

Peer Quality and the Academic Benefits to Attending Better Schools *

Mark Hoekstra[†] Pierre Mouganie[‡] Yaojing Wang[§]

June 7, 2016

Abstract

Despite strong demand for attending high schools with better peers, there is mixed evidence on whether doing so improves academic outcomes. We estimate the cognitive returns to high school quality using administrative data on a high-stakes college entrance exam in China. To overcome selection bias, we use a regression discontinuity design that compares applicants barely above and below high school admission thresholds. Results indicate that while peer quality improves significantly across all sets of admission cutoffs, the only increase in performance occurs from attending Tier I high schools. Importantly, this increase in performance leads to a significant increase in students' access to higher education. Further evidence suggests that the academic gains to high school quality are driven by teacher quality, rather than peer quality or class size.

JEL Classification: I21, I24, I26, J24

Keywords: School Quality, Peer Quality, Teacher Quality, Regression Discontinuity, China

*We would like to thank Jason Lindo, Steven Puller, Lori Taylor, Miguel Urquiza, and Ron Zimmer for their helpful comments and suggestions. We also thank seminar participants at Columbia University and the Association for Education Finance and Policy (AEFP) meetings for helpful comments and discussions. All errors are our own.

[†]Texas A&M University, Department of Economics and NBER, *Email:* markhoekstra@tamu.edu

[‡]American University of Beirut, Department of Economics, *Email:* pm10@aub.edu.lb

[§]Texas A&M University, Department of Economics, *Email:* ywang@econmail.tamu.edu

1 Introduction

A common feature of educational systems around the world is that students sort into high schools and colleges on the basis of ability. In the United States, the sorting at the high school level takes place largely by families moving across neighborhoods and school attendance zones, while in college and in the much of the world the sorting is based on demonstrated academic performance. Across all of these contexts, students and their families exhibit strong revealed preferences for attending more selective high schools and colleges composed of higher achieving peers.

However, while recent research has documented significant returns to college quality (e.g., Andrews, Li, and Lovenheim, 2012; Canaan and Mouganie, 2015; Hoekstra, 2009; Saavedra 2009; Zimmerman, 2014), the literature on the returns to high school quality is less conclusive. Many recent studies find that attending middle and high schools with significantly higher-performing peers does not improve academic performance (Abdulkadiroglu, Angrist, and Pathak, 2014; Clark 2010; Dobbie and Fryer, 2014; Lucas and Mbiti, 2014; Zhang, forthcoming). In contrast, others find that attending more selective schools does result in benefits (Berkowitz and Hoekstra; 2011; Clark and Del Bono, 2014; Ding and Lehrer, 2007; Jackson, 2010; Park, Shi, Hsieh, and An, 2015; Pop-Eleches and Urquiola, 2013). In addition, in some cases the benefits are relatively modest. For example, Pop-Eleches and Urquiola (2013) find that students in Romania who attend schools with peers that are one standard deviation better score only 0.1 to 0.2 standard deviations higher on the national high school exit exam. Overall, the lack of consistent evidence of meaningful effects presents a puzzle. The purpose of this paper is to address this puzzle by examining the returns to high school quality across a range of high schools with different levels of selectivity, and in a context in which we can use measures of other school inputs such as class size and teacher quality to speak to potential mechanisms.

To do so, we apply a regression discontinuity design that exploits a unique institutional feature of the educational system in China. All students in China who wish to attend high school must sit for a national entrance exam, performance on which determines high school eligibility. While some students score barely above these cutoffs, others score just below. Intuitively, we compare the college entrance exam performance of these students to each other, which enables us to distinguish the effect of attending more selective high schools from unobserved confounding factors such as ability and motivation.

This approach has two primary advantages. The first is that the institutional context and administrative student-level data are ideally suited for providing credible estimates of the returns to high school quality. This is in large part because the high school admission

thresholds are determined only after students take the exam, making it difficult—if not impossible—for students to manipulate where they are relative to the cutoff. As a result, it is difficult to imagine a scenario in which students barely above and below the threshold are not otherwise similar to each other. In addition, the institutional framework we study provides for significant variation in high school quality; most students barely above the cutoff choose to attend the more selective school, and very few students just below the cutoff are able to attend that school. This contrasts with other contexts where student preferences and admission rules create much smaller discontinuities in attendance, which limits the interpretation of the estimates to a relatively small set of compliers. Finally, because of the high demand for attending four-year colleges, and because the college entrance exam is the determining factor for university admissions, students are highly incentivized to take the exam. As a result, there is little scope for selection into test-taking to bias the estimates. More importantly, because both teachers and students face strong incentives to do well on the college entrance exam, performance on it should be a good measure of whether students learn more at better schools. This contrasts with other settings, where observed performance outcomes may not be good measures of what students and teachers hope to achieve during high school.

The second main advantage of our study is that we are able to estimate gains in academic performance across a range of high schools that differ in quality but are within the same educational context. Thus, we are able to estimate returns not only for elite schools, but also for less selective schools in the same region, all of whom have discontinuously higher-achieving peers than the schools just below the cutoff. More importantly, this setting enables us to exploit differences across these cutoffs with respect to returns and education inputs in order to provide evidence on the importance of potential mechanisms. In particular, our data include measures of both class size and teacher quality as well as peer quality. Our measure of teacher quality is the concentration of “superior” teachers. This is the top rank of teachers in China, and the only one that cannot be earned based on credentials such as advanced degrees. Instead, it is based on rigorous evaluations of performance, a significant component of which is student performance on the college entrance exam. Data on these potential mechanisms turn out to be important, as the results we document are difficult to reconcile with the hypothesis that peer quality is responsible for returns to school quality. In addition, the fact that we are able to exploit the heterogeneity in returns and inputs within the same educational context reduces worries that any differences in returns are due to differences in behavioral responses, such as those documented by Pop-Eleches and Urquiola (2013).¹

¹The focus on mechanisms underlying the returns to high school quality is one difference between our

Results indicate that across the full range of high schools, there are few academic benefits to attending more selective high schools. Specifically, using a stacked RDD approach similar to that of Pop-Eleches and Urquiola (2013), we document that being barely admitted to a more selective school is associated with an average of a one-fifth standard deviation increase in peer quality. Similarly, we show that there are meaningful increases in peer quality across different admission thresholds throughout the range of high schools. However, we find no evidence that attending these schools with higher-ability peers leads to improved college entrance exam performance, on average. In contrast, when we focus on the return to attending elite Tier I schools, as other recent studies have done in the US (Abdulkadiroglu, Angrist, and Pathak, 2014; Dobbie and Fryer, 2014), we find significant returns. Specifically, we document that attending Tier I schools leads to a 0.16 standard deviation increase in exam performance. Given this exam is the primary determinant of admissions to universities in China, these gains lead to significant increases in students' ability to attend four-year colleges, which has been shown to have substantial returns in China (Giles, Park, and Wang; 2015).

Interpreted in the context of peer quality, these findings present a puzzle. That is, while we document that threshold-crossing is associated with significant increases in peer quality across all schools even outside of the Tier I threshold, the only returns come from attending Tier I rather than Tier II schools. We further document that these findings are difficult to reconcile even by the presence of non-linear peer effects, or by the possibility that only students at top schools are incentivized to do well on the exam. We do so by showing that while there is a significant discontinuity in peer quality across admission thresholds within Tier I schools, there is no evidence of improved performance.

Instead, we find that these results are most consistent with the hypothesis that returns to high school quality are caused by teacher quality, rather than peer quality. Specifically, we find that the only meaningful discontinuity in access to prestigious superior teachers is at the Tier I/Tier II threshold. In contrast, while there are large discontinuities in peer quality across all other thresholds—including within Tier I, within Tier II, and across Tier II and III—there are at most very small discontinuities in access to superior teachers across

paper and that of Park, Shi, Hsieh, and An (2015), who estimate the returns to attending magnet high schools in a different province of China. In addition, there are other important differences. Our data are from the Ministry of Education, while theirs are from schools contacted directly who agreed to share data. Perhaps as a result, we observe college entrance exam scores for 91 percent of students, which is consistent with the officially reported range using aggregate data, compared to 62 percent for Park, Shi, Hsieh, and An (2015). We also observe significantly higher compliance in our data for students barely above and below the threshold. In our sample, threshold-crossing at the Tier I cutoff is associated with a roughly 70 percentage point increase in the likelihood of attendance, which is twice as large as the increase in Park, Shi, Hsieh, and An (2015).

those same cutoffs. Similarly, we find no evidence that the pattern of results could be due to differences in class size; if anything, class size is larger for students who attend Tier I versus Tier II schools.

The finding that teacher quality, rather than peer quality, is likely responsible for returns to attending more selective schools is consistent with previous estimates on the value-added of superior teachers in China. Hannum and Park (2001) estimate that superior teachers improve test scores by 0.17 standard deviations relative to the lowest ranked teachers. By comparison, a back-of-the-envelope calculation suggests that if superior teachers in our context had a value-added that was 0.25 standard deviations better than average, the increased access to those teachers at Tier I schools would explain all of the positive return we estimate. Given the likelihood that teacher quality at Tier I schools may also be better in ways not measured by the superior teacher rank, we believe it is plausible that all the benefits are due to increased teacher quality. This is also consistent with recent evidence in the US highlighting the importance of teacher quality (Chetty, Friedman, and Rockoff, 2014), as well as with Jackson (2013), who finds that peer achievement can only explain a small part of the returns to selectivity in Trinidad and Tobago.

The results of this study may also help explain the mixed findings in the literature, all of which has documented significant increases in peer quality, but only some of which reports evidence of performance gains. The finding here of substantial differences in the returns to school quality within the same educational context suggests that there should be an increased focus on understanding and measuring *why* school quality matters. This is important because different mechanisms have substantially different policy implications. For example, if gains due to selective schooling were due to peer effects, there would be limited scope for enabling more students to benefit from school quality. On the other hand, if gains are driven by differences in teacher quality, then it may be possible to extend the benefits of attending better schools to more students, without reducing returns to others. Results in this study are more consistent with this latter interpretation, since the *only* positive returns to high school quality occur when there is also a significant increase in teacher quality.

2 The Chinese education system

2.1 Overview of Schooling in China

Children in China generally start elementary school at the age of six or seven. After spending six years in elementary school, children then move on to the first part of middle school, which lasts three years (7th to 9th grade) and completes the nine year national

compulsory education requirement. Graduates from junior middle school then choose to pursue either vocational or traditional schooling. Students who take the vocational track rarely go on to traditional colleges.² The traditional education path involves participating in the second part of middle school, which is equivalent to US high schools. Three years of high school are then followed by higher education (university/college) for those who are willing and able to do so.

In China, elementary and middle school education are both free and compulsory. On the other hand, high school education is neither compulsory nor free. However, in most parts of China, the majority of high schools are public and charge relatively low tuition. For example, in the metropolitan area we study, public high school costs around \$200 per year, and can be less if family income is below certain amounts. Around 60 percent of junior middle school graduates in our sample attend high school, while the rest attend vocational schools. Less than 5 percent of students attend private high schools, which are generally not as good as public schools.

There is vigorous competition amongst middle school students to enroll at the selective high schools, and admissions are most competitive at the highest-ranked high schools. Admission to high schools is based on a city-level entrance exam called the Zhongkao, or the HET, which is comprised of seven subjects. These subjects are Chinese language, Mathematics, English language, Physics, Chemistry, Political Science and Physical Education. The weighted sum of these seven subjects is the one and only criterion for high school admission for most students; the only (rare) exceptions are students with special talents, such as athletes. The HET is graded out of a possible 790 points.³

During high school, students usually choose an academic concentration (Arts or Science) at the beginning of their junior year (2nd year) in high school. Some college majors only admit students from one path and others accept both, so this choice can be a combination of personal preference and comparative advantage.

Importantly, other than sorting into different classes by academic concentration, there seems to be relatively little other sorting within high schools. While we do not have data on the extent to which there is sorting by teacher and student ability within a high school, anecdotally it seems to be the case that the only sorting that can sometimes occur is into “special talent classrooms” where additional resources are targeted toward the top 15 percent

²This is in part because the vocational track curriculum does not prepare students for the college entrance test, which is required for admission to traditional colleges. A different test called the “3+Certificate” is required for admission to vocational colleges.

³Chinese, Math and English are each graded out of a possible 150 points, while Politics, Physics and Chemistry are each graded out of 100 points. Physical Education is graded out of a possible 40 points.

of the students in some schools.⁴ In those cases, the other 85 percent of students and teachers are evenly distributed into classrooms with respect to performance.

Similar to the high school admission process, university admission decisions are made almost entirely on the basis of performance on the college entrance exam, called the Gaokao or the CET. This exam is taken after three years of high school, and is required of all students who wish to attend college. High school students concentrating in arts take an exam that includes Chinese language, Mathematics for arts students, English language and a comprehensive arts test consisting of Political Science, History, and Geography. Students concentrating in sciences take an exam that includes Chinese language, Mathematics for science students, English language and a comprehensive science test consisting of Physics, Chemistry, and Biology. The exam for both tracks is graded out of a possible 750 points.⁵ In contrast to high school, students in China are free to attend any university that accepts students from that province—regardless of location—conditional on meeting that university’s threshold score. In addition, eligibility to attend any four-year college in China is determined by specific thresholds set by each province. As a result, students are heavily incentivized to perform well on the CET. Indeed, the desire to do well on this exam is the main reason for the competitive admissions process into high schools, as students hope to position themselves to do well on the college entrance exam and thus attend a selective university.

2.2 High School Choice Mechanism

High school admissions for the metropolitan area we study is centrally operated by each city’s education administrators. In early June, students fill out application forms indicating their ordered preference of high schools. These students then take the high school entrance exam in mid June. High schools predetermine how many students they wish to admit for that year and grant admission based on students’ preferences and test scores. Most metropolitan areas, including ours, use an admission procedure similar to the Boston Mechanism. In the first round of admissions, each high school only considers students who list them as their first choice. Students with HET scores above a certain threshold are accepted and the rest are rejected and placed in a pool of candidates to be considered by the next high school on a student’s list. Only in the event that a high school still has any remaining slots after the first round will it consider admitting students who list them as their second or third choice. Once a student is granted admission by any high school, the selection process ends for that

⁴For example, see <http://zhongkao.gaofen.com/article/448289.htm>.

⁵For the science track, Chinese, Mathematics for the sciences, and English are each graded out of 150 points, while the science comprehensive test is graded out of 300 points.

For the arts track, Chinese, Mathematics for the arts, and English are each graded out of 150 points, while the arts comprehensive test is graded out of 300 points.

student and he/she is not to be considered by any other high school.

For illustration, suppose school A plans to enroll 100 students for that academic year. Further, suppose that there are 80 students who indicate their first preference is to join that school. School A will first admit those 80 students, then proceed to rank students who listed A as their second choice—conditional on not yet being admitted by their first choice. If there are more than 20 of those applicants, school A will take the 20 highest scoring students. If admission slots remain, then School A proceeds to fill the rest of their seats with students who list A as their third choice, and so forth. In the more likely scenario that there are more than 100 students who list School A as their first preference, officials select the highest scoring 100 students. The lowest admitted student’s score—regardless of preference order—is the official cutoff score for school A for that year. High schools go through this process simultaneously, as each student can be admitted by at most one school.⁶

To ensure smoother and more transparent school-student matching, schools are divided into four groups by the city education department. These groups are defined in advance of student applications, and the groupings are made public to all students. We call these groups “tiers” as they divide the schools with respect to quality/selectivity. The best schools are called Tier I schools, the second-best are Tier II, and so on. The composition of each tier is quite stable over time, though sometimes schools change tiers from one year to another to reflect changes in quality. In addition, there are also differences in school selectivity within tiers as well as across tiers. However, the provincial level education bureau sets the curriculum and textbooks for all public schools, independent of tier, and the number of classes during a day is similar across all schools.⁷

⁶Public high schools in our sample are allowed to designate around 10% of their seats as “high priced”. Students enrolled through the high-priced channel pay a one-time fee to the school upon registration, though they receive the same education as the other regular students. This one-time fee is set by the schools and revealed to students before they apply. In urban areas it is usually around 40,000 Yuan (6,600 USD), while in suburban areas it is around 20,000 Yuan (3,300 USD). Schools allocate these high-price slots in a separate but otherwise similar process as that used to allocate the other slots. For example, suppose school A plans to set 90 regular seats and 10 high-priced seats. Then A (regular) and A (high-priced) independently go through the high school admissions process as described above. Students decide which schools, regular or high-priced, to apply at the same time, before the test. Students can even apply to both regular and high-priced of the same school. Thus, all schools with high-priced seats, will have two cutoffs—one for regular students and one for high priced students—and this information is released to the public by both education officials and the media. In our analysis, we keep in the sample all individuals who entered high school through this “high priced” process, though those students are likely “non-compliers” and thus contribute little to the variation we use to identify effects. We do not exclude them because doing so would potentially create imbalances in the composition of students on either side of the cutoff. That is, while a student barely above the traditional cutoff who attends that school would remain in the sample, her nearly identical counterpart who is barely below the traditional cutoff but who is above the high-price cutoff for that school would be excluded, thereby invalidating the design. However, in practice this does not seem to be an issue; in results available upon request we find very similar results even when excluding these students.

