysis of Labor Market Data, edited by J. Heckman and B. Singer, Cambridge University Press.

Hoxby, Caroline M. (2000). *The effects of class size on student achievement: new evidence from population variation,* The Quarterly Journal of Economics, November 2000, 1239-1285.

Hoxby, Caroline M. (2007). *The effects of New York City's charter schools on student achievement*, Working paper.

Krueger, Alan B. (1999). *Experimental estimates of education production functions,* The Quarterly Journal of Economics, 497-532.

Krueger, Alan B. and Diane M. Whitmore (1999). *The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from Project STAR,* Working Paper # 427, Industrial Relations Section, Princeton University.

Lai, Fang, Elisabeth Sadoulet, and Alain de Janvry (2006). *The effect of school and teacher quality on student performance: Using a natural experiment from the middle school reforms in Beijing,* Working Paper.

Lavy, Victor (2002). *Evaluating the effect of teachers' group performance incentives on pupil achievement,* Journal of Political Economy, vol. 110, No. 6, 1286-1317.

Lee, David S. (2003). *Randomized experiments from non-random selection in U.S. House Elections,* NBER working paper # 8441.

Lee, David S., Enrico Moretti, and Matthew J. Butler (2004). *Do voters affect or elect policies? Evidence from the U.S. House,* Quarterly Journal of Economics, 119, 807-860.

Newhouse, David and Kethleen Beegle (2005). *The effect of school type on academic achievement: Evidence from Indonesia* World Bank Policy Research Working Paper 3604.

Porter, Jack (2003). *Estimation in the regression discontinuity model*, Harvard University manuscript

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain (2005). *Teachers, schools, and academic achievement,* Econometrica, vol. 73, No. 2, 417-458.

Rockoff, Jonah E (2004). *The impact of individual teachers on student achievement: Evidence from panel data,* American Economic Review, 94, 2, 247-252.

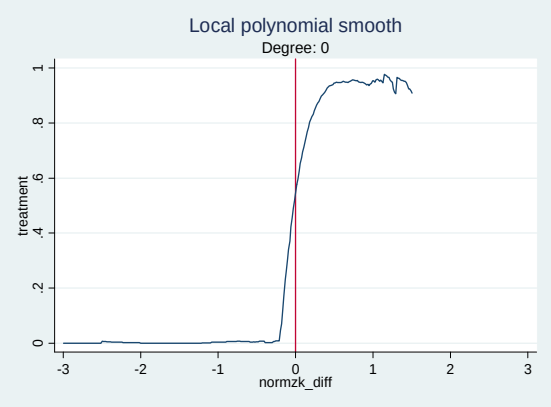
Thistlethwaite, D. and D. Campbell (1960). *Regression-Discontinuity analysis: an alternative to the ex post facto experiment,* Journal of Educational Psychology 51, 309-17.

Urquiola, Miguel (2006). *Identifying class size effects in developing countries: evidence from rural Bolivia ,* The Review of Economics and Statistics, 88(1), 171-177.

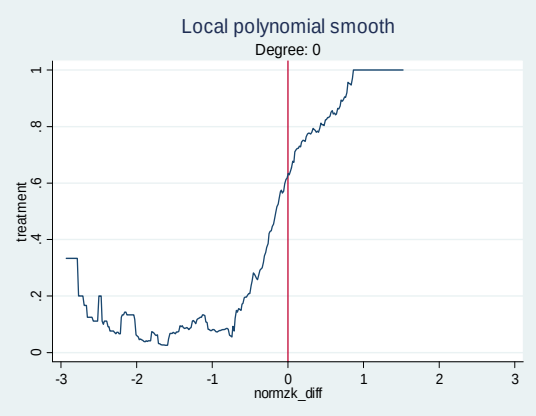
Van der Klaauw, Wilbert (2002). *Estimating the effect of financial aid offers on college enrollment: A regression-discontinuity approach,* International Economic Review, Vol43 (4), November.

