Peer Quality and the Academic Benefits to Attending Better Schools

Mark Hoekstra
Texas A&M University and NBER
markhoekstra@tamu.edu

Pierre Mouganie
American University of Beirut
pm10@aub.edu.lb

Yaojing Wang
Bank of America
yaojing.wang@bankofamerica.com

July 7, 2017

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 by comparing the college entrance exam scores of students in China who were 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. Further evidence suggests the returns to high school quality are driven by teacher quality, rather than peer quality or class size.

We would like to thank Stefanie Fischer, Jason Lindo, Steven Puller, Lori Taylor, Miguel Urquiola, Ron Zimmer, and two anonymous referees for their helpful comments and suggestions. We also thank seminar participants at Columbia University, Uppsala University, the European Economic Association annual meeting (EEA, 2016), the 1st IZA Economics of Education workshop, the Society of Labor Economists (SOLE, 2017) meetings, and the Association for Education Finance and Policy (AEFP, 2016) meetings for helpful comments and discussions. All errors are our own.

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 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, 2016; Canaan and Mouganie, Forthcoming; Hoekstra, 2009; Saavedra 2008; 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; Dee and Lan, 2015; Dobbie and Fryer, 2014; Lucas and Mbiti, 2014; Zhang, 2016). In contrast, others find that attending more selective schools does result in benefits (Berkowitz and Hoekstra, 2011; Clark and Del Bono, 2016; Ding and Lehrer, 2007; Jackson, 2010; Park et al., 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. The purpose of this paper is to contribute to this literature by estimating the returns to school quality across a range of high schools with different levels of selectivity, and to use data on peer quality, class size, and teacher quality to speak to the source of any heterogeneity in effects found within this context.

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.

The primary advantage of this approach 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, because of the high demand for attending four-year colleges, and because the college entrance exam is the determining factor for university admissions, both teachers and students face strong incentives to do well on the college entrance exam. As a result, exam performance should be a good measure of whether students learn more at better schools, which contrasts with outcomes examined in other settings.

Our data and context provide several advantages relative to previous work on school quality in China. For example, we obtained administrative data directly from the Ministry of Education, and observe college entrance exam scores for 91 percent of students, which is consistent with the officially reported range using aggregate data. By comparison, Park et al. (2015) use data requested from individual schools, and are missing outcome scores for nearly 40 percent of the students. A second advantage of our study relative to previous work in China is increased statistical power; we are able to detect effects of attending Tier I schools of magnitudes that lie within the 95 percent confidence intervals of previous studies that report no effects (Dee and Lan, 2015; Zhang, 2016).[1]

Another important advantage of our study relative to the existing literature in both China and elsewhere is that we are able to test directly for discontinuities in important educational inputs across the different admission thresholds. This enables us to provide evidence on the mechanism underlying any heterogeneity in benefits across different admission thresholds. While this approach cannot enable a definitive conclusion regarding the exact mechanism(s) through which school quality impacts achievement, we argue that our findings are suggestive nonetheless. That is in part because we study the heterogeneity in benefits within a single school district, rather than across cities or countries. As a result, many of the factors that could explain the different findings documented in the existing literature–such as differences in institutions or behavioral responses–are less likely to explain the differences across admission thresholds in our setting. In addition, we are able to test directly for discontinuities in three major educational inputs expected to be of first-order importance, including peer quality, class size, and teacher quality. Importantly, our measure of teacher quality is the concentration of "superior" teachers, which 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.

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. 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 the most selective group of schools in our sample, Tier I schools, we find

[1] We note that the focus of Zhang (2016) is on extending the local average treatment effect (LATE) framework to contexts where there is imperfect matching between the instrument and the treatment and outcome variables. This method is then applied to the question of whether attending elite middle schools affects middle school exit exam scores.

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 labor market 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 academic benefits 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 the benefits from attending more selective high schools 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 threshold. In contrast, while there are large discontinuities in peer quality across all other thresholds—including within Tier I, within Tier II, and at the Tier II threshold—there are at most very small discontinuities in access to superior teachers across those same cutoffs. We also 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 importance of teacher quality in explaining the benefits (or lack thereof) of attending more selective schools is also confirmed by a school-level analysis. For each admission cutoff in our sample, we estimate discontinuities in CET scores as well in peer quality, teacher quality, and class size, and show that school value-added is only explained by discontinuities in teacher quality.