⁷Some schools may choose to keep their students longer after hours, such as nights or weekends, par-

All schools are also ranked nationally according to the following designations (from best to worst): National demonstrative high schools (“Guojia Shifanxing Gaozhong”), Provincial first class schools (“Shengyiji Xuexiao”), Municipal first class schools (“Shiyiji Xuexiao”), District level first class schools (“Quyiji Xuexiao”) and ordinary schools (“Putong Zhongxue”). All Tier I schools in our sample are designated as national demonstrative high schools. This ranking system was introduced by the Ministry of Education during the State’s ninth “Five-Year Plan” period (1996-2000). To earn this title of “National Demonstrative High School, a school must meet certain criteria regarding curriculum design, school facilities, teacher quality and student performance.⁸

Schools in the first tier begin the admissions process. After Tier I schools fill all their seats, Tier II schools will start admitting students, then Tiers III and IV. Accordingly, students list their preferences by tier. For each tier, a student has four ordered school choices. Importantly, because there are fewer than four Tier I schools in the districts we analyze in this study, students are able to list and rank each of the Tier I schools. The order of choice is important as most competitive schools fill their slots solely with students who have them listed as their first choice. Students generally understand this and as a result most list their preferences by perception of school quality.

School choice is different for students from different parts of the city we analyze. Specifically, the city is divided geographically into twelve administrative districts, which define the region in which students have choice regarding high school. Of the twelve districts, eight of them are mostly urban areas and are geographically small. Students from these eight districts can go to high schools in their own district but also have access to schools in the other seven urban districts. Specifically, a student residing in an urban area can choose from almost all urban Tier I schools—regardless of district—in addition to schools in Tier II to IV from their own district. On the other hand, students from suburban districts on average have a choice set of only two Tier I schools and 11 Tier II through IV schools. Further, students residing in the four suburban districts can only choose high schools in their own district. As a result of the more limited choice sets facing students in the four suburban districts, the admission system generates much more significant discontinuities with respect to school selectivity and ability levels. The students in these districts also have much more uniform preferences over school quality, given the significant differences across the limited

ticularly in preparation for the college entrance exam. While we are unable to acquire data on this, our understanding is that the practice is quite common across schools, and thus is unlikely to differ across tiers in a discontinuous way.

⁸For example, according to the national demonstrative school assessment scheme on the provincial department of education website, at least 30% of the teachers must have either a graduate degree or superior teacher title; student crime rate must be lower than 1%; and at least 25% of students must meet the provincial elite college cutoff in the CET and 60% must meet the four-year college cutoff.

set of schools. For these reasons, we restrict our analysis to these suburban districts.

2.3 Type of teachers in high school

A unique and important aspect of high school education in China is the clear distinction of teachers by rank. There are different titles (ranks) for high school teachers, and salaries increase with these ranks. The three professional ranks for all public school teachers, regardless of class level are “elementary”, “intermediate” and “superior”. Further, within the intermediate rank, there are two smaller categories; “second class” and “first class”.

One automatically becomes an “elementary” rank teacher upon employment in the teaching sector. However, if that person holds a master’s degree, then they start at the intermediate second class rank. Teachers with a doctoral degree start at the intermediate first class level. After two years as an elementary rank teacher, a teacher then applies for promotion to the intermediate second class level. After four years within this rank, they are able to apply for the intermediate first class rank. Finally, after achieving a first class rank and after a period of no less than five years,⁹ a teacher is permitted to apply for the superior teacher rank.

After a teacher applies for promotion, and after the approval of the school they work for, city education officials put together a committee to start an evaluation process assessing a teacher’s performance along several dimensions such as teaching, publication level, integrity, and various other aspects in his/her field. A committee will quantitatively grade a candidate on these aspects.¹⁰ For example, having a PhD degree is worth 3 points, a Masters degree is worth 2 points, and a Bachelor’s degree is worth 1 point. There are five categories and 100 total points: Degree (3 points), Tenure (7 points), Experience in Current Position (22 points), Performance in Current Position (38 points), Research Papers (10 points) and Awards and Contribution (20 points).¹¹ Within the Performance category, there are four sub-categories, one of which is based on teaching outcomes. In this subcategory, a candidate can score up to 13 points if their students’ average test scores are high (8 points), their students improve a lot on a particular subject (3 points) and they form a unique and effective teaching style (2 points). The assessment ends with an oral exam. If more than 2/3 of the committee members vote for approval, then a teacher will be approved for the promotion. Similar to tenure in the U.S., once a teacher is promoted to a higher rank, they generally do not get demoted. For example, in our sample during the time period we study, no teachers were

⁹Master degree holders can apply for promotion to first class rank after two years instead of the usual four. Doctorate degree holders only need to have two years of teaching experience to be eligible for superior teacher promotion.

¹⁰http://www.gzedupg.com/zhicheng_detail.php?pid=10&id=398.

¹¹The assessment table can be found at: <http://www.gzedupg.com/download/sjh2008362fb2.xls>

demoted for failing to meet certain assessment tests.

Salaries differ by rank, so teachers have an incentive to get promoted. Teacher salaries in China consist of two main parts: 1) state (base salary) and 2) local (city, district and school level). The local part of teacher salary varies extensively from city to city and even from school to school and can be based on performance. In addition, more selective schools often pay more, and thus can recruit higher quality teachers. The state base salary is determined by one's professional rank and title ("Gangwei Gonzi") as well as years of service ("Xinji Gonzi") and has a nationwide set of standards. For instance, superior teachers receive an additional "Gangwei Gonzi" salary of 930 RMB (\$150) to 1180 RMB (\$190) per month. On the other hand, first class teachers receive an additional "Gangwei Gonzi" salary of 680 RMB (\$110) to 780 RMB (\$126) per month. Superior teachers' "Xinji Gonzi" portion of their salary starts from level 16 (317 RMB per month), while first class teachers start from level 9 (181 RMB per month). As a result, promotion to superior rank from first class results in a base salary increase of at least 19 percent, and as much as 51 percent.

An important question is whether teachers of higher rank in China have higher value-added when it comes to the college entrance test scores of their students. While our understanding of the promotional process leads us to believe that this would likely be the case in particular for superior-rank teachers, to our knowledge there are two empirical studies that speak directly to this question. Using lottery data from Beijing middle schools, Lai, Sadoulet, and de Janvry (2011) show that teacher rank is highly correlated with estimated school fixed effects, suggesting that a significant part of school quality is due to teacher rank. In addition, Hannum and Park (2001) find that teachers with superior rank increase math and language test scores each year by 0.08 and 0.25 standard deviations, respectively.¹² They conclude that the teacher quality ranks reveal significant information about teacher quality that is not contained by measures such as the teacher's degree attainment and years of experience.

3 Data

We use student-level administrative data from a large capital city of a densely populated province of more than 7,000 square kilometers in China. As a condition of using the data, we are unable to reveal the name of the province and city. The city has a population of

¹²This calculation is based on results reported in Tables 6A (math) and 6B (language). In column 3 of each table, they report the coefficient on "Teacher Qualification 2", which corresponds to superior teachers in our setting. Those coefficients are 0.14 and 0.44, and are interpreted as the exposure to an average 1.79 years of teaching (page 23). Thus, rescaling each estimate by 1.79 to recover a per-year effect gives us estimates of 0.08 and 0.25

more than 10 million and a per capita GDP of more than \$20,000. The two suburban areas we study in this paper have a total population of more than 2 million. GDP per capita is around \$16,000, which is lower than the urban part of the city but still higher than the national average.

Our data come from the education bureau authorities of the city. The authorities merged student data of those who took the High School Entrance Test (HET) and attended one of the traditional high schools in 2007 with those who took the College Entrance Test (CET) in 2010, resulting in a sample size of 49,674 students.¹³ For each student, we observe both their HET and CET scores and some student characteristics including the middle school and high school attended, gender, age, middle and high school district, and parents' occupations. Because some high school students do not take the CET, in Section 5.3 we test for selection into taking this exam and perform bounding exercises to ensure our estimates are not biased by selection into test-taking.

Our data only contain individuals attending traditional high school since those attending vocational schooling are not generally eligible to take the college entrance examination. Further, we restrict our sample to suburban districts, where students have a limited choice set compared to students in urban districts. In addition, our analysis focuses on the two suburban districts that have at least one school that is exclusively Tier I.¹⁴ Our final sample consists of 15,367 students taking the high school entrance exam (HET) in the year 2007 and the college entrance exam (CET) in 2010.

Data on school and teacher characteristics were collected from government reports and official school websites, as well as (in rare cases) recruitment pamphlets for the few schools that do not have websites. These data include the size of the schools, which include the geographic size of the school, the number of students, classes, teachers and superior teachers. We link these data to our student data using school identifiers.

Within the two suburban areas we analyze, students generally have at most 15 high schools to choose from. The tier of each school is clear and widely known. The main determinant of which high school is attended is the score on the high school entrance exam. We observe detailed administrative data on test scores by subject and the eventual high

¹³According to official records, a total of 59,591 students registered for the CET in this city in 2010, which includes the current high school seniors as well as those who already graduated and wished to take the test again.

¹⁴There are four suburban areas where students must attend high schools within the same district. One of them does not have any Tier I schools, while another has five high schools that are simultaneously classified as Tier I and Tier II. Because much of our analysis is focused on the returns to Tier I schools, we leave both of these districts out of the sample, though later on we perform robustness checks showing that our estimates are largely unchanged when we include the latter district, regardless of how we classify those five Tier I/Tier II schools.

school attended by the student. As a result, we are able to measure school peer quality by calculating the average score on the HET for students in each school.

The main outcome of interest is a student's total score on the college entrance exam. In addition, we also examine eligibility to attend a four year college. We can do so because eligibility for entry into a four year college is centrally determined by whether a student crosses the lowest threshold score imposed by a specific university. This threshold is common to all students in a province regardless of which city they reside in within the province. As a result, while we do not observe the university a student ultimately attends, we are able to determine whether students are eligible to enter any four year college using their final CET scores.

Descriptive statistics for all students who sat for the 2007 high school entrance exam are reported in Table 1. These statistics are reported for the whole sample and by high school tier.¹⁵ The average scores on the HET and CET are 615 and 487 points, with standard deviations of 59 and 100, respectively. These scores increase with the level of tier as one would expect. Just over half of the high school students (53%) are female. For the full sample, 48% of the students choose to major in arts, though that figure ranges from 35% in Tier I to 66% in the last two tiers. Very few students attend private high schools (1%). 42% of students in our sample are eligible to go to a four year college. This number is as high as 81% for students attending Tier I high schools and drops to 26% and 5% for Tiers II to IV. Further, students eligible to attend an elite college are almost exclusively composed of Tier I high school graduates. Higher tier schools tend to be larger in size. Class size also tends to be slightly larger; Tier I schools have an average of 55 students per class compared to 53 and 51 students per class for Tier II and Tier III/IV schools respectively. More selective schools have a significantly higher superior teacher ratio; 38% of teachers in Tier I schools are superior, compared to 16% for Tier II schools and 7% for the lowest two tiers. Finally, 55% of all teachers in our sample are female, which is roughly constant across tiers.

4 Identification Strategy

4.1 Single Cutoff: The Academic Return to Attending Tier I Schools

As mentioned earlier, the high schools we are analyzing are divided into four tiers with the first tier containing the best set of high schools within a district or town. Accordingly,

¹⁵We only have one Tier IV school in our district. As a result, we combine the summary statistics for Tier III and Tier IV schools.

we use a regression discontinuity design (Lee and Lemieux, 2010; Imbens and Lemieux, 2008) to estimate the causal impact of elite high school attendance (defined as going to a Tier I high school) on college entrance exam scores and college attendance. The key identifying assumption underlying an RD design is that all determinants of future outcomes vary smoothly across the Tier I high school admissions threshold. This is likely to hold, as precisely manipulating the overall exam score would be extremely difficult, if not impossible. This is because the cutoff scores for each high school are only determined after the exams are administered and graded. These cutoffs are determined based on high school applicants' percentile ranks, which are only calculated after the tests are graded. As a result, students and graders do not know where the admission thresholds for each school lie until after the test is taken and graded. In addition, graders do not observe any identifying information on students, so it is not possible for them to artificially increase the grades of certain students.

All students in our data attend a school in one of two suburban towns.¹⁶ Accordingly, we have two Tier I cutoffs in our data—one for each town. In order to summarize the effects of attending an elite school, we pool data across both towns. Formally, we estimate the following reduced-form equation:

$$Y_i = \alpha + h(S_i) + \tau D_i + \delta X_i + \epsilon_i, \quad (1)$$

Where the dependent variable Y is the outcome of interest. D is a dummy variable indicating whether a student i crosses the town-specific score threshold for attending a Tier I high school.¹⁷ S represents student high school entrance test (HET) scores in 2007 measured in points relative to the cutoff score of each town. Formally, $S_i = grade_i - \overline{grade}_z$ for all individuals within a town facing a common Tier I cutoff z . The function $h(\cdot)$ captures the underlying relationship between the running variable and the dependent variable. We also allow the slopes of the fitted lines to differ on either side of the admissions threshold by interacting $h(\cdot)$ with the treatment dummy D . X is a vector of controls that should improve precision by reducing residual variation in the outcome variable, but should not significantly change the treatment estimate if our identifying assumption holds. The term ϵ represents unobservable factors affecting outcomes. Finally, the parameter τ gives us the average effect of having the opportunity to access a Tier I high school for each outcome of interest.

In our analysis, we specify $h(\cdot)$ to be a linear function of S and estimate the equation over a narrower range of data, using local linear regressions with a uniform kernel. This approach can be viewed as generating estimates that are more local to the threshold and does not impose any strong functional assumptions on the data. As a result, the preferred specifi-

¹⁶As mentioned earlier, students may only attend high school in the suburban town in which they reside.

¹⁷We have two thresholds, with each representing a different town.

cations in this paper are drawn from local linear regressions with the optimal bandwidths chosen by a robust data driven procedure as outlined in Calonico, Cattaneo and Titiunik (2014)—henceforth CCT. We also present results for a variety of bandwidths relative to the optimal bandwidth as has become standard in the RD literature (Lee and Lemieux, 2010). Further, standard errors are clustered at the high school score (HET) level, as suggested in Lee and Card (2008).

4.2 Multiple Cutoffs: The Academic Return to Attending Better Schools

Within each of the two suburban towns in our sample, we rank schools according to their posted admissions cutoff score for that year (2007). This yields 23 quasi-experiments as each cutoff results in a potential RD analysis.¹⁸ Following Pop-Eleches and Urquiola (2013), we focus on regressions that pool data across all school entry cutoffs. Specifically, we stack the data such that each student within a certain town serves as a separate observation for each cutoff.¹⁹ Due to repeated observations, we cluster our standard errors at the individual level. Formally, our reduced form regression from this procedure takes the following form:

$$Y_{iz} = \alpha + h(S_{iz}) + \omega D_{iz} + \phi X_i + \epsilon_i \quad (2)$$

Here the subscript i still refers to students and the subscript z refers to all possible high school cutoffs facing an individual within a town (i.e. $z = 1, \dots, H-1$; where H represents the total number of high schools in that town ordered from worst to best based on their respective cutoff scores). ω gives us the ITT estimate of having the opportunity to go to a better school, regardless of tier. Further, the running variable is defined as $S_{iz} = grade_{iz} - \overline{grade}_z$ for all individuals within a town facing numerous cutoffs z. As a result, equation (2) takes the same form as equation (1) except for the fact that each individual can be observed multiple times depending on his/her relative position to a high school cutoff. However, regressions restricted to students scoring close to the cutoffs rarely use student-level observations more than once.

¹⁸We have an average of around 12 schools for each town resulting in 11 different cutoffs within each town.

¹⁹For instance, our smallest town has 12 different schools, leading to 11 separate cutoffs. This town also contains 4,025 students. For that town, our procedure produces a dataset of $(4,025 \times 12) 48,300$ observations.