Figure 1 Student assignment

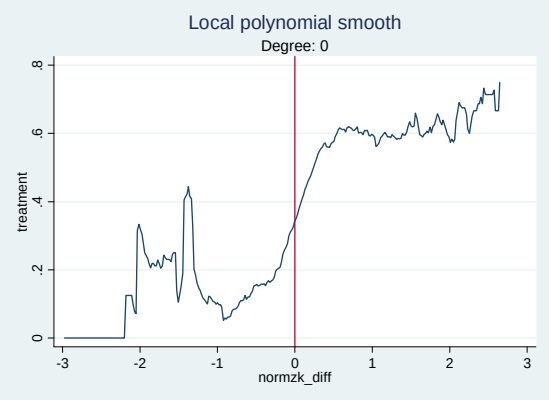
Panel A. County16



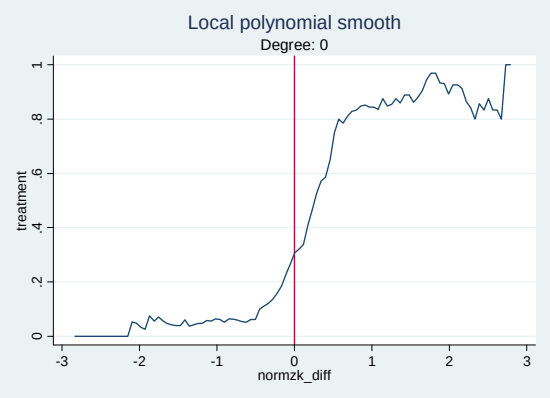
Panel B. County 18



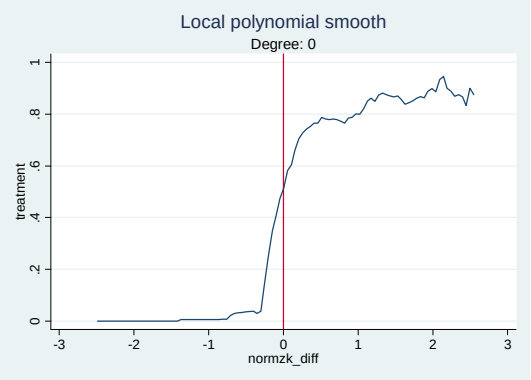
Panel C. County 19



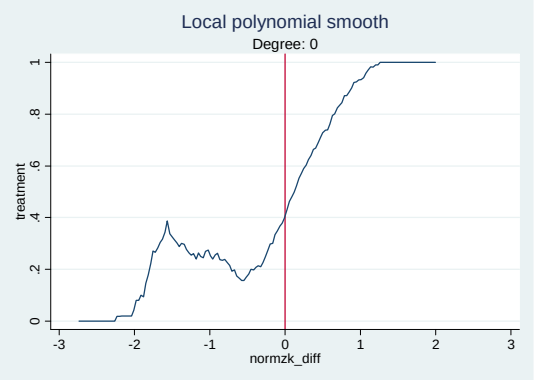
Panel D. County 20



Panel E. County 26



Panel F. County 28



Panel G. County 29

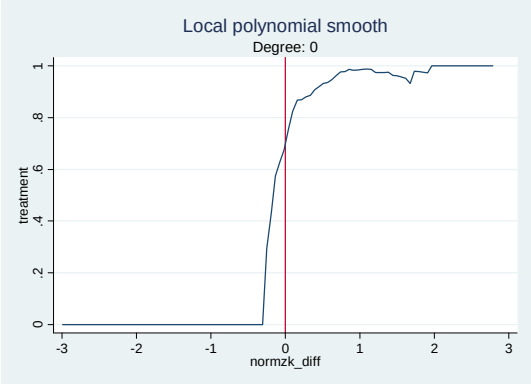
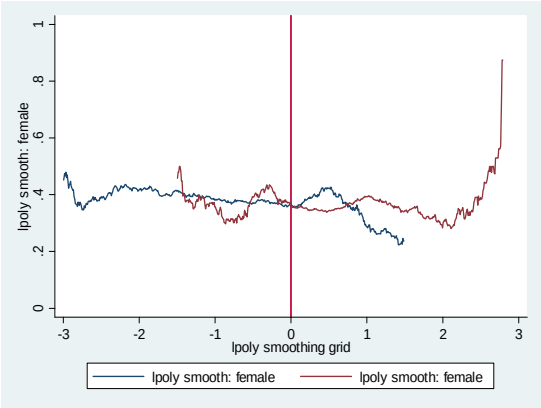