These findings have important implications both for the literature and for public policy. 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, our findings indicate 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, the results provide some evidence that at least in this context, improving teacher quality may be the most effective way policymakers can replicate the benefits associated with attending more selective schools.

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 two districts 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"

[2]This 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.

[3]Chinese, 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.

where additional resources are targeted toward the top 15 percent 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 two districts we study is centrally operated by each district'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 school districts, 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 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

[4]For example, see http://zhongkao.gaofen.com/article/448289.htm.

[5]For 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.

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]

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

[6]Public 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.

[7]Some schools may choose to keep their students longer after hours, such as nights or weekends, particularly 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.

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.

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 one of those eight districts 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 the four 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 set of schools. For these reasons, we restrict our analysis to these suburban districts.

2.3 High School Teachers

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

[8]For 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 selective college cutoff in the CET and 60% must meet the four-year college cutoff.

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 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

[9]Master 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.

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

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

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 more than 10 million and a per capita GDP of more than $20,000. The two districts 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. We do so because students in suburban districts must

[12] This 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

[13] According 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.

attend a school in that district, which results in significant sorting across schools by peer ability. In contrast, because students in urban districts can attend a school in *any* urban district, there is insufficient sorting to generate a discontinuity in peer quality across admission thresholds. 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 12,259 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 districts 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 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

[14]There are four suburban districts 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.

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

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. Accordingly, we use a regression discontinuity design (Lee and Lemieux, 2010; Imbens and Lemieux, 2008) to estimate the causal impact of Tier I high school attendance 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 districts.[16] Accordingly, we have two Tier I cutoffs in our data—one for each district. In order to summarize the effects of attending Tier I school, we pool data across both districts. 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 district-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

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

[17] We have two thresholds, with each representing a different district.

the cutoff score of each district. Formally, $S_i = grade_i - \overline{grade_z}$ for all individuals within a district facing a common Tier I cutoff z. The function h(.) 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(.) 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(.) 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 specifications 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, for single cutoff results, 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 districts 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 district serves as a separate observation for each cutoff.[19] Due to repeated observations when using multiple cutoffs, we cluster our standard errors at the individual level. Formally, our reduced form regression from this procedure takes the following form:

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

Here the subscript i still refers to students and the subscript z refers to all possible high school cutoffs facing an individual within a district (i.e. z= 1,...,H-1; where H represents the total number of high schools in that district ordered from worst to best based on their respective cutoff scores).

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

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

ω gives us the intent-to-treat 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 district 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.

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 would be able to manipulate the assignment variable in 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 Figure 1(a) and Figure 1(b), which show the density function for the stacked RDD across all admission cutoffs as well as for only the Tier I admission threshold. Importantly, based on the the McCrary (2008) density test statistic reported in the figures we fail to reject the null hypothesis of no discontinuity. 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 Figures 1(c) and 1(d), and indicate that there is little evidence that underlying student ability varies discontinuously across the thresholds. Estimates shown in Appendix Table A1 are also close to zero and statistically insignificant across a range of bandwidths.

[20] In 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 summarized 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 2(a) shows results for the likelihood of attending the more selective school, while Figure 2(b) shows the discontinuity in peer quality, defined as the average high school entrance exam score of students in the school in which the student enrolled. These figures show compelling visual evidence that threshold-crossing leads 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 Panels A and B of Table 2. Results are shown using bandwidths ranging from 2.5 times optimal bandwidth to three-quarters of optimal bandwidth. Estimates are also shown with and without controls. Local linear 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 2(c) 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 more selective schools with significantly higher-performing peers, they do not achieve at higher levels as a result. Estimates across a range of bandwidths and specifications in Panel C 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 2(d) and indicate that attending better schools does not lead to improved access to four-year colleges. Corresponding estimates in Panel D of Table 2 are similar. In short, there is little evidence that attending more selective schools with better peers 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 Tier I Schools

We now turn to examining the returns to attending Tier I schools, which is the most selective set of schools in our data. We do so since much of the prior literature has focused on the benefits from attending the most selective schools. To put the Tier I schools in our setting in the broader context of schools studied by the existing literature, we note that in our data the top one-third of students attend a Tier I school. This makes the Tier I schools in our setting more selective than the schools studied by Dee and Lan (2015) (top 60 percent) and Park et al. (2015) (top 50 percent), and similar to those studied by Clark and Del Bono (2016) (top 30 percent) and Zhang (2016) (top 21 percent). In contrast, by this same measure Tier I schools are much less selective than the Kenyan schools studied by Lucas and Mbiti (2014) and the New York City schools studied by by Abdulkadiroglu et al. (2014) and Dobbie and Fryer (2014) that are attended by the top 5 percent.[21]