4.3 Tests of Identification

As described above, given the nature of the school assignment mechanism and the way in which it is implemented, we find it unlikely that students are able to manipulate the assignment variable in such a way that would invalidate the research design. However, we still provide empirical tests in order to assess whether the data appear consistent with the identifying assumption that no other determinants of achievement vary discontinuously across the threshold.

First, we ask whether there is any evidence of bunching around the admission threshold. Under our identifying assumption, there should be no such bunching. In contrast, if students or graders could manipulate scores relative to the cutoff, we might expect to see too few students just short of the cutoff, and too many students barely exceeding the cutoff.

Results are shown in Panels A and B of Figure 1, which show the density function for the stacked RDD across all admission cutoffs as well as for only the Tier I admission threshold. Both show no evidence of bunching around the cutoff, consistent with the identifying assumption.

In addition, we also test whether observed determinants of achievement are smooth across the threshold. If the identifying assumption holds, we expect all such variables to vary smoothly across the admission thresholds. On the other hand, if students or graders are able to manipulate scores around the cutoff, then we might expect to see evidence of different types of students on either side of the cutoff. Covariates in our data set include age, gender, and district and middle school fixed effects.

Rather than focusing on these covariates individually, we instead use those covariates to predict college entrance test scores for each student. We then ask whether those predicted scores are smooth across the cutoff.²⁰ We do this in part because using this weighted average of characteristics corresponds most closely to what we care about - whether underlying ability to do well on the college entrance exam varies smoothly across the cutoff. In addition, the predicted performance measure can more easily quantify the role of middle schools and district attended by the students.

Results are shown in Panel C of Figure 1, and indicate that there is little evidence that underlying student ability varies discontinuously across the threshold. Estimates shown in Appendix Table A1 are also close to zero and statistically insignificant across a range of bandwidths.

²⁰In Table A1, we also show estimates for age and gender separately.

5 Results

5.1 Effects of School Quality Across All Admission Thresholds

We begin by examining the impact of attending better schools using all of the admission thresholds in our data. Specifically, we seek to document that threshold-crossing is associated with increases in peer quality, and then ask whether threshold-crossing leads to improved performance on the high school exit exam.

The results are shown graphically in Figure 2. These figures take the same form as those after them in that open circles represent local averages of the outcome over a 4 point score range. We show results using a bandwidth of 50 points on either side of the cutoff. The running variable is defined as the number of points above the admission threshold. Consequently, a value of zero on the x-axis implies that the student barely met the admission threshold for the school.

Figure 2A shows results for peer quality, defined as the average high school entrance exam score of students in the school in which the student enrolled. Using a local quadratic fit and a bandwidth of 50, we find that average peer quality significantly increases at the threshold. Specifically, threshold crossing leads to an improvement in peer quality of 13.5 percent of a standard deviation. Thus, there appears to be compelling visual evidence that threshold-crossing does lead students to attend “better” schools, with higher-performing peers. While this relationship is deterministic given the way in which admission decisions are made, it does reflect that given the opportunity, on average students choose to enroll in schools with higher-achieving peers.

Corresponding regression estimates are shown in Panel A of Table 2. Results are shown using bandwidths ranging from three-quarters of the optimal bandwidth to 2.5 times that optimal bandwidth, where optimal CCT bandwidth was calculated as 14 points. Estimates are also shown with and without controls. Estimates for the effect of threshold-crossing on peer quality range from 0.15 to 0.2 standard deviations; all estimates are statistically significant at the one percent level.

Figure 2B shows results for the main outcome of interest, the college entrance test score. This score is far and away the main determinant of whether a student is able to attend a four-year college, and how selective that four-year college will be. Results indicate that even though students barely above the cutoff attend significantly better schools, they do not achieve at higher levels as a result. Estimates across a range of bandwidths and specifications in Panel B of Table 2 range from -0.017 to 0.006 of a standard deviation in CET scores. None of the estimates are statistically significant at conventional levels.

However, one might be concerned that average scores may not reflect benefits to attending

better schools if students and teachers are aiming to improve scores primarily over one part of the distribution. Since an important goal of many students is to earn a score high enough to gain entry into a four-year college, we focus on an outcome that measures whether the college entrance exam score achieved exceeded the cutoff for attending four-year college in the province. Results are shown in Figure 2C and indicate that attending better schools does not lead to improved access to four-year colleges. Corresponding estimates in Panel C of Table 2 are similar. In short, there is little evidence that attending better schools improves cognitive ability or college attendance, on average. In addition, we also test for heterogeneity by gender. Results are shown in Appendix Figure A2, and indicate that while peer quality across the threshold is higher for both boys and girls, neither group experiences a cognitive return or increase in college attendance.

5.2 Effects of Attending Elite Tier I Schools

Given that much of the recent literature has focused on the returns to attending *elite* high schools (e.g., Abdulkadiroglu, Angrist, and Pathak, 2014; Dobbie and Fryer, 2014), we now turn to examining the returns to attending elite schools in our sample. Specifically, we examine the returns to attending Tier I, relative to Tier II schools. This cutoff is one of several used to identify effects in the previous section.

Results are shown graphically in Figure 3. Panel A shows the likelihood of attending a Tier I high school for those just above and just below the admission threshold. Results indicate that while only around 10 percent of applicants just below the cutoff attend Tier I schools, more than 70 percent of those just above the cutoff attend Tier I schools. We note that the likely reason some students (i.e., noncompliers) are able to attend despite missing the cutoff is due to the high price admission slots allocated by the schools, as well as exceptions to the admission policy granted to some applicants such as athletes. Corresponding local linear estimates in Panel A of Table 3 range from 63 to 67 percentage points; all estimates are significant at the 1 percent level.

Panel B of Figure 3 shows that threshold-crossing leads to significant increases in peer ability of approximately 0.29 standard deviations of the college entrance exam. This reflects that on average Tier I schools are attended by significantly higher ability peers, though the schools may also be better in other ways (we return to this issue later). Corresponding regression estimates shown in Panel B of Table 3 range from 0.30 to 0.37 standard deviations, all of which are significant at the 1 percent level. Thus, our results indicate that being eligible to attend a Tier I school results in roughly a 66 percentage point increase in the likelihood of attending a Tier I school, and an increase in peer quality of one-third of a standard deviation.

Importantly, estimates for both the likelihood of attending Tier I schools and peer quality are nearly unchanged when adding controls measuring age, gender, and district and middle school fixed effects.

Panel C of Figure 3 shows that in contrast to the results across all admission thresholds, being eligible to attend an elite Tier I high school leads to a nearly one-tenth standard deviation increase in achievement on the college entrance test. Corresponding regression estimates are shown in Panel C of Table 3. The smallest estimate is that for the narrowest bandwidth (0.75 of the optimal bandwidth), which is 0.07 standard deviations and is significant at the 5 percent level. Estimates for bandwidths between 1 and 2.5 times optimal bandwidth range from 0.07 to 0.09 standard deviations and are all significant at the 1 percent level. The addition of controls does not affect estimates in a meaningful way, consistent with the identifying assumption.

Panel D of Figure 3 shows that this increase in the college entrance test scores also results in increased eligibility to attend four-year colleges. Estimates in Panel D of Table 3 range from 5 to 14 percentage points, though only estimates for larger bandwidths are statistically significant at conventional levels.

While we do not have student-level data on long-run outcomes such as college attendance, aggregate data suggest that eligible students enroll at four-year colleges at high rates. Specifically, data from our province indicate that 63 percent of students who exceeded the eligibility threshold enrolled at four-year colleges in the province. This has important implications for long-term outcomes; Giles, Park, and Wang (2015) estimate a 37 percent return to attending four year college in China. Similarly, Li, Liu, Ma, and Zhang (2005) estimate the per-year return to attending college is as high as 10 percent. Consequently, while we lack the data to estimate the long-term returns directly, the existing literature suggests that the long-run gains to attending Tier I schools are significant.

In addition, in Figure 4 we investigate whether returns to attending Tier I schools are different for boys than for girls. The results are quite striking; while being barely eligible for Tier I schools leads both boys and girls to attend schools with significantly higher-performing peers, only boys experience a cognitive return. Unfortunately, it is difficult for us to ascertain why this is, though we return to the issue in the next section when we discuss explanations for the overall pattern of results.

Finally, we can also report local average treatment effects of attending Tier I schools by rescaling the intent-to-treat estimates by the estimated discontinuity in the likelihood of attending a Tier I school across the admission threshold. Results are shown in Table 4. Both males and females are roughly 60 to 65 percentage points more likely to attend a Tier I school if they are (barely) across the threshold. This speaks to the strong revealed preference for

attending more selective schools in this context. In addition, Panel B of Table 4 shows both intent-to-treat and local average treatment estimates of the difference in peer quality across the cutoff. Specifically, we estimate that attending Tier I schools results in an increase in peer quality of 0.48 standard deviations for boys and girls. Similarly, results indicate that Tier I school attendance leads to a 0.3 standard deviation increase in CET scores for boys, and increases their eligibility to attend a four-year college by 19 percentage points, with the effects remaining statistically insignificant for girls.

In summary, our analysis yields two findings. First, across all admission thresholds, while being barely admitted results in enrollment at schools with significantly better peers, it does not result in cognitive returns, on average. Second, admission at Tier I schools leads students to attend significantly better schools, which *does* improve cognitive outcomes, a return driven by boys. Importantly, we can reject the null hypothesis that these effects are equal.²¹ These two apparently contradictory findings present a puzzle similar to that in the existing literature, which has documented mixed findings with respect to returns to high school quality. Thus, while the next section tests the robustness of these findings, after that we return to the question of why there are returns to Tier I high schools in China, but not to other “better” schools.

5.3 Threats to Identification

One potential threat to identification is if attending better schools leads students to select a different academic track, a decision that is made in the second year of high school. For example, if (barely) going to a Tier I high school increases the likelihood of a student choosing a scientific track, then that difference, rather than a broader sense of improved school quality, could drive our results. To test for this, we check whether the probability of choosing an arts versus science track is discontinuous at the threshold for attending a Tier I high school. Results are shown in Appendix Figure A3a and A3b. Both figures show that the likelihood of majoring in arts versus science is smooth across all admission thresholds (Figure A3a) as well as the cutoff for Tier I schools (Figure A3b).²²

In addition, we also test whether differential grade repetition across the admission cutoff could bias our estimates. For example, if Tier I high schools were more likely to have their

²¹For example, the estimate using optimal bandwidth with controls in column 5 of Table 2 indicates that a 0.187 standard deviation increase in peer quality is associated with a -0.014 standard deviation increase in CET scores. Rescaling these estimates (and standard errors) by a factor of two results in a change in peer quality roughly equivalent to the 0.380 increase across the Tier I threshold shown in Panel B of Table 3. However, the rescaled upper bound of the 95 percent confidence interval for the estimate in Panel B of Table 2 is only 0.047, which is considerably lower than the estimate of 0.073 in column 5 of Panel B in Table 3.

²²In results available upon request, we also check whether the likelihood of majoring in arts varies by gender. The results remain unchanged.

worst students repeat a grade, then perhaps the improvement in CET scores we document could be due to age or quantity of schooling, rather than school quality. While grade repetition is uncommon in China, we test explicitly for this explanation by examining whether exact age at the time of taking the CET is smooth across the admission threshold.²³ Results are shown in Appendix Figures A3c and A3d, which show that age is smooth across both cutoffs, indicating that grade repetition is unlikely to explain our findings.

Perhaps a more worrisome potential source of bias is the possibility of selection into taking the college entrance exam. That is, if barely being admitted to a better school (or an elite school) made it more or less likely for the student to take the college entrance test, then our estimates could be biased. We address this concern by first testing for selection into test-taking, and then using those estimates to bound our estimates. We begin by matching our dataset to data containing the population of high school test takers, regardless of whether they took the high school entrance exam. In this way, students who do not match are identified as not having taken the CET. Our match rate is high; we estimate that 91 percent of students entering high school sat for the CET exam. This is in line with official aggregate data for the districts in our sample, which indicate that 90 to 94 percent of students take the CET exam over the years we study.

We then examine whether going to a better quality high school leads students to take the CET at different rates. Results are shown in Figure 5. Results indicate that while going to *any* better school is not associated with differential college entrance test-taking, barely attending an elite high school does appear to lead to a higher rate of test-taking. Additional results in Figure 5f indicate that this increased test-taking is driven by girls; in contrast, the rate of test-taking is constant across the admission threshold for boys (Figure 5d).²⁴ Corresponding estimates in Table 5 yield the same conclusion: while there is no evidence of selection into test-taking for boys and girls when looking at better schools overall, students who are barely admitted to elite schools are one to two percentage points more likely to take the CET, though these estimates are not statistically significant across all bandwidths.

The simplest explanation for this finding is that some marginal students are induced to take the CET when barely attending Tier I schools, when they would not have if they had attended lower-tier schools. This would likely work against our finding that attending Tier I schools leads to improved CET performance. In addition, we note that the positive returns to

²³Grade repetition is rare in part because students are not allowed to repeat their senior year of high school. For other years, in order to repeat a year a student must fail three classes after taking a make-up exam and must gain the approval of school and city-level administrators.

²⁴This could also potentially explain results from the previous section indicating that the likelihood of observing a female in the sample is higher at the Tier I cutoff for some bandwidths. See Appendix Table A1.

attending Tier I schools were driven by boys, while Figure 5d shows no evidence of selection into test-taking for boys. This provides additional comfort that the type of selection into test-taking that we observe cannot explain our findings.

Nevertheless, we perform formal bounding exercises to assess the degree to which this selection into test-taking could affect our results. Specifically, we use the trimming procedure suggested by Lee (2009). The intuition behind this test is as follows. To find a lower bound (worst case scenario) for the estimated impact of treatment on college exam scores, we assume that only the best students attending the most selective high schools, who would have otherwise dropped out, select into the exam. Thus, dropping the top distribution of the treatment group makes it comparable to the control group. Formally, we drop the top distribution of students within each bin (i.e within a bandwidth of 4 HET points).²⁵ Further, the share of students to be trimmed from each bin in the treatment group is derived from the treatment estimate of the likelihood of selecting into the college entrance exam.²⁶ A similar procedure—trimming the bottom performing students in the treatment group—is used to estimate the upper bound.

Table 6 summarizes the updated local linear regressions by comparing previous college test score RD estimates with those estimated using the trimming analysis. Bootstrapped standard errors are reported in parentheses for the lower and upper bound estimates. For consistency and comparability, we use the same bandwidths as the college entrance exam score regressions in Table 3. We present lower and upper bound estimates for all bandwidths. Results indicate that the lower and upper bound estimates of the return to attending an elite high school remain positive and significant, and range from 5 to 12 percent of a standard deviation, compared to original estimates ranging from 7 to 9 percent of a standard deviation. For example, bounds corresponding to the estimate using optimal bandwidth are shown in Column 5. Our primary estimate was 0.073 standard deviations, while the estimated lower and upper bounds are 0.054 and 0.101 standard deviations, respectively, both of which are statistically significant at the 5 percent level. Thus, we conclude that selection into the CET does not bias our results in a meaningful way.

²⁵In results available upon request, we also run this exercise by dropping the top distribution of students in the treatment group, regardless of distance to threshold. The results remain statistically similar.

²⁶For example, using a local linear regression of bandwidth =50, we estimate that students are 1.69 percentage points more likely to select into the college entrance exam at the cutoff. To estimate the total percent of students to be trimmed, we merely divide 1.69 by the mean proportion of test takers for the control group at the threshold (90.5%). This results in us trimming 1.85 percent of the data in the treatment group.

6 Interpretation: The Role of Peer Quality, Class Size, and Teacher Quality

We now turn to the question of *why* there are returns to quality for Tier I high schools, but not for others. Perhaps the simplest explanation is that there are smaller or nonexistent differences in peer quality across the non-Tier I cutoffs that are obscured when the results are aggregated, as they were in Figure 2. To examine whether or not that is the case, we ask whether there is a significant difference in peer quality across cutoffs *other than* the Tier I cutoff. Results are shown in Figure 6, which stacks together all cutoffs other than the Tier I cutoff.²⁷ Results in Figure 6a indicate that while barely admitted students attend schools with students who scored one-quarter of a standard deviation higher on the high school entrance exam, Figure 6b indicates that they score only 0.019 standard deviations higher on the college entrance exam, which is both economically small and statistically indistinguishable from zero. Thus, it is clear that while peer quality does increase discontinuously across the non-Tier I admission cutoffs, it is equally clear that there is no evidence of improvement on the CET.