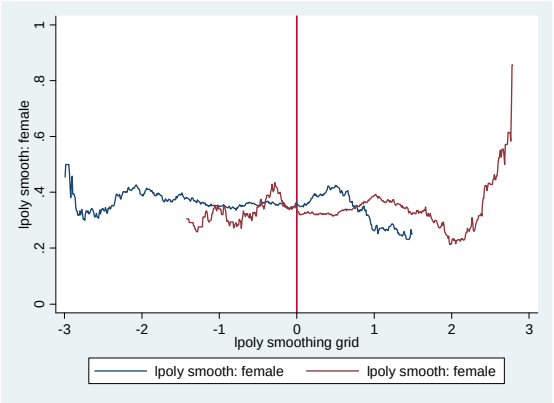
Figure 2 Pre-assumption test

Panel A Female

Whole sample

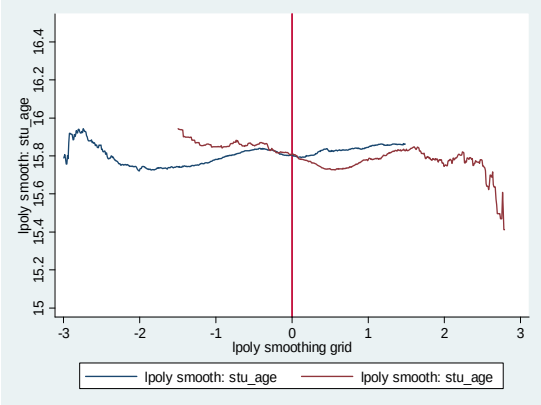


Sub-sample



Panel B Student age

Whole sample



Sub-sample

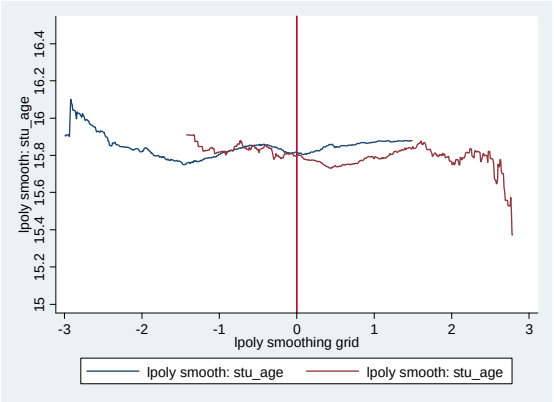
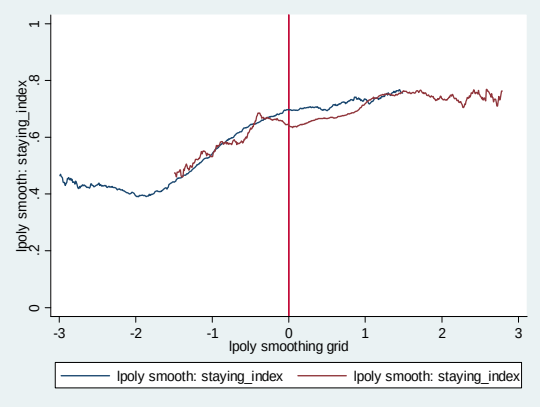
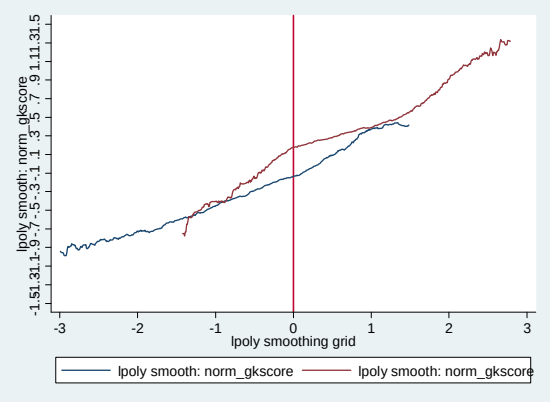