Results are shown graphically in Figure 3. Figure 3(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. We note that these first-stage estimates are significantly larger than most in the existing literature, suggesting that there is a higher rate of sorting and compliance in our setting than in others.[22]

Figure 3(b) shows that threshold-crossing leads to large increases in peer ability of around one-third of a standard deviation. This reflects that on average Tier I schools are attended by

[21]In New York City and Boston, the only students to take the entrance exam are those who wish to attend the more selective schools. Consequently, to assess the degree of selectivity, we compare average SAT scores for students in those schools to the statewide distribution of scores. In addition, we note that the two Boston exam schools also studied by Abdulkadiroglu et al. (2014) are less selective than their NYC counterparts; their students rank at the 40th and 80th percentiles of the state SAT distribution.

[22]For example, the estimated first stages are 20 percentage points for Zhang (2016), 28 percentage points for Dee and Lan (2015), 33 to 52 percentage points for Park et al. (2015), 40 percentage points for the least selective NYC exam school studied by Abdulkadiroglu et. al (2014) and Dobbie and Fryer (2014), 49 percentage points for Lucas and Mbiti (2014), 70 percentage points for the least selective Boston exam school studied by Abdulkadiroglu et. al (2014), and 75 percentage points for Clark and Del Bono (2016).

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.

Figure 3(c) shows that in contrast to the results across all admission thresholds, being eligible to attend a 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.

Figure 3(d) 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, and Zhang (2012) 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 in Panel C indicate that Tier I school attendance leads to a 0.155 standard deviation increase in overall CET scores and an 11.9 percentage point increase in the likelihood of attending a four-year college, both of which are driven by effects for boys rather than girls. Importantly, while we find these estimates to be statistically and economically significant, we note that effects of this magnitude are not statistically detectable in previous analyses of selective schools in China.[23]

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.[24] 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 this context, 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 A3(a) and A3(b). Both figures show that the likelihood of majoring in arts versus science is smooth across all admission thresholds (Figure A3(a)) as well as the cutoff for Tier I schools (Figure A3(b)).[25]

[23]For example, all the 95 percent confidence intervals for the main estimates in row one of Table 6 in Zhang (2016) contain both zero and our local average treatment effect estimate of 0.155. Similarly, while Dee and Lan (2015) do not report local average treatment effects of attending the Chinese magnet schools they study, we can approximate them by rescaling their reduced-form estimates and standard errors in the last row of Table 2 by the magnitude of the first stage on enrollment shown in row one. The resulting estimates are not statistically different from zero or from our estimate of 0.155.

[24]We test for the equality of the estimates using seemingly unrelated regression; the chi-squared statistic is 7.73 and resulting p-value is 0.0054.

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

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 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.[26] Results are shown in Appendix Figures A3(c) and A3(d), which show that age is smooth across both cutoffs, indicating that grade repetition is unlikely to explain our findings.

One might also be concerned that the different estimation strategies (stacked versus single cutoff) could itself drive the differences in findings at the Tier I cutoff versus other cutoffs. However, we find no evidence that this is the case.[27]

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 a Tier I school) made it more or less likely for the student to take the college entrance test, then our estimates could be biased. To test for selection into sitting for the exam, we match 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.

Results are shown in Figure 5, and indicate that while going to *any* better school is not associated with differential college entrance test-taking, barely attending a Tier I high school does appear to lead to a higher rate of test-taking. Additional results in Figure 5(f) 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 5(d)).[28] 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

[26]Grade 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.

[27]For example, if we create a placebo tier cutoff using the highest scoring cutoff within Tier II and estimate effects the same way as we did for the Tier I cutoff, we find no evidence of a discontinuity in CET (CCT estimate = -0.017, se=0.038) despite a significant difference in peer quality (estimate = 0.124, se=0.029). The same is true if we create a placebo tier cutoff separating the most selective schools from the least selective schools within Tier III (CCT estimate for CET = -0.026, se=0.038). In addition, if estimation strategy were the cause of the discontinuity in CET at the Tier I cutoff, we would expect to find a discontinuity in the CET at the Tier II cutoff as well. As shown in Figure A4(b), we find no such discontinuity. Finally, we note that we estimate very similar effects using the stacked RDD approach even if we use a bandwidth of 10, which ensures that the use of repeated observations is extremely rare. In that case, while the estimated effect on peer quality is 0.204 (se=0.031), the estimate for CET scores is still insignificant at 0.043 (se=0.038).