Results for each cutoff separately are shown in Appendix Figure A4. While splitting the sample in this way leads to reduced statistical power, results are consistent across all three different sets of admission thresholds in showing significant increases in peer quality but no evidence of return. Panel (a) shows the discontinuity in peer quality at the Tier II cutoff of about one-fifth of a standard deviation, while panel (b) shows that there is no evidence of performance gains to barely attending the better school. Similarly, panel (c) shows the discontinuity in peer quality of 0.3 standard deviations across the cutoffs within Tier I (i.e., attending a more selective Tier I school versus a less selective Tier I school), while panel (d) shows that there is no performance gain across that cutoff. Finally, panel (e) documents a 0.09 standard deviation increase in peer quality across the cutoffs within Tier II (i.e., more selective versus less selective schools within Tier II), while panel (f) reveals no positive cognitive return for this group of students. Thus, Figure A4 shows that while there are significant improvements in peer quality across all non-Tier I admission thresholds in the school quality distribution, there are no cognitive returns to attending the more selective schools.

Combined with our main results reported earlier, these findings indicate that while barely attending Tier I schools (and the 0.35 standard deviation increase in peer quality) results in an improvement of 7 to 9 percent of a standard deviation in CET scores, barely attending

²⁷Because we exclude the Tier I cutoff, we cannot use a bandwidth greater than 18 points. This is because for one of our districts, the Tier I cutoff is 634 points, while the next cutoff after that is 615 points.

better schools across all other cutoffs except the cutoff from Tier II to Tier I (and the 0.2 standard deviation average increase in peer quality) does not result in better CET scores. Thus, these results suggest that peer quality is unlikely to explain the differences in cognitive returns across the Tier I and non-Tier I high schools that we observe.

A more nuanced explanation, however, is that perhaps nonlinear returns to peer quality could explain the observed pattern of findings. For example, one might argue that peer quality benefits high-ability students more than low-ability students. We also view that as inconsistent with the evidence. Specifically, we note that even within Tier I—where all students are relatively high-ability—we find a significant increase in peer quality without observing an increase in cognitive performance. Specifically, Figure A4c shows that those students who barely attend better Tier I schools versus worse Tier I schools experience peer quality that is 0.3 standard deviations higher, while Figure A4d shows that these students do not perform better on the CET exam.

A related possibility is that perhaps students sort differently into peer groups across the different cutoffs. For example, if the barely admitted students at the non-Tier I cutoffs were to primarily associate with lower-performing students at those better schools, but the barely admitted students at the Tier I cutoff were to associate with higher-performing students (i.e., they mix better, or more randomly), then that could explain the heterogeneity in returns. While our data do not allow us to directly test this, given the consistency of findings across cutoffs shown in Figure A4, we view this explanation as implausible. As a result, we interpret the pattern of findings as inconsistent with the hypothesis that the returns to school quality in this context are driven by peer quality.

A second potential interpretation is that the difference in returns is due to differential behavioral responses by students, such as those documented by Pop-Eleches and Urquiola (2013) in Romania. For example, in response to attending better schools students could feel marginalized by being the worst students relative to others. Alternatively, parents of barely admitted students could respond to changing the amount of private tutoring they purchase for their child.²⁸ While we are unable to test directly for these behavioral responses, we view them as unlikely to explain the heterogeneity in findings. That is because it is difficult to imagine how parents (or students) react differently across the Tier I cutoff than they do across cutoffs within Tier I, within Tier II, or across the Tier II/Tier III cutoff. Thus,

²⁸An estimated 30 percent of students in China receive some form of outside tutoring, which is significantly less than the 70 percent in Korea (For additional details, see <https://xa.yimg.com/kq/groups/17389986/899104917/name/CSFB+China+Education+Sector.pdf>). In addition, the length of the school day and week leaves relatively little time for outside tutoring. This is true especially for seniors, who attend school on Saturday, and who often attend school on weekdays to 9 pm, regardless of the selectivity of the school.

while we cannot rule out this type of differential-response explanation directly, we view it as unlikely to explain the results.

Next, we turn to whether the heterogeneity in returns could be driven by differences in class size, which has received considerable attention in the literature (e.g., Angrist and Lavy, 1999; Krueger, 2003). For example, if class size is discontinuously smaller across the Tier I cutoff, but not across other cutoffs, it could explain the heterogeneity in returns that we observe. Results for the Tier I cutoff are shown in Figure 7a, and indicate that students who barely attend Tier I schools are in significantly *larger* classes (56 versus 54 students). In contrast, Figure 8a shows that across all other cutoffs, class size is no different. If anything, that suggests that return to attending Tier I schools should be lower than the return across other cutoffs. As a result, the observed heterogeneity of returns is not easily explained by differences in class size.

Another potential explanation is teacher quality, which has been shown to lead to significant increases in achievement in other settings. While we do not have the necessary data to estimate teacher value-added in our setting, we do observe the proportion of teachers with the “superior teacher” ranking, which is the top ranking a teacher can receive outside of an exceptionally rare “special grade teacher” rank, and the only ranking (other than special grade teacher) that one cannot automatically qualify for with tenure and advanced degrees.²⁹ In addition, Lai, Sadoulet, and de Janvry (2011) use lottery data from Beijing middle schools to show that teacher ranks are highly correlated with the estimated school fixed effects. Similarly, Hannum and Park (2001) report that teachers with the highest rank in their sample increase achievement by 0.17 standard deviations relative to teachers with the next highest rank, and conclude that the quality ranks used in the Chinese schooling system contain a substantial amount of information on teacher quality that is not contained in conventional measures such as education of the teacher and years of experience.

Results for the Tier I admission threshold are shown in Figure 7b, and indicate that there is a large discontinuity of 10.8 percentage points in the proportion of superior teachers. In contrast, Figure 8b shows a much smaller discontinuity (estimated at 2.1 percentage points) in the proportion of superior teachers at the non-Tier I admission thresholds. Appendix Figures A5d, A5e, and A5f further break down the non-Tier I thresholds and show no evidence of a discontinuity in the proportion of superior teachers at the Tier II cutoff, and only very small discontinuities in teacher quality at cutoffs within Tiers I and II. This pattern is thus broadly consistent with our results on cognitive returns shown above; the small improvements in teacher quality at non-Tier I cutoffs are associated with small improvements in CET scores, while the large improvement in teacher quality at the Tier I cutoff is associated

²⁹In our sample, only 5 of the teachers (0.13%) have achieved special grade teacher rank, or “Teji Jiaoshi”.

with a large increase in CET scores.³⁰ This suggests that returns to high school quality in this context are due to teacher quality, rather than to peer quality. This conclusion is also broadly consistent with work by Jackson (2013), who reports that peer achievement can only explain a small fraction of the school selectivity effect in Trinidad & Tobago.³¹ In addition, suggestive evidence on the difference in the proportion of superior teachers by subject area around the Tier I cutoff helps explain why boys seem to benefit more from attending Tier I schools than girls.³²

To assess whether the differential exposure to superior teachers can explain all of the improved achievement we observe, we perform a back-of-the-envelope calculation. We estimate that the proportion of superior teachers increases by 10.8 percentage points at the Tier I cutoff. Given our findings on the return to attending Tier I schools shown in Panel C of Table 3, we estimate that if superior teachers increased achievement by 0.25 standard deviations relative to their counterparts, then the additional superior teachers would explain all of the estimated return to attending Tier I schools.³³ This effect is roughly twice as large as the increase in achievement found to result from a one standard deviation increase in teacher quality (Chetty, Friedman, and Rockoff, 2014; Aaronson et al. 2007; Kane, Rockoff, and Staiger 2008; Rockoff 2004; Rivkin, Hanushek, and Kain 2005). In addition, we emphasize that there are likely other improvements in teacher quality at the Tier I admission threshold

³⁰When rescaling to account for the differences in the discontinuities in teacher quality across the Tier I and non-Tier I cutoffs, the implied effects of teacher quality are quite similar. The estimated discontinuity in teacher quality is 5.1 times larger for the Tier I cutoff than the non-Tier I cutoff (10.8 percentage points in Figure 7b versus 2.1 percentage points in Figure 8b). Multiplying 5.1 by the estimated discontinuity in CET scores for the non-Tier I cutoff of 0.019 shown in Figure 5b results in an estimate of 0.097, which is similar to the estimated discontinuity of 0.086 in CET scores at the Tier I cutoff, as shown in Figure 3c.

³¹It is also consistent with how little discretion principals at the schools studied by Abdulkadiroglu, Angrist, and Pathak, (2014) apparently have in deciding which teachers to hire, according to our discussion with one of the authors. As a result, if all of the returns to school quality are driven by teacher quality, we would not expect positive effects in that context.

³²While we were unable to obtain data on teacher subject area by rank for all schools, we were able to obtain this information for two Tier I schools and two Tier II schools. The difference in the proportion of superior teachers between these Tier I and Tier II schools is 9.5 percentage points, which is similar to the discontinuity at the Tier I cutoff shown in Figure 7 of 10.8 percentage points. However, the difference between the Tier I and Tier II schools in the proportion of superior math and science teachers of 20.5 percentage points is much larger than the difference in the proportion of superior arts teachers (8.1 percentage points). As a result, if the benefits from better schools were solely due to better teachers, we would expect there to be larger effects for boys because they major in science at much higher rates than girls (71 versus 33 percent.) Similarly, when we estimate the benefits to attending Tier I schools by major, we find that the overall benefits are driven by students who major in science. In contrast, we find no evidence that teacher gender is different from Tier I to Tier II, for either all teachers or for superior teachers.

³³Estimates in Panel C of Table 3 indicate that Tier I schools increase average student achievement by approximately 0.08 standard deviations. Assuming that all of that return comes from increased access to superior teachers implies rescaling those estimates by 0.108, which indicates that superior teachers increase scores by 0.74 standard deviations. Dividing by three given the three years of high school results in estimates of 0.25.

that are more difficult to measure. Thus, our view is that it is likely that the academic benefits from attending Tier I schools are due to increases in teacher quality.

In summary, the additional exercises in this section indicate that the returns to high school quality in our setting are unlikely to be caused by peer quality or class size. Rather, the empirical evidence suggests that the returns are likely due to differences in teacher quality - as proxied by superior teacher rank. Thus, while we cannot rule out with complete certainty that the heterogeneous returns across the cutoffs are due to differences in some unobserved input other than teacher quality, we think the interpretation most consistent with our findings is that the benefits from attending better schools are due to teacher quality.

7 Conclusion

This paper estimates the cognitive benefits due to attending higher quality high schools. It does so by using a regression discontinuity design that compares the academic outcomes of students who are barely eligible and ineligible to enter better quality high schools in China. Results indicate that across the distribution of school quality, the only positive returns to school quality are for those who attend Tier I, rather than Tier II schools. Importantly, we document that this is true despite the fact that admission threshold-crossing across the continuum of high school quality is associated with significant increases in peer quality. As a result, we conclude that at least in this setting, positive cognitive returns to high school quality are unlikely to be due to peer quality.

We provide additional evidence that the returns to attending elite Tier I schools are due to teacher quality. Specifically, we document that students who (barely) attend Tier I schools are significantly more likely to be taught by teachers of superior rank. A back-of-the-envelope calculation suggests this increased exposure to teachers of superior rank can explain the entire cognitive return to Tier I schools if those teachers increase achievement by around 0.25 standard deviations compared to their counterparts. This implies that for the increased access to superior teachers to explain all of the cognitive benefits, they must be just over two standard deviations higher quality than their counterparts. While this is large, it is not implausible given the existing research on teacher quality in China. Hannum and Park (2001) estimate that teachers of superior rank increase test scores by around 0.17 standard deviations, while Lai, Sadoulet, and de Janvry (2011) show that teacher rank is highly correlated with estimates of school quality. In addition, we note that other less-easily-measured forms of teacher quality may also be improving across the Tier I cutoff, which could be responsible for some of the improvement in achievement.

These findings have important implications for the literature on the returns to school

quality. First, by demonstrating the heterogeneity of returns in a single educational context that cannot be explained by differences in peer quality, the results here highlight the importance of measuring additional education inputs. Thus, while peer quality has been and remains a straightforward proxy for school quality, the results of this paper highlight that it is not a sufficient statistic for school quality, and that focusing on peer quality can make it difficult to reconcile seemingly inconsistent findings.

In addition, to the extent that returns to high school quality more generally are not due to peer quality, it has important implications for how to increase academic achievement. If the benefits to attending better schools were due to better peers, it would be difficult to extend those benefits more broadly since there is a limited set of high-performing peers. In contrast, the results presented here demonstrate that at least in this context, policymakers may be able to do other things—such as improve teacher quality—to replicate school quality and improve educational outcomes.

References

- Aaronson, D., Barrow, L., Sander, W., 2007. Teachers and Student Achievement in the Chicago Public High Schools. *Journal of Labor Economics* 25(1) 95-135.
- Andrews, R.J., Li, J., Lovenheim, M.F., 2012. Quantile Treatment Effects of College Quality on Earnings: Evidence from Administrative Data in Texas. *NBER Working Paper 18068*.
- Abdulkadiroglu, A., Angrist, J., Pathak, P., 2014. The Elite Illusion: Achievement effects at Boston and New York Exam Schools. *Econometrica* 82(1): 137-196.
- Angrist, J., Lavy, V., 1999. Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. *The Quarterly Journal of Economics* 114(2): 533-575.
- Berkowitz, D., Hoekstra M., 2011. Does high school quality matter? Evidence from admissions data. *Economics of Education Review* 30 (2011) 280-288.
- Calonico, S., Cattaneo, M.D., Titiunik, R., 2014. Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica* 82 (6): 2295-2326.
- Canaan, S., Mouganie, P., 2015. Returns to Education Quality for Low Skilled Students: Evidence from a Discontinuity. *Working Paper*. Available at SSRN 2518067.
- Chetty R., Friedman, J.N., Rockoff, J.E., 2014. Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *American Economic Review* 104(9): 2593-2632.
- Clark, D., 2010. Selective Schools and Academic Achievement. *B.E Journal of Economics Analysis and Policy: Advances in Economic Analysis and Policy* 10(1).
- Clark, D., Del Bono, E., 2014. The Long-Run Effects of Attending an Elite School: Evidence from the UK. *IZA Discussion Paper No.8617* November 2014.
- Ding, W., Lehrer, S, F., 2007. Do Peers Affect Student Achievement in China's Secondary Schools? *Review of Economics and Statistics* 89(2): 300-312.
- Dobbie W., Fryer, R.G., 2014. The Impact of Attending a School with High-Achieving Peers: Evidence from the New York City Exam Schools. *American Economics Journal: Applied Economics* 6(3): 58-75.
- Giles, J., Park, A., Wang, W., 2015. The Great Proletarian Cultural Revolution, Disruptions to Education, and the Returns to Schooling in Urban China. *IZA Discussion Paper No. 8930*, March 2015.
- Hannum, E., Park, A., 2001. Do Teachers Affect Learning in Developing Countries?: Evidence from Matched Student-Teacher Data from China. *Conference Rethinking Social Science Research on the Developing World in the 21st Century*.
- Hoekstra, M., 2009. The Effect of Attending the Flagship State University on Earnings: A Discontinuity-Based Approach. *The Review of Economics and Statistics* 91 (4), 717-24.