Figure 3 Effects of school quality

Panel A Having college entrance score



Panel B College entrance examination score



Panel C College eligibility

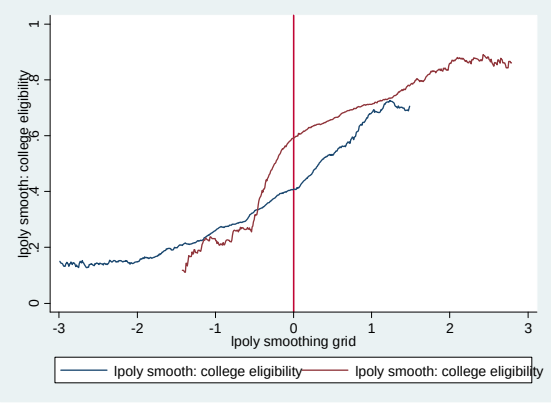


Table 1. Sample description

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
County	Enter year	Num. of schools	Num. of students	Highest cutoff line	Magnet cutoff line	Magnet average entering score	Magnet students (%)	Magnet students with high school entrance score lower than cutoff line(%)	Regular students with high school entrance score higher than cutoff line(%)	Students with missing college entrance score(%)	Students not admitted by college(%)	Students admitted by college(%)	Students with missing college admission information(%)	Students with missing gender(%)
16	1997	3	531	560	560	597.35	51.32%	0.00%	3.86%	68.42%	0.00%	0.00%	100.00%	0.00%
16	1998	5	1195	564	564	613.75	27.28%	1.53%	8.52%	64.69%	0.00%	0.00%	100.00%	0.00%
16	1999	4	762	554.5	554.5	609.19	42.52%	0.00%	12.56%	56.96%	0.00%	0.00%	100.00%	0.00%
16	2000	3	784	560	560	599.59	34.69%	1.10%	8.98%	44.64%	0.00%	0.00%	100.00%	0.00%
16	2001	3	1032	568	568	597.85	37.75%	0.00%	2.64%	34.85%	0.00%	0.00%	100.00%	0.00%
18	1997	9	463	530	530	547.49	46.44%	26.51%	13.71%	0.00%	0.00%	0.00%	100.00%	0.00%
19	2000	5	1805	540	540	566.20	39.26%	18.90%	39.47%	18.72%	0.00%	0.00%	100.00%	0.00%
20	1998	5	1137	518.5	518.5	558.14	34.92%	14.61%	16.62%	51.36%	0.00%	0.00%	100.00%	0.00%
26	1999	3	391	543	543	579.95	55.87%	0.91%	34.68%	46.68%	0.00%	0.00%	100.00%	0.00%
26	2001	3	812	532	532	575.38	47.60%	1.03%	22.77%	46.62%	0.00%	0.00%	100.00%	0.00%
28	2000	4	1572	553	553	563.09	48.47%	28.87%	23.70%	25.45%	11.64%	17.94%	70.42%	0.00%
29	1997	2	386	438	438	501.42	81.09%	2.56%	35.62%	25.91%	40.67%	33.16%	25.91%	0.00%
29	1998	2	572	430	430	495.91	86.21%	0.00%	10.13%	13.09%	69.28%	17.45%	13.09%	0.00%

Table 2. Summary statistics

Variable	College entrance exam score	Eligible for college	Female	Age	Magnet	Eligible
All students						
Mean	NA	NA	0.374	15.785	0.444	0.503
S.D.	NA	NA	0.484	0.369	0.497	0.500
Observation	NA	NA	11448	11448	11448	11448
Students having college entrance exam scores						
Mean	0.000	0.495	0.352	15.803	0.485	0.563
S.D.	0.996	0.500	0.478	0.386	0.500	0.496
Observation	7108	7108	7108	7108	7108	7108

Note:

(1) Magnet =1 if the student enters good high school, which is defined as the school whose students' average entering grade is the highest in this county and year

(2) Eligible =1 if the student's entering grade is higher than the cutoff line of the good high school defined above