[28]This 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.

better schools overall, students who are barely admitted to Tier I 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 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.

Because the discontinuity in the likelihood of taking the test is most likely due to differences in whether to sit for the exam, rather than attrition from the district, we address this concern by defining new outcomes as indicators for whether each student has exceeded a given CET score threshold. These outcomes include scoring in the top 25 percent, 50 percent, or 75 percent of the overall distribution, as well as scoring high enough to exceed the cutoff for college eligibility (the 48th and 69th percentiles of the CET distribution for science and arts, respectively). Results are shown in Table 6, where Panel A shows results for the original sample that excludes non-test takers. Panel B shows results for all students, and assigns zeros to those who did not take the exam and thus did not exceed the threshold. Results from Panel A indicate that attending Tier I schools results in a 7 to 8 percentage point increase in college eligibility, and shifts students from the bottom quartile into the second and third quartile. Importantly, results from Panel B are very similar, and indicate that threshold-crossing is associated with being significantly more likely to score in the top 50 and 75 percent of the overall distribution.

In addition, we also perform formal bounding exercises to assess the degree to which this selection into test-taking could affect our results. Specifically, we use a trimming procedure in the spirit of that suggested by Lee (2009).[29] Results are shown in Appendix Table A4, and indicate that the lower and upper bound estimates of the return to attending Tier I high school remain positive and significant, and range from 5 to 12 percent of a standard deviation, compared to

[29]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 4 HET points of the cutoff, which results in more extreme bounds than when we drop the top distribution of students in each bin in the treatment group regardless of distance to threshold. 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. For example, using a local linear regression of bandwidth equal to 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. A similar procedure—trimming the bottom performing students in the treatment group—is used to estimate the upper bound.

original estimates ranging from 7 to 9 percent of a standard deviation.

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. In particular, we focus primarily on the three determinants of achievement that the existing literature has demonstrated to be most important: peer quality, class size, and teacher quality.

With respect to peer quality, we note that 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 summarized in Figure 6, which stacks together all cutoffs other than the Tier I cutoff.[30] Corresponding estimates in appendix Table A2 indicate that while barely admitted students attend schools with students who scored a statistically significant 0.16 standard deviations higher on the high school entrance exam, they score only 0.020 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. Figure A4(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 around one-third of a standard deviation 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) reveals around a ten percent of a 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 results in an improvement of 7 to 9 percent of a standard deviation in

[30] Because 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.

CET scores, barely attending better schools across all other cutoffs does not result in better CET scores. However, there are large and significant discontinuities in peer quality across both sets of cutoffs. 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, Appendix Figure A4(c) 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 A4(d) 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 schools with better peers, 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.[31] Unfortunately, our data do not allow us to test for these behavioral responses directly. However, we note that for this to explain the heterogeneity in our findings, parents or students would have to respond differently across the Tier I cutoff than they would across cutoffs within Tier I, within Tier II, or across Tier II and III.

Next, we turn to whether the heterogeneity in returns can be driven by differences in school resources. In particular, while we are unable to obtain data on per-pupil spending at the school level, we do have data on arguably the two most important inputs through which school resources could directly affect achievement. The first of these is class size, which has received considerable

[31] An 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.

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 7(a), and indicate that students who barely attend Tier I schools are in significantly *larger* classes (56 versus 54 students). In contrast, Figure 8(a) shows that across all other cutoffs, class size is no different. If anything, that suggests that the 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.

The final input we examine 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.[32] In addition, Lai, et al. (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 7(b), and indicate that there is a large discontinuity of 10.8 percentage points in the proportion of superior teachers. In contrast, Figure 8(b) 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 A5(d), A5(e), and A5(f) 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 with a large increase in CET scores.