- Imbens, G.W., Lemieux, T., 2008. Regression discontinuity designs: A guide to practice. *Journal of Econometrics* 142 (2), 615-35.
- Jackson, C.K, 2010. Do Students Benefit From Attending Better Schools? Evidence From Rule-based Student Assignments in Trinidad and Tobago. *Economic Journal* 120(549): 1399-1429.
- Jackson, C.K, 2013. Can higher-achieving peers explain the benefits to attending selective schools? Evidence from Trinidad and Tobago. *Journal of Public Economics* 108: 63-77.
- Kane, T.J., Rockoff, J.E., Staiger, O., 2008. What Does Certification Tell Us about Teacher Effectiveness? Evidence from New York City. *Economics of Education Review* 27(6):615-31.
- Krueger, A.B., 2003. Economic Considerations and Class Size. *Economic Journal* 113(485): F34-F63.
- Lai, F., Sadoulet, E., de Janvry, A., 2011. The Contributions of School Quality and Teacher Qualifications to Student Performance: Evidence from a Natural Experiment in Beijing Middle Schools. *Journal of Human Resources* 46: 123-153.
- Lee, D.S, 2009. Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *Review of Economic Studies* 76(3): 1071-1102.
- Lee, D.S., Card, D. 2008. Regression discontinuity inference with specification error. *Journal of Econometrics* 142 (2), 655-674.
- Lee, D.S., Lemieux, T., 2010. Regression discontinuity designs in economics. *Journal of Economic Literature* 48 (2), 281-355.
- Li, H., Liu, P. W., Ma, N., Zhang, J., 2005. Does Education Pay in Urban China? Estimating Returns to Education Using Twins. *Discussion Papers 00015, Chinese University of Hong Kong, Department of Economics*.
- Lucas, A.M., Isaac M.M., 2014. Effects of School Quality on Student Achievement: Discontinuity Evidence from Kenya. *American Economic Journal: Applied Economics* 6(3): 234-63.
- Park, A., Xinzhen, S., Xuenhi, An., Hsieh C., 2015. Magnet High Schools and Academic Performance in China: A Regression Discontinuity Design. Forthcoming in *Journal of Comparative Economics*.
- Pop-Eleches, C., Urquiola M., 2013. Going to a Better School: Effects and Behavioral Responses. *American Economic Review* 103(4): 1289-1324.
- Rivkin, S.G., Hanushek, E.A., Kain, J.F., 2005. Teachers, Schools and Academic Achievement. *Econometrica* 73(2): 417-58
- Rockoff, J.E., 2004. The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data. *American Economic Review: Papers and Proceedings* 94(2): 247-52

Saavedra, J., 2009. The Learning and Early Labor Market Effects of College Quality: A Regression Discontinuity Analysis. *Rand Corporation Working Paper*.

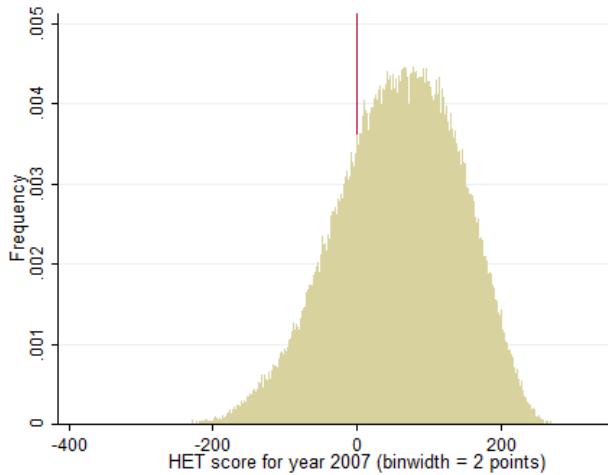
Zimmerman, S., 2014. The Returns to College Admission for Academically Marginal Students. *Journal of Labor Economics* 32 (4): 711-754.

Zhang, H., Forthcoming. Identification of Local Average Treatment Effects under Imperfect Matching: with an Application to Chinese Elite Schools. *Journal of Public Economics*.

A Figures

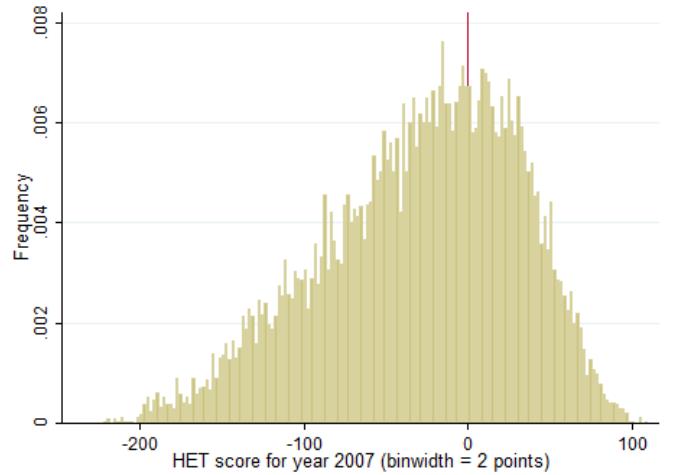
Figure 1: Testing the validity of the RD design for both empirical strategies

Attending Better Schools (All Cutoffs)



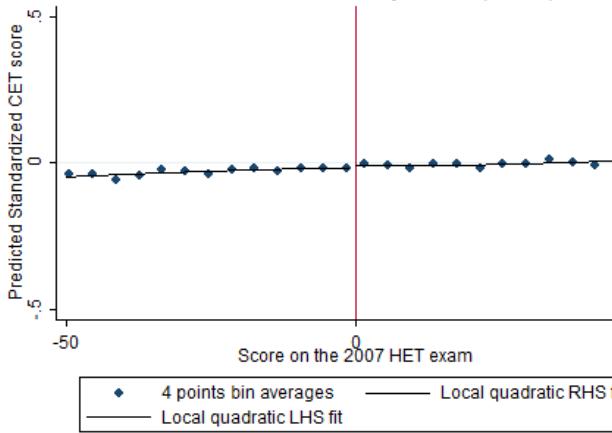
(a) Distribution of HET scores

Attending a Tier I school



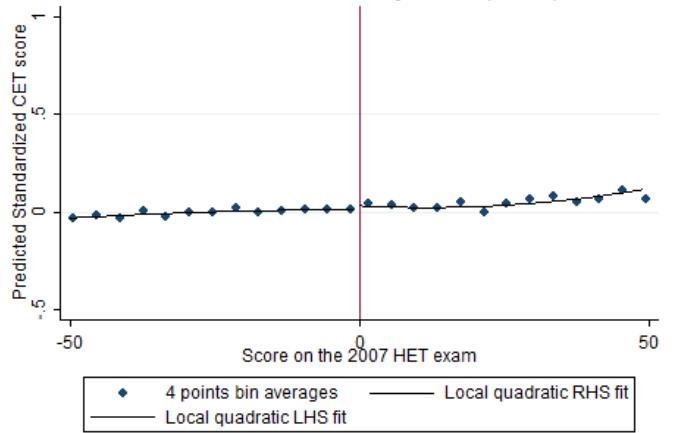
(b) Distribution of HET scores

Estimated Discontinuity: 0.006(0.007)



(c) Smoothness of baseline covariates

Estimated Discontinuity: 0.019(0.016)



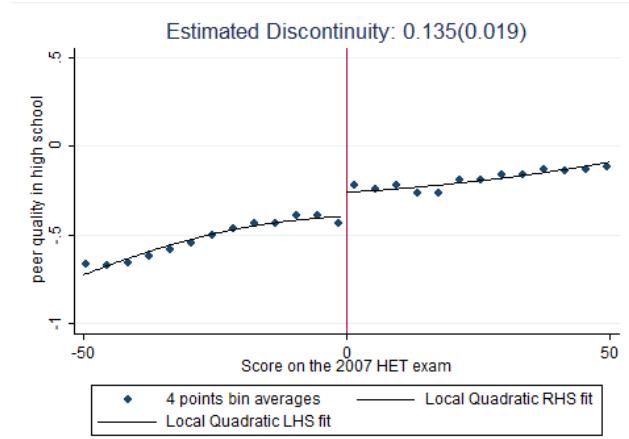
(d) Smoothness of baseline covariates

Notes: Sample includes students who took the HET exam in the year 2007.

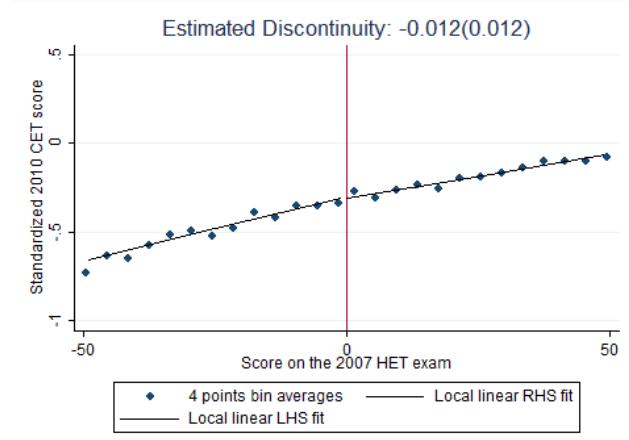
Bins for histogram represent an average count of 2 score points.

Predicted score based on the following controls: sex, gender, district fixed effects, middle school fixed effect.

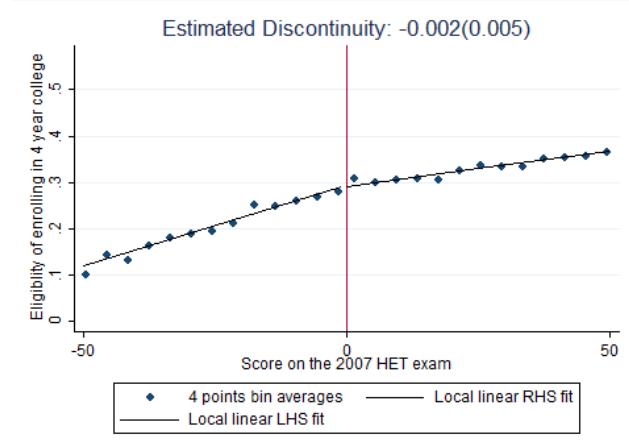
Figure 2: Local polynomial “stacked RD” estimates for attending better schools



(a) Peer quality based on scores on high school entrance exam



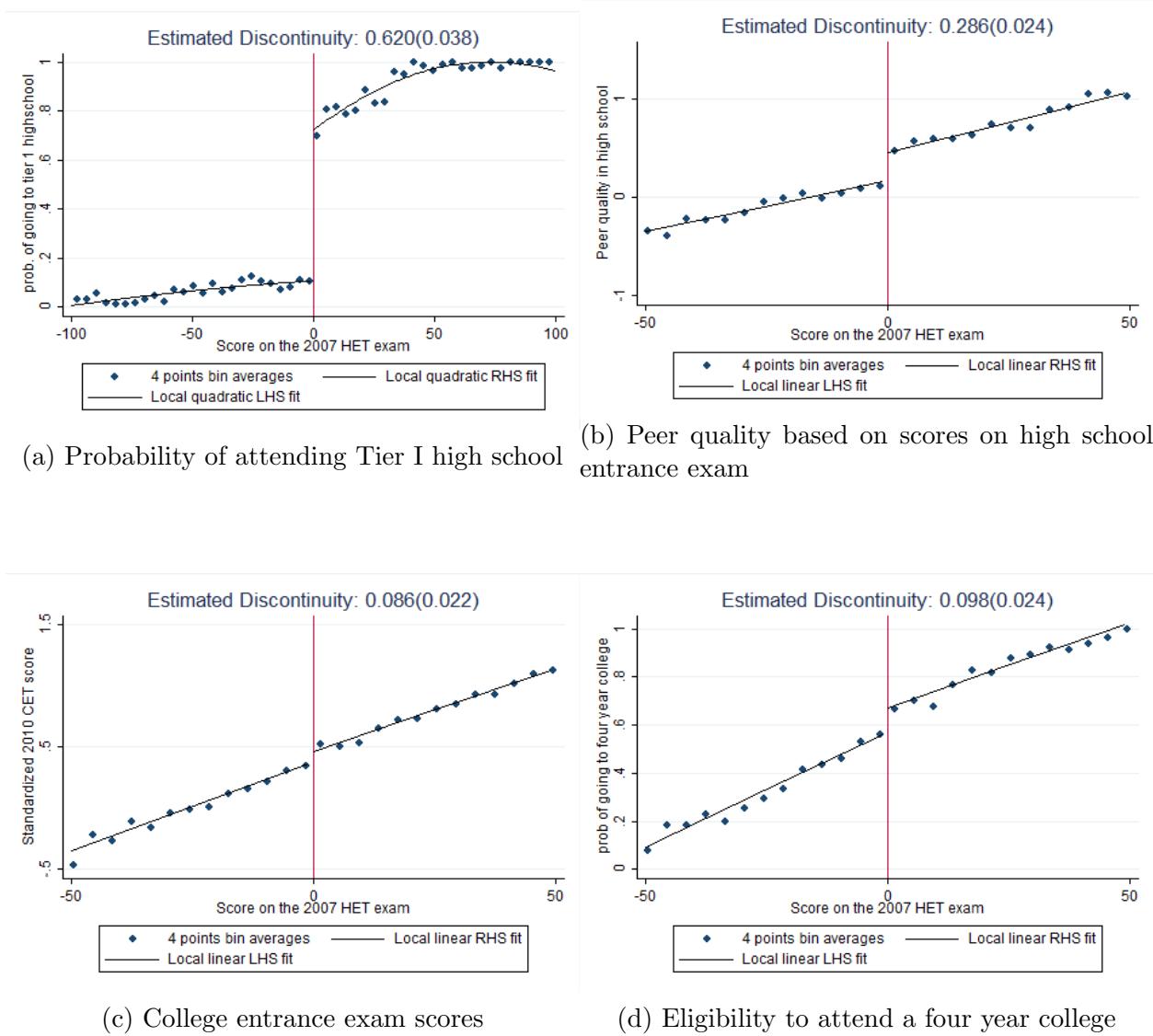
(b) College entrance exam test scores



(c) Eligibility to attend a four year college

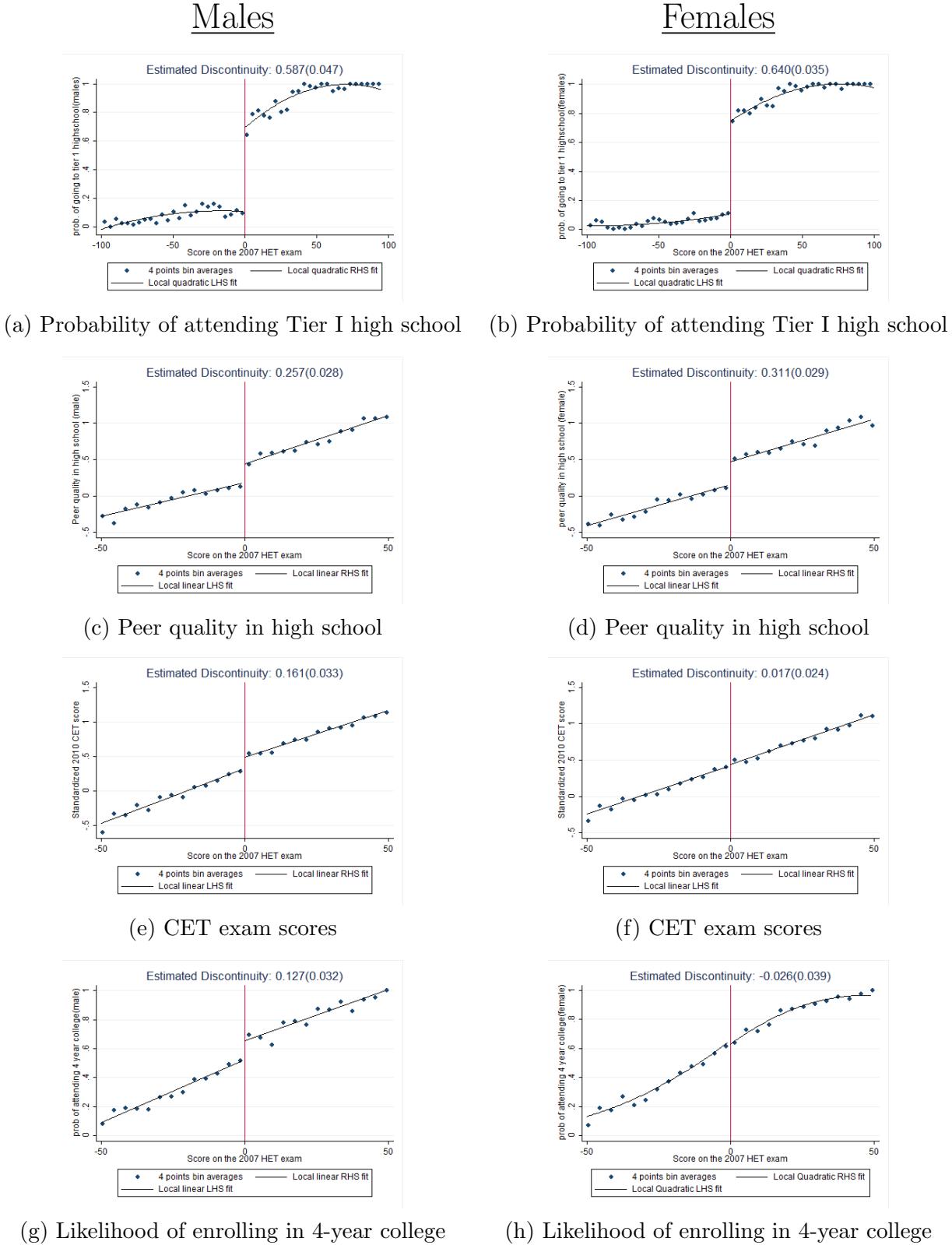
Notes: Sample includes students who took the high school entrance exam (HET) in the year 2007. Since we observe individuals with multiple cutoffs, we cluster at the student ID level.