Table 3. Determinants of Attending a Magnet School

Dependent variable: Magnet	(1)	(2)	(3)	(4)
Eligible	0.490	0.397	0.478	0.397
	(0.016)***	(0.015)***	(0.019)***	(0.019)***
Female	NO	YES	NO	YES
Age	NO	YES	NO	YES
County-year fixed effects	YES	YES	YES	YES
Middle school fixed effects	NO	YES	NO	YES
4th order polynomial terms of high school entrance exam score	YES	YES	YES	YES
Observations	11448	11448	7108	7108
R-squared	0.55	0.61	0.53	0.59
F value (Eligible=0)	990.38	682.36	633.17	444.92

Standard errors in parentheses

Table 4. Pre-assumption tests

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Female	Female	Age	Age	Female	Female	Age	Age
Eligible	0.017		0.021		0.029		0.010	
	(0.019)		(0.014)		(0.023)		(0.020)	
Magnet(Eligible as IV)		0.042		0.054		0.072		0.025
		(0.049)		(0.036)		(0.059)		(0.050)
4th order polynomial terms of high school entrance exam score	YES	YES	YES	YES	YES	YES	YES	YES
County-year fixed effects	YES	YES	YES	YES	YES	YES	YES	YES
Middle school fixed effects	YES	YES	YES	YES	YES	YES	YES	YES
Observations	11448	11448	11448	11448	7108	7108	7108	7108
R-squared	0.03	0.03	0.35	0.35	0.05	0.05	0.37	0.37

* significant at 10%; ** significant at 5%; *** significant at 1%

Columns 1 to 4 include all observations in main sample, columns 5 to 8 include students with college entrance exam scores.

Standard errors in parentheses.

Table 5. Determinants of Observing a College Entrance Examination Score

	(1)	(2)	(3)	(4)	(5)	(6)
Magnet	-0.054			-0.064		
	(0.012)***			(0.012)***		
Eligible		-0.048			-0.037	
		(0.017)***			(0.017)**	
Magnet(Eligible as IV)			-0.098			-0.093
			(0.035)***			(0.043)**
Female	NO	NO	NO	YES	YES	YES
Age	NO	NO	NO	YES	YES	YES
Middle school fixed effect	NO	NO	NO	YES	YES	YES
4th order polynomial terms of high school entrance exam score	YES	YES	YES	YES	YES	YES
County-year fixed effect	YES	YES	YES	YES	YES	YES
Observations	11448	11448	11448	11448	11448	11448
R-squared	0.18	0.17	0.17	0.22	0.22	0.22

Standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

Note:

Table 6. Difference between students having and not having college entrance exam scores

Having a college entrance examination score=1		
	(1)	(2)
High school entrance exam score	0.061	0.067
	(0.010)***	(0.007)***
Female	YES	YES
Age	YES	YES
Middle school fixed effect	YES	YES
County-year fixed effect	YES	YES
Observations	5081	6367
R-squared	0.19	0.28

Robust standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

Note:

(1) Column 1 uses the sample of magnet school students, and column 2 uses the sample of regular school students.

Table 7. Effects of school quality on students' college entrance score and college eligibility

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent Variable : College entrance exam score					
Magnet	0.157			0.118		
	(0.027)***			(0.029)***		
Eligible		0.119			0.083	
		(0.041)***			(0.042)**	
Magnet(Eligible as IV)			0.248			0.210
			(0.085)***			(0.106)**
Female	NO	NO	NO	YES	YES	YES
Age	NO	NO	NO	YES	YES	YES
Middle school fixed effect	NO	NO	NO	YES	YES	YES
4th order polynomial terms of high school entrance exam score	YES	YES	YES	YES	YES	YES
County-year fixed effect	YES	YES	YES	YES	YES	YES
Observations	7108	7108	7108	7108	7108	7108
R-squared	0.26	0.26	0.26	0.30	0.30	0.30
	Dependent Variable : College entrance score is higher than lowest cutoff line=1					
Magnet	0.056			0.041		
	(0.015)***			(0.016)**		
Eligible		0.087			0.074	
		(0.022)***			(0.023)***	
Magnet(Eligible as IV)			0.183			0.187
			(0.047)***			(0.058)***
Female	NO	NO	NO	YES	YES	YES
Age	NO	NO	NO	YES	YES	YES
Middle school fixed effect	NO	NO	NO	YES	YES	YES
4th order polynomial terms of high school entrance exam score	YES	YES	YES	YES	YES	YES
County-year fixed effect	YES	YES	YES	YES	YES	YES
Observations	7108	7108	7108	7108	7108	7108
R-squared	0.23	0.23	0.22	0.25	0.25	0.24

Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Note:

(1) Magnet =1 if the student enters good high school, which is defined as the school whose students' average entering grade is the highest in this county and year

(2) Eligible=1 if the student's entering grade is higher than the cutoff line of the good high school defined above

Table 8 .Total effects of entering good school on the probability to enter college

	Dependent Variable : College entrance score is higher than lowest cutoff line=1					
	(1)	(2)	(3)	(4)	(5)	(6)
Magnet	0.010			-0.006		
	(0.012)			(0.013)		
Eligible		0.047			0.037	
		(0.017)***			(0.018)**	
Magnet(Eligible as IV)			0.096			0.094
			(0.036)***			(0.045)**
Female	NO	NO	NO	YES	YES	YES
Age	NO	NO	NO	YES	YES	YES
Middle school fixed effect	NO	NO	NO	YES	YES	YES
County-year fixed effects	YES	YES	YES	YES	YES	YES
4th order polynomial terms of high school entrance exam score	YES	YES	YES	YES	YES	YES
Observations	11448	11448	11448	11448	11448	11448
R-squared	0.17	0.17	0.17	0.20	0.20	0.20

Standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

(1) Magnet =1 if the student enters good high school, which is defined as the school whose students' average entering grade is the highest in this county and year

(2) Eligible=1 if the student's entering grade is higher than the cutoff line of the good high school defined above

(3) Students with missing college entrance score are assumed not to enter colleges

Table 9. Different characteristics between magnet schools and regular schools

	(1)	(2)	(3)	(4)
	Ratio of teachers having advanced title	Ratio of teachers having education of four year college	Avg. high school entrance exam score of peer students	Class size
Magnet school=1	0.132	0.410	1.532	9.045
	(0.032)***	(0.083)***	(0.173)***	(4.124)**
County-year fixed effects	YES	YES	YES	YES
Observations	50	47	51	45
R-squared	0.38	0.46	0.72	0.45
	(5)	(6)	(7)	(8)
	No. of students	Campus area (10000 square meters)	No. of books in library (10000 units)	Does equipment satisfy criteria
Magnet school=1	1,098.305	5.302	6.609	0.284
	(156.727)***	(0.719)***	(1.095)***	(0.177)
County-year fixed effects	YES	YES	YES	YES
Observations	46	46	44	40
R-squared	0.69	0.71	0.62	0.56

Standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 10. Determinants of Educational Attainment Excluding Fuzzy Implementation of Cutoff Scores

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent Variable : Having a college entrance exam score=1					
Magnet	0.076			0.109		
	(0.021)***			(0.025)***		
Eligible		-0.010			0.011	
		(0.025)			(0.027)	
Magnet(Eligible as IV)			-0.014			0.020
			(0.035)			(0.046)
Female	NO	NO	NO	YES	YES	YES
Age	NO	NO	NO	YES	YES	YES
Middle school fixed effect	NO	NO	NO	YES	YES	YES
Observations	6470	6470	6470	6470	6470	6470
R-squared	0.14	0.14	0.14	0.21	0.21	0.21
	Dependent Variable : College entrance exam score					
Magnet	0.392			0.430		
	(0.051)***			(0.066)***		
Eligible		0.368			0.309	
		(0.062)***			(0.068)***	
Magnet(Eligible as IV)			0.488			0.511
			(0.081)***			(0.112)***
Female	NO	NO	NO	YES	YES	YES
Age	NO	NO	NO	YES	YES	YES
Middle school fixed effect	NO	NO	NO	YES	YES	YES
Observations	3452	3452	3452	3452	3452	3452
R-squared	0.30	0.29	0.30	0.35	0.35	0.35
	Dependent Variable : College entrance score is higher than lowest cutoff line=1					
Magnet	0.179			0.197		
	(0.027)***			(0.033)***		
Eligible		0.194			0.172	
		(0.033)***			(0.035)***	
Magnet(Eligible as IV)			0.257			0.284
			(0.044)***			(0.058)***
Female	NO	NO	NO	YES	YES	YES
Age	NO	NO	NO	YES	YES	YES
Middle school fixed effect	NO	NO	NO	YES	YES	YES
Observations	3452	3452	3452	3452	3452	3452
R-squared	0.27	0.27	0.27	0.30	0.30	0.30

Standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%

Note: (1) Magnet =1 if the student enters good high school, which is defined as the school whose students' average entering grade is the highest in this county and year

(2) Eligible=1 if the student's entering grade is higher than the cutoff line of the good high school defined above

(3) All specifications include county-year fixed effects, and 4 order polynomial function of normalized high school entrance scores

(4) In this sample, the county-years with more than 10% magnet students with high school entrance sore lower than cutoff lines are dropped

Table 11. Effects of Entering a Magnet School on Student Courses of Study

	(1)	(2)	(3)	(7)	(8)	(9)
	Dependent Variable : Taking natural sciences exam=1, taking liberal arts exam=0					
Magnet	0.029			0.023		
	(0.013)**			(0.014)		
Eligible		0.003			0.006	
		(0.020)			(0.021)	
Magnet(Eligible as IV)			0.006			0.015
			(0.043)			(0.053)
Female	NO	NO	NO	YES	YES	YES
Age	NO	NO	NO	YES	YES	YES
Middle school fixed effect	NO	NO	NO	YES	YES	YES
County-year fixed effects	YES	YES	YES	YES	YES	YES
4 order polynomial function of normalized high school entrance score	YES	YES	YES	YES	YES	YES
Observations	6899	6899	6899	6899	6899	6899
R-squared	0.09	0.09	0.09	0.13	0.13	0.13
	Dependent Variable : Taking natural sciences exam=1, taking other type of exams=0					
Magnet	0.028			0.024		
	(0.014)**			(0.015)		
Eligible		0.011			0.013	
		(0.021)			(0.021)	
Magnet(Eligible as IV)			0.022			0.032
			(0.043)			(0.053)
Female	NO	NO	NO	YES	YES	YES
Age	NO	NO	NO	YES	YES	YES
Middle school fixed effect	NO	NO	NO	YES	YES	YES
County-year fixed effects	YES	YES	YES	YES	YES	YES
4 order polynomial function of normalized high school entrance score	YES	YES	YES	YES	YES	YES
Observations	7108	7108	7108	7108	7108	7108
R-squared	0.10	0.10	0.10	0.13	0.13	0.13

Standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

(1) Magnet =1 if the student enters good high school, which is defined as the school whose students' average entering grade is the highest in this county and year

(2) Eligible=1 if the student's entering grade is higher than the cutoff line of the good high school defined above

Table 13 Heterogeneity tests

	(1)	(2)	(3)	(4)	(5)	(6)
	Having a college entrance exam score	College entrance exam score	College eligibility	Having a college entrance exam score	College entrance exam score	College eligibility
Magnet (Eligible as IV)	-0.080	0.224	0.175	-0.059	-1.143	0.739
	(0.044)*	(0.107)**	(0.058)***	(0.286)	(1.186)	(0.565)
Female	-0.037	-0.039	-0.053	YES	YES	YES
	(0.014)***	(0.034)	(0.018)***			
Magnet*Female (Eligible*Female as IV)	-0.038	-0.044	0.039			
	(0.024)	(0.061)	(0.033)			
Age	YES	YES	YES	-0.005	-0.174	-0.001
				(0.012)	(0.063)***	(0.028)
Magnet*Age (Eligible*Age as IV)				-0.002	0.090	-0.037
				(0.018)	(0.077)	(0.037)
County-year fixed effects	YES	YES	YES	YES	YES	YES
Middle school fixed effects	YES	YES	YES	YES	YES	YES
4th order polynomial terms of high school entrance exam score	YES	YES	YES	YES	YES	YES
Observations	11448	7108	7108	11448	7108	7108
R-squared	0.22	0.30	0.24	0.22	0.30	0.24

Robust standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%