These findings are further summarized in Figure 9, which shows estimated discontinuities in the probability of attending the better school, average peer quality, performance on the CET, class size, and ratio of superior to normal teachers for five different cutoffs. Figure 9 shows that while there are significant increases in peer quality across the different sets of admission thresholds, the only cutoff to show benefits to attending (Tier I) is also the only cutoff where we estimate a significant increase in teacher quality. In addition, a back-of-the-envelope calculation suggests it is plausible that differential exposure to superior teachers can explain all of the improved achievement

[32]In our sample, only 5 of the teachers (0.13%) have achieved special grade teacher rank, or "Teji Jiaoshi".

we observe 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 the same as the increase in achievement found to result from a two standard deviation increase in teacher quality (Chetty, Friedman, and Rockoff, 2014; Aaronson, Barrow, and Sander, 2007; Kane, Rockoff, and Staiger, 2008; Rockoff 2004; Rivkin, Hanushek, and Kain 2005). We also emphasize that there are likely other improvements in teacher quality at the Tier I admission threshold that are more difficult to measure.

The (discontinuous) concentration of superior rank teachers within Tier I high schools is consistent with Lankford, Loeb, and Wykoff (2002) and Jackson (2009), who show that more qualified and higher value-added teachers sort into schools based on student ability and demographics. In addition, superior rank teachers may also prefer the prestige associated with working at a Tier I school in our sample, all of which are recognized as national demonstrative schools. Financial considerations could also play a role, as part of teacher salary is determined locally and can vary at the school level.

Finally, we perform one additional exercise to further examine the role of peers, class size, and superior teachers in explaining the heterogeneity in academic benefits. Specifically, we estimate the discontinuity in CET scores at the admission cutoff for each of the 23 schools, and regress those estimates on estimated discontinuities in peer quality, teacher quality, and class size.[34] In doing so, we ask how well each potential mechanism explains the school-specific value-added in CET scores. The advantage of this approach is that it uses all of the across-school variation in both value-added and mechanisms. The downside is that because these are individual school estimates, rather than stacked RDD estimates, each discontinuity is estimated with significantly more error.

The resulting estimates are shown in Table 7. Consistent with our previous findings, the only factor that is correlated with academic benefits in a (marginally) significant way is teacher quality, as measured by the discontinuity in the proportion of superior rank teachers. The estimate in column (1) implies that a one standard deviation increase in the proportion of superior teachers (0.16) results in a 0.078 (0.16*0.489) standard deviation increase in CET scores. In column (2), we show results when we include all four suburban districts and show that the relationship between teacher quality and school value-added is somewhat larger. Our view is that these findings are suggestive of a relationship between teacher quality and school value-added.

In summary, we conclude that there is suggestive evidence that the academic benefits from attending Tier I schools are due to increases in teacher quality. The interpretation of the effects as

[33]Estimates 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 the estimate of 0.25.

[34]We thank an anonymous referee for this suggestion.

being due to teacher quality is 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.[35] 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.[36] It is more difficult for us to determine why there is more sorting of teachers across the Tier I cutoff than across the Tier II cutoff or the cutoffs within Tiers I and II. We hypothesize that while some of this is likely due to perceptions of increased prestige associated with teaching at a Tier I school, it may also be due to increased teacher pay at the Tier I cutoff. As discussed earlier, one element of teacher pay is determined locally and can vary from school to school, though unfortunately our data do not allow us to test directly for discontinuities in teacher salaries. In addition, we note that while we think the results here are most consistent with the interpretation that heterogeneity in benefits is due to differences in teacher quality, we cannot rule out the role of unmeasured determinants of CET scores that are correlated with teacher quality.

7 Conclusion

This paper estimates the cognitive benefits due to attending more selective 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. Specifically, we find that attending a Tier I school with peers who score one-third of a standard deviation higher on the high school entrance exam is associated with a 7 percent of a standard deviation increase in performance on the college entrance exam. In contrast, we find that even though threshold-crossing elsewhere is also associated significant increases in peer quality, we find no evidence of any academic benefits. 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.

[35]It is also consistent with how little discretion principals at the schools studied by Abdulkadiroglu et al. (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.

[36]While 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.

We provide additional evidence suggesting that the returns to attending Tier I schools are due to teacher quality. We do so in part by estimating discontinuities for each admission cutoff in our data, and then showing that these school-specific discontinuities in CET scores are explained only by discontinuities in teacher quality, rather than peer quality or class size. In addition, the only discontinuity in access to superior teachers within or across tiers is at the Tier I cutoff, which is also where we observe the only discontinuity in CET scores. 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 around 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 et al. (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.