Figure 3: Local polynomial RD estimates for attending Tier I schools



Notes: Sample includes students who took the high school entrance exam in the year 2007. Standard errors clustered at score level.

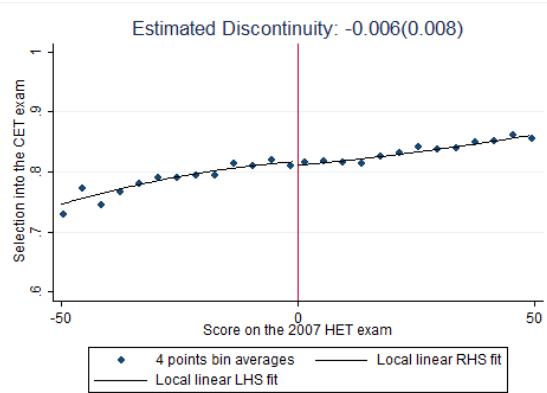
Figure 4: Local polynomial RD estimates for attending Tier I schools, by gender



Notes: Sample includes students who took the high school entrance exam in the year 2007. Standard errors clustered at score level.

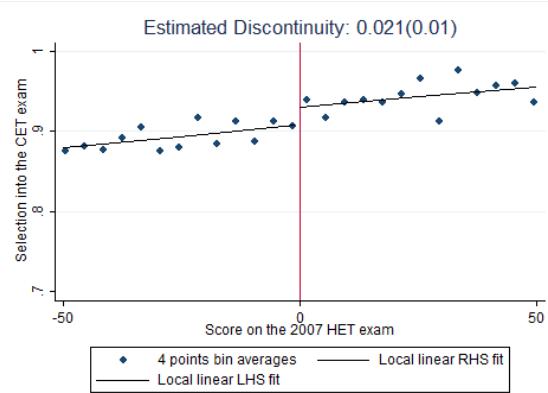
Figure 5: Selection into the college entrance exam

Attending Better Schools (All Cutoffs)

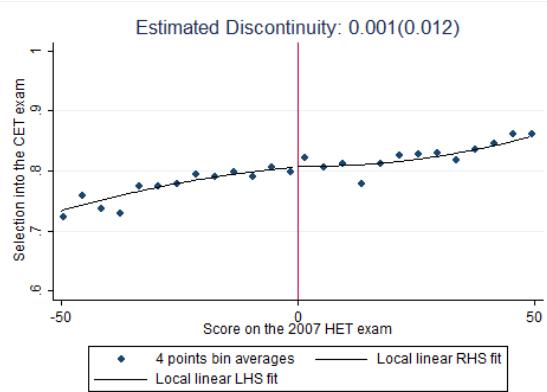


(a) Selection into the CET exam

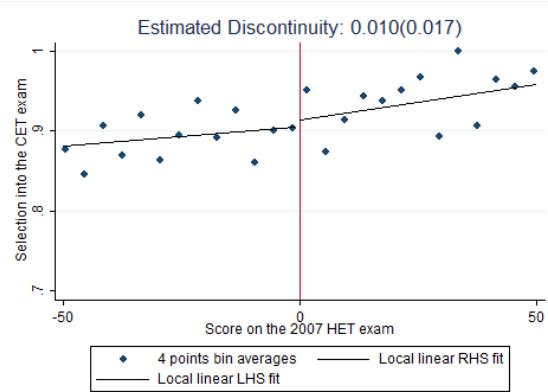
Attending a Tier I School



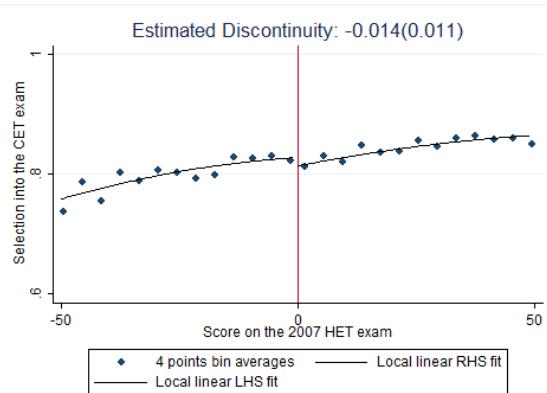
(b) Selection into the CET exam



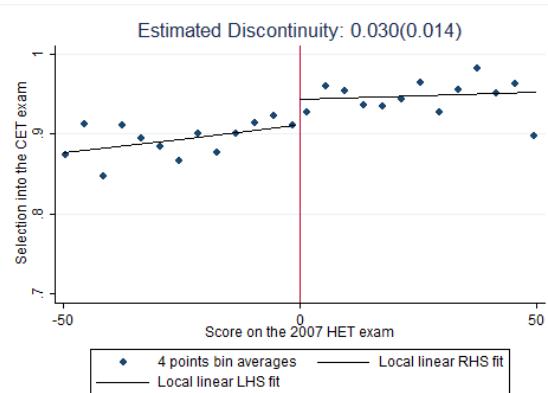
(c) Selection into the CET exam (Males only)



(d) Selection into the CET exam (Males only)



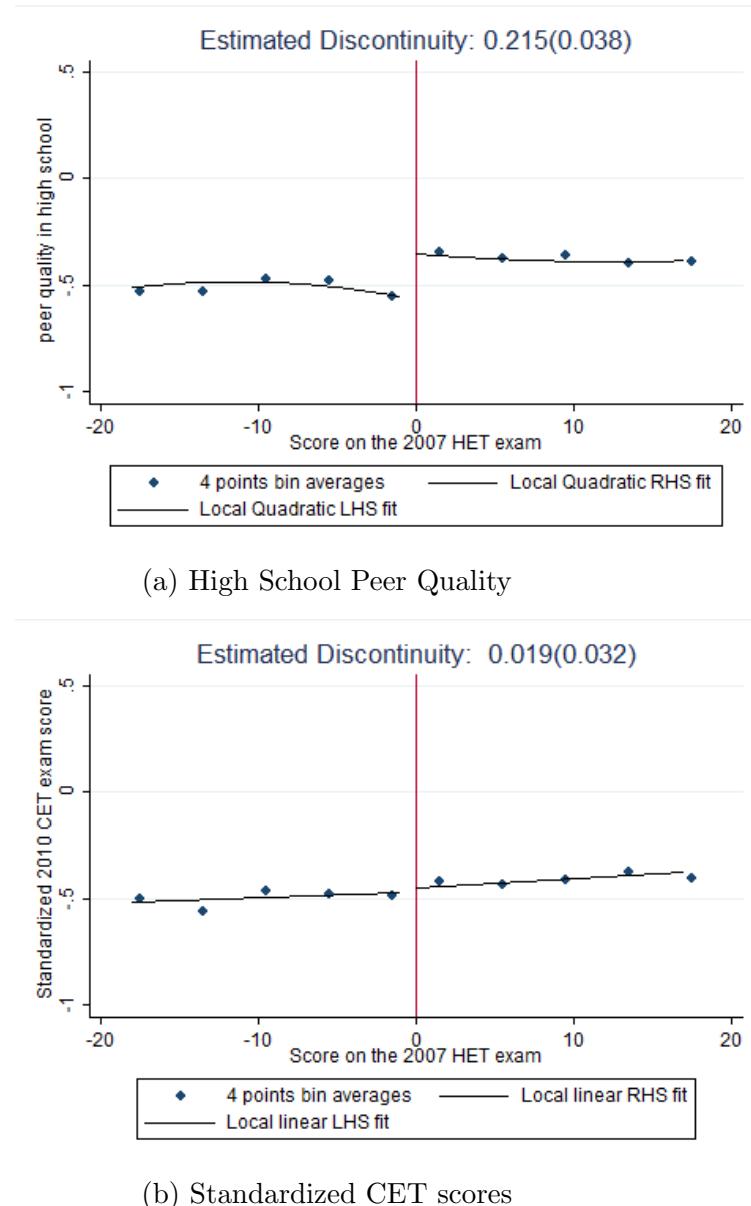
(e) Selection into the CET exam (Females only)



(f) Selection into the CET exam (Females only)

Notes: Sample includes students who took the high school entrance exam in the year 2007 (including those with no college entrance exam scores).

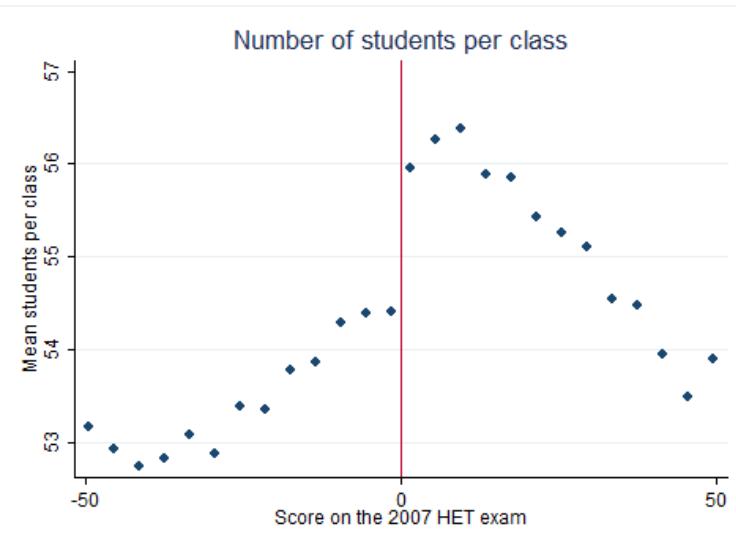
Figure 6: Local polynomial “stacked RD” estimates with all cutoffs except the Tier I cutoff



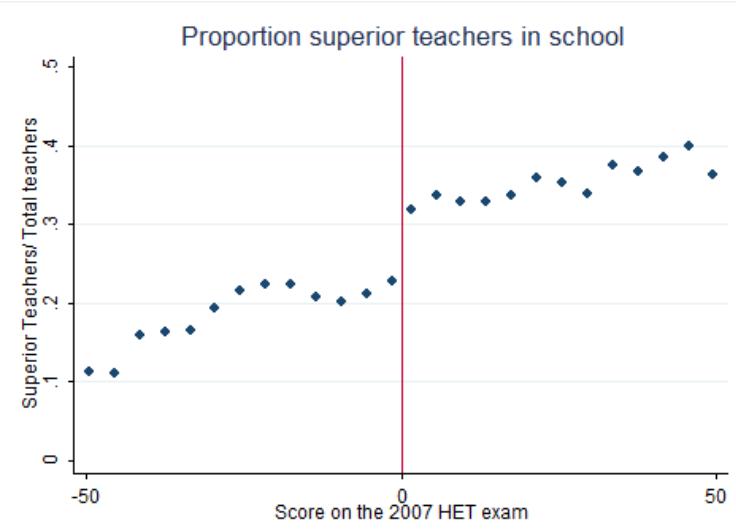
Notes: Sample includes students who took the high school entrance exam (HET) in the year 2007. Since we observe individuals with multiple cutoffs, we cluster at the student ID level.

In order to exclude the Tier I admission threshold, we use a maximum bandwidth of 18 points.

Figure 7: Class size and teacher quality RD estimates for attending a Tier I school



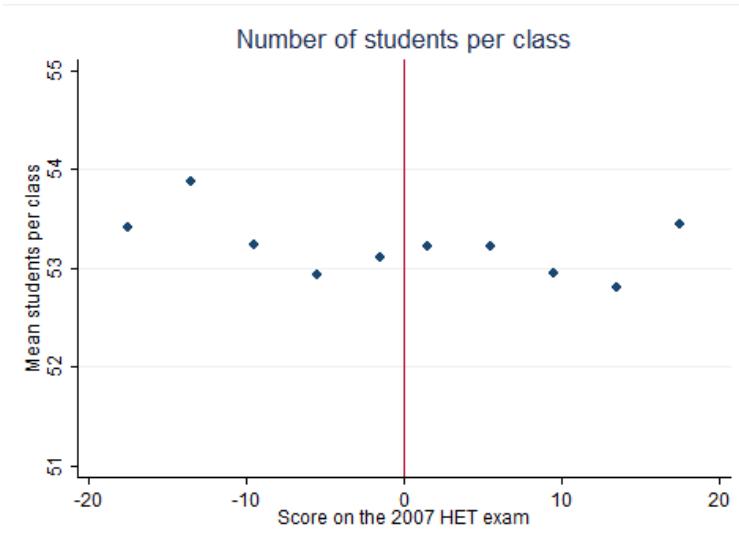
(a) Number of students per class



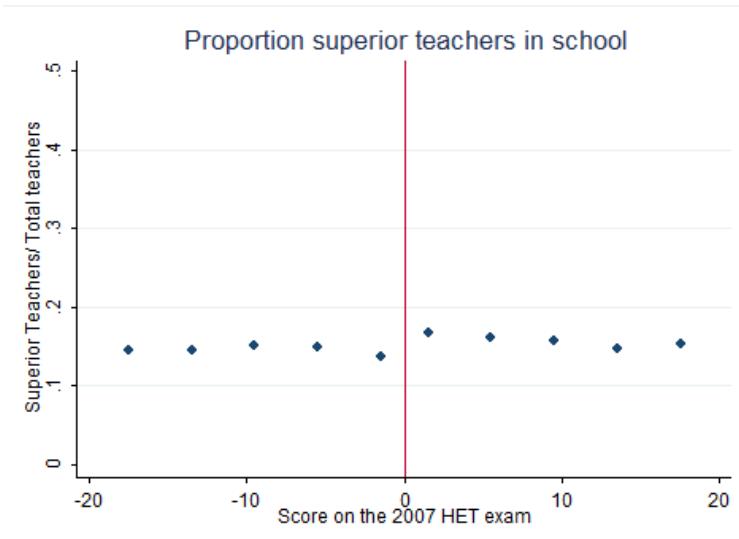
(b) Proportion superior teachers

Notes: Sample based off of school level data.

Figure 8: Class size and teacher quality “stacked RD” estimates for all admission cutoffs except for the Tier I cutoff



(a) Number of students per class



(b) Proportion superior teachers

Notes: Sample based off of school level data.

In order to exclude the Tier I admission threshold, we use a maximum bandwidth of 18 points.

B Tables

Table 1: Descriptive Statistics

	(1) Whole Sample	(2) Tier I Schools	(3) Tier II Schools	(4) Tier III/IV Schools
High school entrance exam scores	614.74 (59.52)	669.02 (31.91)	602.07 (39.62)	537.47 (43.17)
College entrance exam scores	487.57 (99.69)	567.69 (60.57)	464.40 (83.37)	386.14 (79.81)
Proportion female	0.53	0.54	0.52	0.53
Proportion majoring in arts in high school	0.48	0.35	0.52	0.66
Proportion private schools	0.010	0.006	0.015	0.018
Eligible for four year college	0.42	0.81	0.26	0.05
Eligible for elite college	0.08	0.22	0.012	0.002
Proportion female teachers	0.55	0.55	0.56	0.56
School Size (in Mu*)	128.49 (42.26)	145.97 (47.71)	123.24 (34.70)	99.14 (27.90)
Number of students per class	53.62 (5.14)	55.02 (2.55)	53.24 (2.26)	51.54 (11.62)
Ratio of superior teachers	0.22 (0.16)	0.38 (0.08)	0.16 (0.13)	0.07 (0.05)
Number of schools	25	4	12	9
Number of Students	12250	4306	5900	2044

Notes: *1 Chinese Mu = 7176 sq feet.

Data taken from two rural districts in the Province for students taking the high school entrance exam in 2007.
Standard errors (for non-binary variables) in parentheses.

Table 2: “Stacked RD” estimates for attending better schools across all admission cutoffs

Bandwidth	2.5 CCT (1)	2 CCT (2)	1.5 CCT (3)	1.25 CCT (4)	CCT (5)	0.75 CCT (6)
Panel A:						
Discontinuity in high school peer quality	.189*** (.025)	.204*** (.028)	.149*** (.021)	.188*** (.024)	.187*** (.027)	.205*** (.031)
With Controls	.187*** (.023)	.203*** (.026)	.148*** (.020)	.188*** (.023)	.187*** (.026)	.204*** (.030)
Observations	33,414	26,806	20,221	17,369	13,570	9,619
Panel B:						
Discontinuity in CET exam scores	−.002 (.016)	−.005 (.020)	−.017 (.013)	−.009 (.016)	−.007 (.020)	.006 (.025)
With Controls	−.006 (.016)	−.009 (.019)	−.019 (.013)	−.014 (.016)	−.014 (.019)	−.004 (.024)
Observations	66,530	53,930	41,599	34,334	27,777	21,135
Panel C:						
Discontinuity in likelihood of enrolling in 4-year college	−.001 (.008)	−.004 (.005)	−.003 (.006)	−.001 (.007)	−.001 (.009)	.009 (.012)
With Controls	−.002 (.007)	−.004 (.005)	−.004 (.006)	−.001 (.007)	−.003 (.009)	.005 (.011)
Observations	70,493	57,330	42,524	37,029	29,682	22,065
Score Polynomial	Two	Two	One	One	One	One

Notes: Sample includes students who took the college entrance exam in the year 2007.

Controls include: Age, gender, district fixed effects and middle school fixed effects. Optimal Bandwidth selected using the CCT bandwidth selector proposed in Calonico et al. (2015).

Optimal BW = 14 for high school peer quality regressions. Optimal BW = 29 for CET exam regressions. Optimal BW = 31 for likelihood of enrolling in four year degree regressions.

Since we observe individuals with multiple cutoffs, we cluster at the student ID level.

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

Table 3: RD estimates for attending Tier I schools

Bandwidth	2.5 CCT (1)	2 CCT (2)	1.5 CCT (3)	1.25 CCT (4)	CCT (5)	0.75 CCT (6)
Panel A: Discontinuity in probability of attending Tier I school						
Tier I school	.631*** (.031)	.632*** (.036)	.671*** (.046)	.649*** (.052)	.642*** (.064)	.628*** (.079)
With Controls	.637*** (.031)	.637*** (.037)	.677*** (.047)	.656*** (.053)	.654*** (.064)	.637*** (.078)
Observations	7,167	6,046	4,654	3,901	3,133	2,389
Panel B: Discontinuity in high school peer quality						
	.301*** (.025)	.314*** (.028)	.372*** (.032)	.357*** (.035)	.358*** (.043)	.345*** (.051)
With Controls	.315*** (.025)	.329*** (.029)	.395*** (.033)	.378*** (.037)	.380*** (.044)	.359*** (.051)
Observations	6,680	5,578	4,278	3,642	2,886	2,237
Panel C: Discontinuity in CET exam scores						
	.082*** (.017)	.089*** (.019)	.083*** (.022)	.085*** (.023)	.073*** (.025)	.069** (.028)
With Controls	.081*** (.016)	.086*** (.017)	.084*** (.021)	.090*** (.021)	.073*** (.022)	.065** (.027)
Observations	9,909	8,870	7,352	6,454	5,298	4,112
Panel D: Discontinuity in likelihood of enrolling in 4-year college						
	.135*** (.023)	.097*** (.024)	.063** (.028)	.049 (.030)	.051 (.035)	.056 (.043)
With Controls	.136*** (.023)	.101*** (.024)	.066** (.028)	.048 (.030)	.048 (.034)	.053 (.041)
Observations	8,478	7,236	5,824	4,859	3,941	3,030
Score Polynomial	One	One	One	One	One	One

Notes: Sample includes students who took the college entrance exam in the year 2007. Controls include: Age, gender, district fixed effects and middle school fixed effects. Optimal Bandwidth selected using the CCT bandwidth selector proposed in Calonico et al. (2015).

Optimal BW = 20 for likelihood of attending Tier I school regressions. Optimal BW = 18 for high school peer quality regressions. Optimal BW = 34 for CET exam regressions. Optimal BW = 25 for likelihood of enrolling in four year degree regressions.

Standard errors clustered at the score level. *** p <0.01 ** p <0.05 * p <0.1

Table 4: Local linear intent-to-treat and local average treatment effect estimates for attending Tier I schools

Treatment effect	ITT	LATE	ITT	LATE	ITT	LATE
Gender	All		Males		Females	
Panel A: First stage						
Likelihood of attending Tier I school	.632*** (.036)	—	.609*** (.045)	—	.652*** (.033)	—
Panel B: Discontinuity in school inputs						
High school peer quality	.303*** (.026)	.486*** (.023)	.290*** (.030)	.484*** (.029)	.317*** (.032)	.489*** (.035)
Panel C: Discontinuity in outcomes						
College entrance exam test scores	.094*** (.024)	.155*** (.040)	.174*** (.034)	.306*** (.061)	.015 (.027)	.020 (.042)
Eligibility to attend a 4-year college	.071*** (.027)	.119*** (.043)	.103*** (.036)	.191*** (.059)	−.042 (.042)	−.060 (.063)
Observations	6046	6046	2813	2813	3233	3233

Notes: Sample includes students who took the high school entrance exam in the year 2007.

All regressions include controls: gender, age, district fixed effects and junior high school fixed effects.

For ease of comparison, all local linear regressions use an equal bandwidth of 40 points on either side of the cutoff.

Standard errors are clustered at the score level.

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

Table 5: Regression discontinuity estimates for selection into the college entrance exam

Bandwidth	2.5 CCT (1)	2 CCT (2)	1.5 CCT (3)	1.25 CCT (4)	CCT (5)	0.75 CCT (6)
Panel A: (Going to a better school)						
Selecting into the CET entrance exam (All)	-.011 (.008)	-.007 (.008)	-.001 (.011)	.004 (.012)	-.000 (.013)	.010 (.014)
Females only	-.015 (.009)	-.017 (.011)	-.012 (.014)	-.008 (.016)	-.001 (.017)	-.003 (.020)
Males only	-.004 (.010)	.004 (.012)	.010 (.016)	.017 (.017)	.001 (.019)	.025 (.021)
Observations (females)	18226	14615	10944	9076	7461	5552
Observations (males)	16522	13177	9840	8246	6768	5082
Panel B: (Going to a top school)						
Selecting into the CET entrance exam (All)	.007 (.008)	.020** (.009)	.020** (.010)	.017 (.011)	.015 (.012)	.025* (.014)
Females only	.017 (.013)	.037*** (.014)	.021 (.015)	.019 (.017)	.011 (.019)	.025 (.022)
Males only	-.005 (.015)	.001 (.017)	.019 (.019)	.015 (.021)	.019 (.024)	.023 (.027)
Observations (females)	5386	4663	3791	3214	2607	1983
Observations (males)	4800	4166	3375	2862	2363	1822
Score Polynomial	One	One	One	One	One	One

Notes: Sample includes students who took the high school entrance exam in the year 2007 with known high school cutoffs (including those who did not sit for the 2010 college entrance exam).

Optimal Bandwidth selected using the CCT bandwidth selector proposed in Calonico et al. (2015).

Optimal BW = 12 for likelihood of taking CET exam (going to a better school school).

Optimal BW = 29 for likelihood of taking CET exam (going to a top school).

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

Table 6: Bounding analysis for the estimated impact of attending Tier I schools

Bandwidth	84 points (1)	68 points (2)	51 points (3)	43 points (4)	34 points (5)	26 points (6)
Panel A: Selection into the CET exam						
CET exam	.003 (.009)	.009 (.009)	.020** (.009)	.017* (.010)	.013 (.011)	.017 (.012)
Constant (Control mean)	.926	.916	.903	.904	.908	.902
Proportion to be trimmed	0.32%	0.98%	2.2%	1.88%	1.43%	1.88%
Panel B: College entrance exam test scores (Original regression estimates) [†]						
College entrance exam test scores (Lower bound estimates)	.082*** (.017)	.089*** (.019)	.083*** (.022)	.085*** (.023)	.073*** (.025)	.069** (.028)
College entrance exam test scores (Upper bound estimates)	.078*** (.020)	.077*** (.020)	.060** (.025)	.063** (.026)	.054** (.027)	.048* (.028)
Observations	9909	8870	7352	6454	5298	4112
Observations after trimming	9851	8826	7268	6391	5258	4068
Score Polynomial	One	One	One	One	One	One

Notes: Sample includes students who took the high school entrance exam in the year 2007 with known high school cutoffs (including those who did not sit for the 2010 college entrance exam).

Optimal Bandwidth selected using the CCT bandwidth selector proposed in Calonico et al. (2015).

To ease comparison with our previous estimates, we use the same bandwidths predicted by the CCT for the original college entrance score regressions.

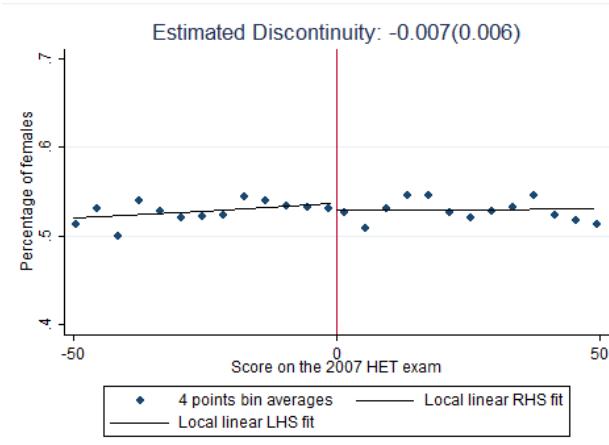
Bootstrapped standard errors reported for the upper and lower bound estimates

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

C Appendix Figures

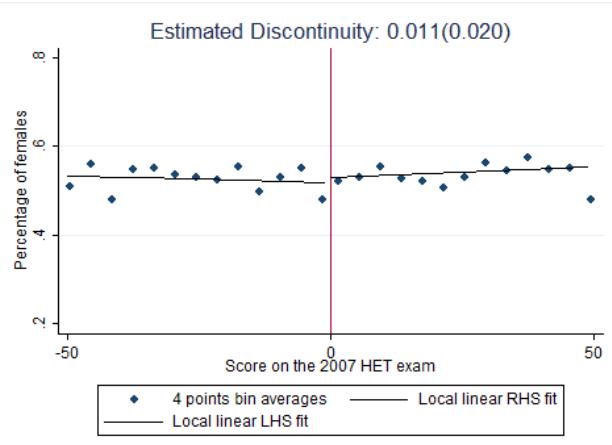
Figure A1: Smoothness of baseline covariates

Attending Better Schools (All Cutoffs)

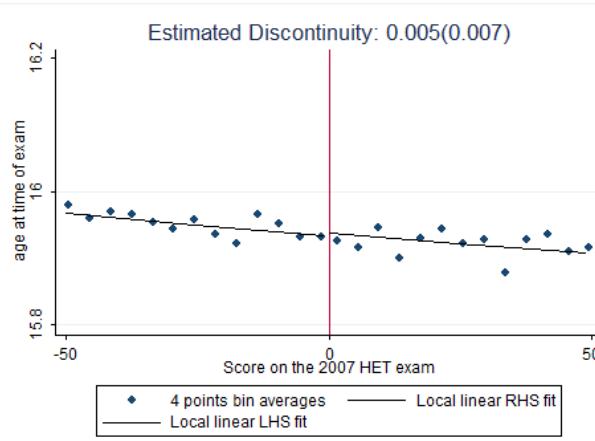


(a) Probability of being female

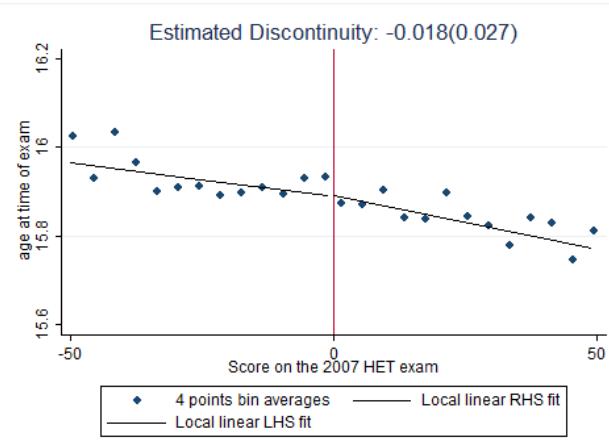
Attending a Tier I school



(b) Probability of being female



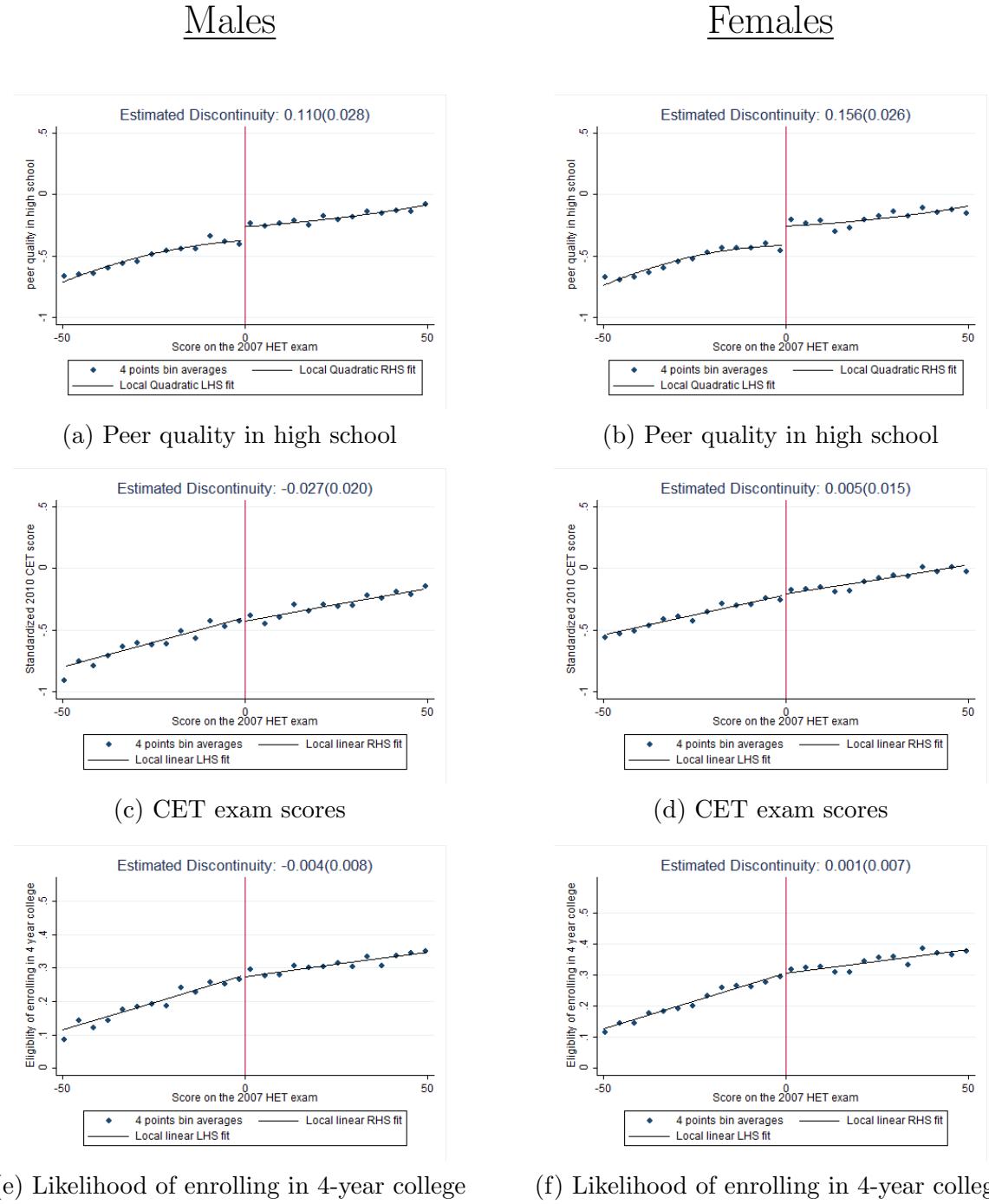
(c) Age at time of high school exam



(d) Age at time of high school exam

Notes: Sample includes students who took the high school entrance exam in the year 2007.

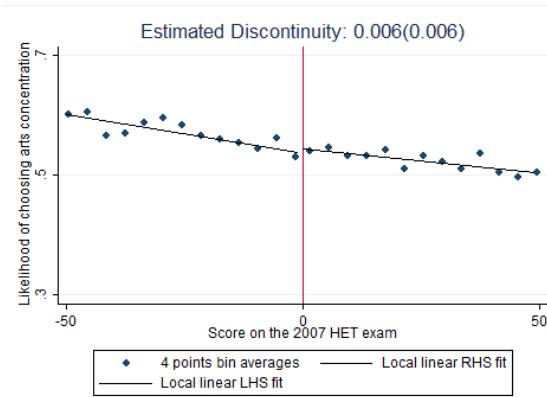
Figure A2: Local polynomial “stacked RD” estimates for attending better schools, by gender



Notes: Sample includes students who took the high school entrance exam in the year 2007. Due to repeated observations, standard errors clustered at individual level.