We note that it remains an open question as to whether the findings here extend to other contexts, and in particular to contexts outside of Asia where the education system may be substantially different. With that caveat, however, we think an important contribution of this study is to highlight the importance of measuring additional education inputs when studying school quality. Most existing studies have focused on peer quality as a measure of school quality, since that is a readily available measure of selectivity. But the results here show that the heterogeneity in benefits across schools is difficult to explain with differences in peer quality. In addition, the extent to which the benefits from attending more selective schools are due to peer quality versus other school inputs is critical for public policy. If the benefits were due to the presence of higher-performing peers, it would be difficult to extend those benefits more broadly since there is a limited set of those peers. On the other hand, if the benefits are due to factors such as teacher quality, as suggested in this setting, then it may be possible to replicate the success of selective schools.

References

Aaronson, Daniel, Lisa Barrow, and William Sander. 2007. Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics* 25, no. 1:95-135.

Andrews, Rodney J., Jing Li, and Michael F. Lovenheim. 2016. Quantile treatment effects of college quality on earnings. *Journal of Human Resources* 51, no. 1:200-238.

Abdulkadiroglu, Atila, Joshua Angrist, and Parag Pathak, 2014. The elite illusion: Achievement effects at Boston and New York exam schools. *Econometrica* 82, no. 1:137-196.

Angrist, Joshua, and Victor Lavy. 1999. Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics* 114, no. 2:533-575.

Berkowitz, Daniel, and Mark Hoekstra. 2011. Does high school quality matter? Evidence from admissions data. *Economics of Education Review* 30, no. 2:280-288.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2014. Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica* 82, no. 6:2295-2326.

Canaan, Serena, and Pierre Mouganie. Forthcoming. Returns to education quality for low skilled students: Evidence from a discontinuity. *Journal of Labor Economics.*

Chetty, Rajj, John N. Friedman, and Jonah E. Rockoff. 2014. Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review* 104, no. 9:2593-2632.

Clark, Damon. 2010. Selective schools and academic achievement. *B.E Journal of Economics Analysis and Policy: Advances in Economic Analysis and Policy* 10, no. 1.

Clark, Damon, and Emilia Del Bono. 2016. The long-run effects of attending an elite school: Evidence from the United Kingdom. *American Economic Journal: Applied Economics* 8, no. 1:150-176.

Dee, Thomas, and Xiaohuan Lan. 2015. The achievement and course-taking effects of magnet schools: Regression-discontinuity evidence from urban China. *Economics of Education Review* 47, (2015):128-142.

Ding, Weili, and Steven F. Lehrer. 2007. Do peers affect student achievement in China's secondary schools? *Review of Economics and Statistics* 89, no. 2:300-312.

Dobbie , Will, and Roland G. Fryer. 2014. The impact of attending a school with high-achieving peers: Evidence from the New York city exam schools. *American Economic Journal: Applied Economics* 6, no. 3:58-75.

Giles, John, Albert Park, and Meiyan Wang. 2015. The great proletarian cultural revolution, disruptions to education, and the returns to schooling in urban China. *IZA Discussion Paper No. 8930.*

Emily, Hannum, and Albert Park, 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, Mark. 2009. The effect of attending the flagship state university on earnings: A discontinuity-based approach. *The Review of Economics and Statistics* 91, no. 4:717-724.

Imbens, Guido W., and Thomas Lemieux. 2008. Regression discontinuity designs: A guide to practice. *Journal of Econometrics* 142, no. 2:615-635.

Jackson, C. Kirabo. 2009. Student Demographics, Teachers sorting, and teacher quality: Evidence from the end of school desegregation. *Journal of Labor Economics* 27, no. 2:213-256.

Jackson, C. Kirabo. 2010. Do students benefit From attending better schools? Evidence From rule-based student assignments in Trinidad and Tobago. *Economic Journal* 120, no. 549:1399-1429.

Jackson, C. Kirabo. 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, Thomas J., Jonah E. Rockoff, and Douglas O. Staiger. 2008. What does certification tell us about teacher effectiveness? Evidence from New York city. *Economics of Education Review* 27, no. 6:615-631.

Krueger, Alan B. 2003. Economic considerations and class size. *Economic Journal* 113, no. 485:F34-F63.

Lai, Fang, Elisabeth Sadoulet, and Alain De Janvry. 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, no. 1:123-153.

Lankford, Hamilton, Loeb, Susanna, and Wykoff, James, 2002. Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis* 24, no. 1:37-62.