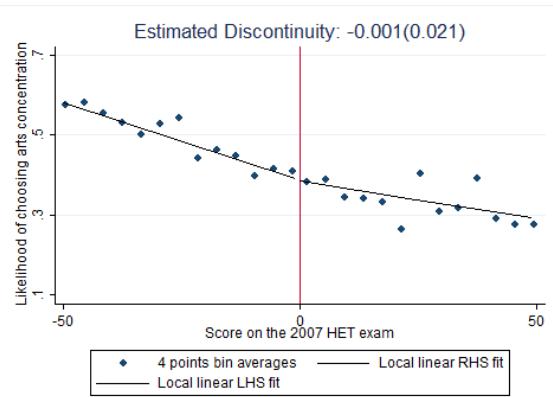
Figure A3: Threats to interpretation

Attending Better Schools

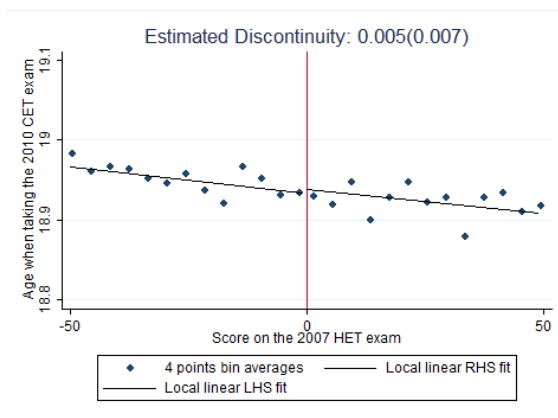


(a) Likelihood of majoring in arts versus sciences in High school.

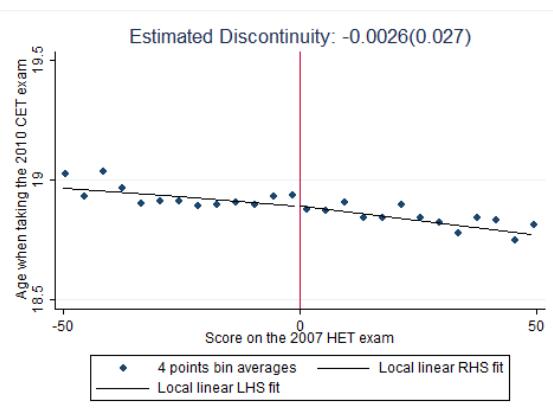
Attending a Tier I school



(b) Likelihood of majoring in arts versus sciences in High school.



(c) Exact age when taking the 2010 CET exam.

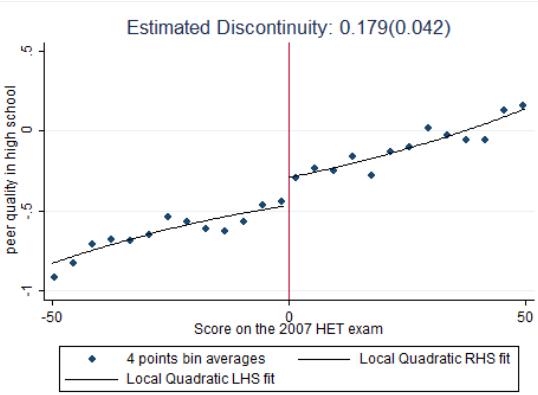


(d) Exact age when taking the 2010 CET exam.

Notes: Sample includes students who took the CET exam in the year 2010.

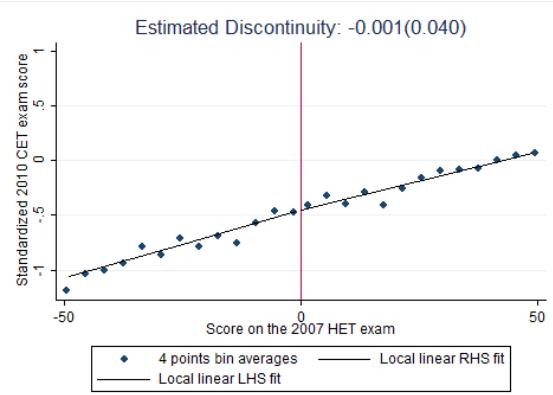
Figure A4: Local polynomial peer quality and test score estimates for other cutoffs

High School Peer Quality

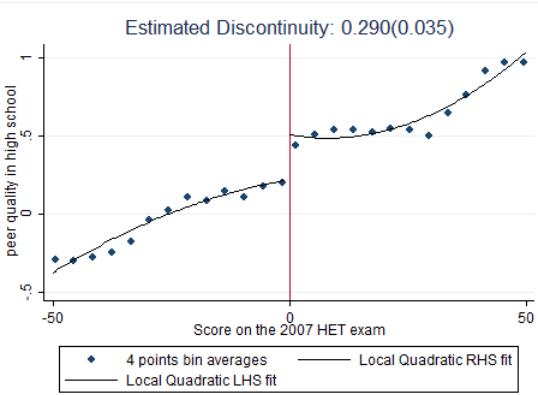


(a) Tier II school cutoff

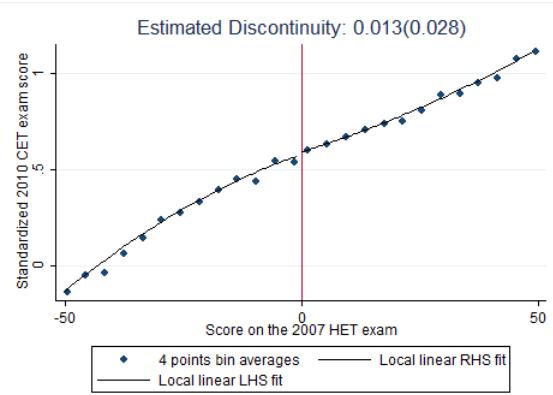
Standardized CET scores



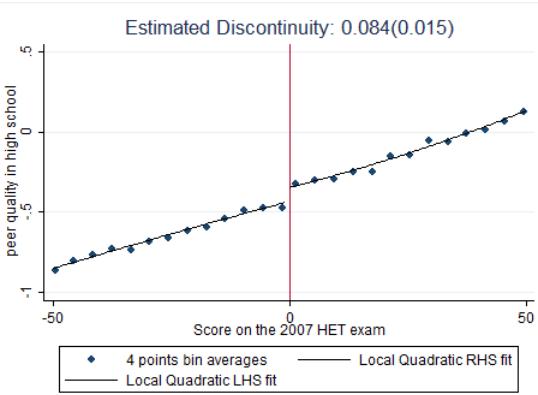
(b) Tier II school cutoff



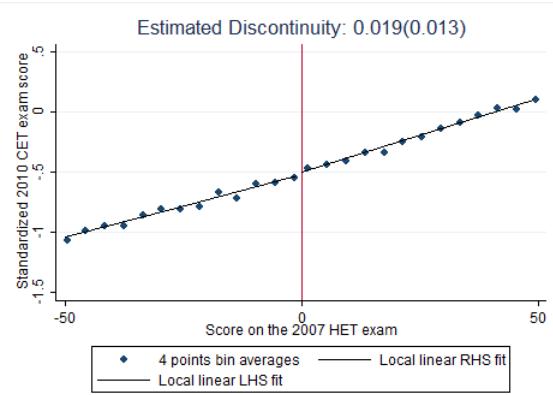
(c) Within Tier I school cutoffs



(d) Within Tier I school cutoffs



(e) Within Tier II school cutoffs



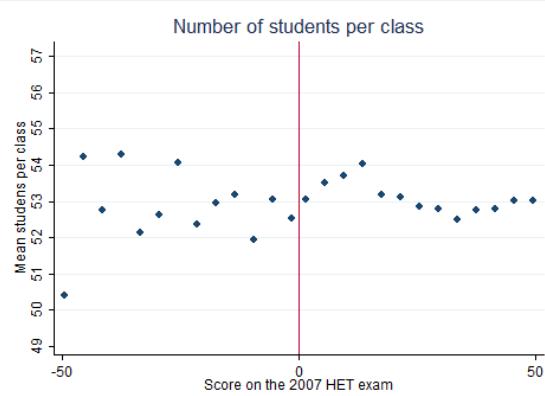
(f) Within Tier II school cutoffs

Notes: Sample includes students who took the HET exam in the year 2007.
48 Standard errors clustered at the score level.

Figure A5: Class size and teacher quality for other cutoffs

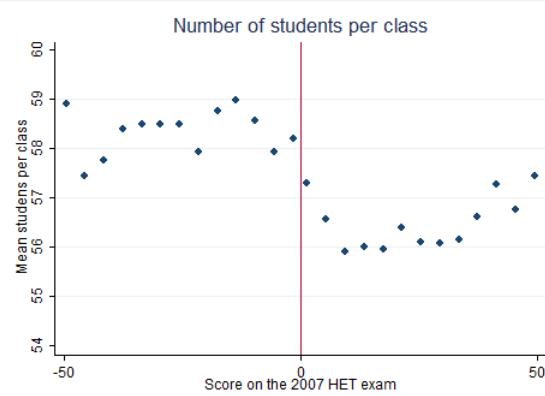
64

Tier II cutoff



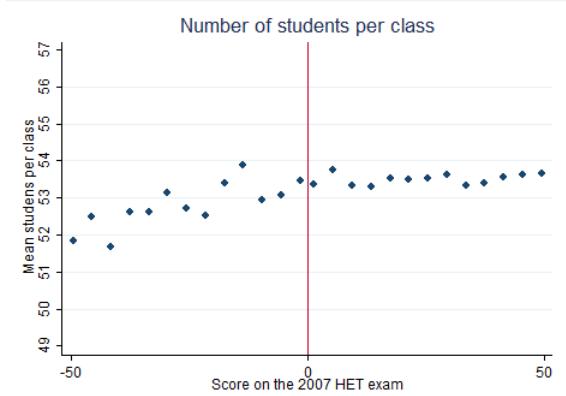
(a) Number of students per class

Within Tier I cutoffs



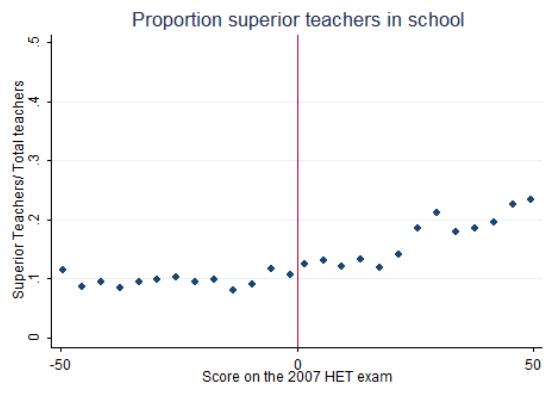
(b) Number of students per class

Within Tier II cutoffs



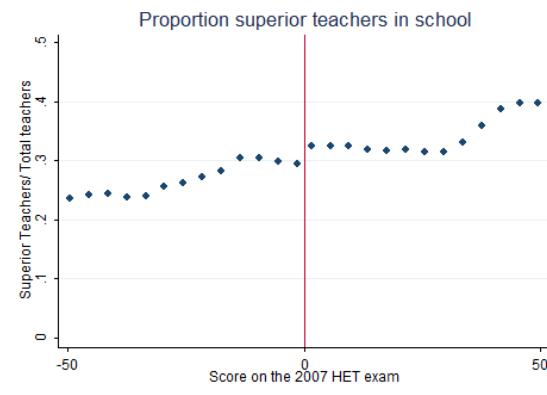
(c) Number of students per class

Proportion superior teachers in school



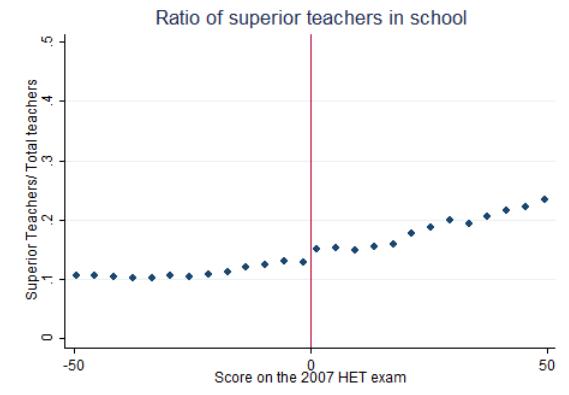
(d) Proportion superior teachers

Proportion superior teachers in school



(e) Proportion superior teachers

Ratio of superior teachers in school



(f) Proportion superior teachers

Notes: Samples based off of school level data.

D Appendix Tables

Table A1: Local polynomial RD estimates for baseline covariates

Bandwidth	2.5 CCT (1)	2 CCT (2)	1.5 CCT (3)	1.25 CCT (4)	1 CCT (5)	0.75 CCT (6)
Panel A: (Going to a better school)						
Predicted CET score	-.001 (.004)	.001 (.005)	-.001 (.006)	.001 (.006)	.003 (.007)	.007 (.008)
Likelihood of being a female	-.004 (.006)	-.006 (.007)	-.007 (.008)	-.007 (.008)	-.008 (.009)	-.001 (.010)
Age when taking HET entrance exam	.003 (.008)	.005 (.009)	.002 (.010)	.006 (.011)	.006 (.013)	-.004 (.014)
Observations for predicted score	75977	62474	47880	40676	32499	24931
Panel B: (Going to a top-school)						
Predicted CET score	-.007 (.013)	-.005 (.014)	.014 (.014)	.027 (.016)	.017 (.017)	.022 (.021)
Likelihood of being a female	.011 (.019)	.017 (.020)	.019 (.021)	.029 (.022)	.049** (.023)	.049* (.027)
Age when taking HET entrance exam	-.011 (.021)	-.012 (.022)	-.003 (.027)	-.006 (.029)	-.024 (.032)	-.051 (.036)
Observations for predicted score	6680	5578	4278	3642	2886	2237
Score Polynomial	One	One	One	One	One	One

Notes: Sample includes students who took the high school entrance exam in the year 2007 with known high school cutoffs.

Predicted HET score based on the following controls: sex, gender, district fixed effects, middle school fixed effect.

Optimal BW = 34 for predicted score, 34 for probability of being a female, 40 for age when taking HET exam (Going to a better school).

Optimal BW = 18 for predicted score, 26 for probability of being a female, 34 for age when taking HET exam (Going to a top school).

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

Table A2: “Stacked RD” estimates for all cutoffs except the Tier I cutoff

Bandwidth	18 points (1)	16 points (2)	14 points (3)
Panel A:			
Discontinuity in high school peer quality	.163*** (.026)	.158*** (.028)	.180*** (.030)
With Controls	.166*** (.02)	.161*** (.03)	.183*** (.03)
Panel B:			
Discontinuity in CET exam scores	.020 (.032)	−.002 (.035)	.014 (.037)
With Controls	.028 (.03)	.007 (.03)	.018 (.04)
Score Polynomial			
Observations	14,624	13,055	11,432

Notes: Sample includes students who took the college entrance exam in the year 2007.

Controls include: Age, gender, district fixed effects and middle school fixed effects.

The maximum possible bandwidth under this identification strategy is 18 test score points.

Since we observe individuals with multiple cutoffs, we cluster at the student ID level.

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

Table A3: Robustness Check—Adding third district to the sample

Treatment effect	Original Sample	Add district 183 (5 Tier I schools)	Add district 183 (1 Tier I school)
Panel A: Going to a better school			
High school peer quality	0.170*** (0.021)	0.180*** (0.017)	0.180*** (0.017)
College exam scores	-0.016 (.014)	-0.014 (0.022)	-0.014 (0.022)
Likelihood of enrolling in four year college	-.002 (.006)	.002 (.011)	.002 (.011)
Observations	37,961	53,334	53,334
Panel B: Going to a top school			
First Stage	.632*** (.036)	.477*** (.025)	.661*** (.035)
High school peer quality	0.303*** (0.026)	0.210*** (0.024)	0.350*** (0.028)
College exam scores	0.094*** (0.024)	0.083*** (0.022)	0.078*** (0.024)
Likelihood of enrolling in four year college	.071*** (.027)	.075*** (.026)	.054*** (.021)
Observations	6,046	8,056	8,056

Notes: Sample includes students who took the high school entrance exam in the year 2007.

All regressions include controls: gender, age, district fixed effects and junior high school fixed effects.

For ease of comparison, all regressions use an equal bandwidth of 40 score points on either side of the cutoff.

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