Lee, David S. 2009. Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *Review of Economic Studies* 76, no. 3:1071-1102.

Lee, David S., and David Card. 2008. Regression discontinuity inference with specification error. *Journal of Econometrics* 142, no. 2:655-674.

Lee, David S., and Thomas Lemieux, 2010. Regression discontinuity designs in economics.*Journal of Economic Literature* 48, no. 2:281-355.

Li, Hongbin, Pak Wak Liu, and Junsen Zhang. 2012. Estimating returns to education using twins in urban China. *Journal of Development Economics* 97, no. 2:494-504.

Lucas, Adrienne M., and Isaac M. Mbiti. 2014. Effects of school quality on student achievement: Discontinuity evidence from Kenya. *American Economic Journal: Applied Economics* 6, no. 3:234-63.

McCrary, Justin. 2008. Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics* 142, no. 2:698-714.

Park, Albert, Shi Xinzheng, Chang-tai Hsieh, and Xuehui An. 2015. Magnet high schools and academic performance in China: A regression discontinuity design. *Journal of Comparative Economics* 43, no. 4:825-843.

Pop-Eleches, Christian, and Miguel Urquiola. 2013. Going to a better school: Effects and behavioral responses. *American Economic Review* 103 no. 4:1289-1324.

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

Rockoff, Jonah E. 2004. The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review: Papers and Proceedings* 94, no .2:247-52

Saavedra, Juan E. 2008. The learning and early labor market effects of college quality: A regression discontinuity analysis. *Unpublished manuscript, University of Southern California.*

Zimmerman, Seth D. 2014. The returns to college admission for academically marginal students. *Journal of Labor Economics* 32, no. 4:711-754.

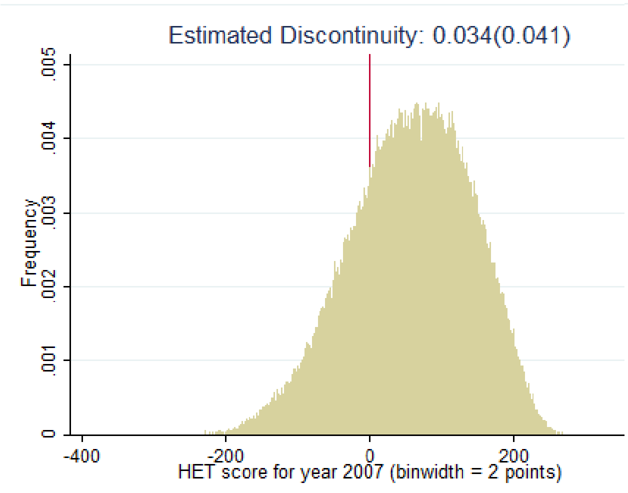
Zhang, Hongliang. 2016. Identification of local average treatment effects under imperfect matching with an application to Chinese elite schools. *Journal of Public Economics.* 142, (2016):56-82.

A Figures

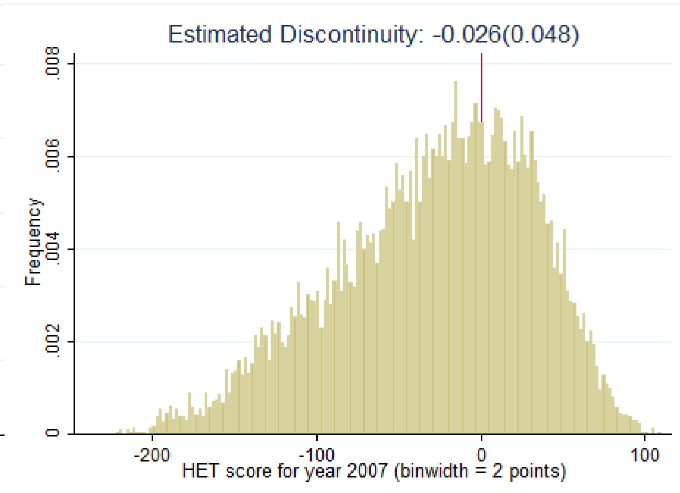
Figure 1: Testing the validity of the RD design for both empirical strategies

Attending Better Schools (All Cutoffs)

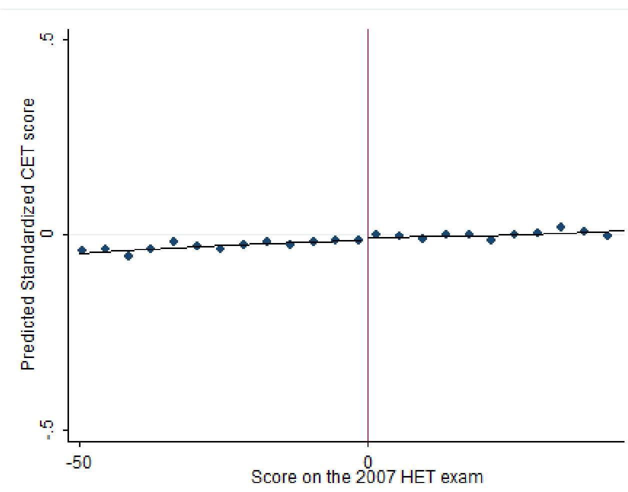
Attending a Tier I school



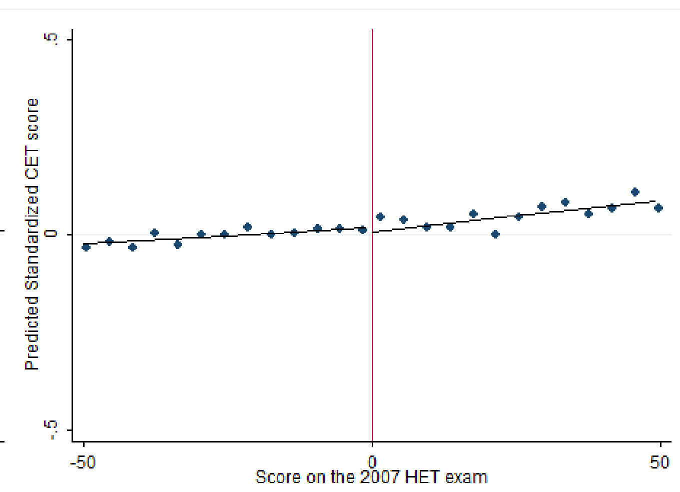
(a) Distribution of HET scores



(b) Distribution of HET scores



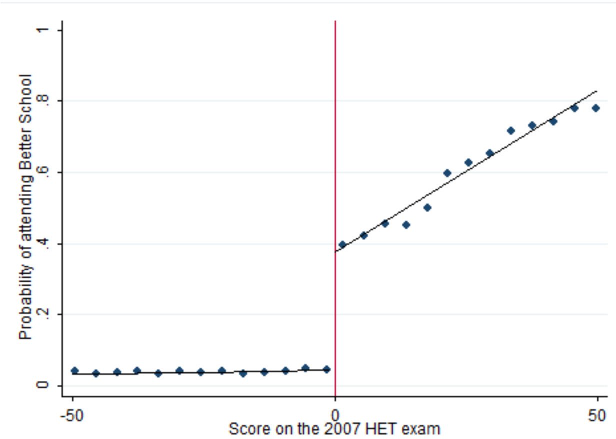
(c) Smoothness of baseline covariates



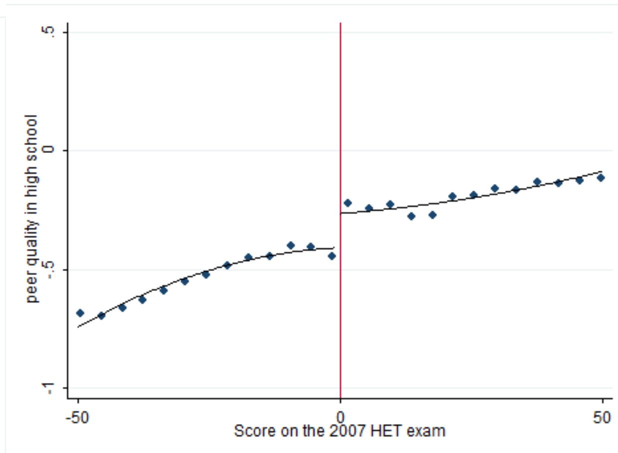
(d) Smoothness of baseline covariates

Notes: Sample includes students who took the high school entrance exam in the year 2007 and the college entrance exam in 2010.
Bins for both histograms represent an average count of 2 score points. Estimated discontinuity for histograms computed using the McCrary density test.
Predicted score based on the following controls: age, gender, district fixed effects, and middle school fixed effects.

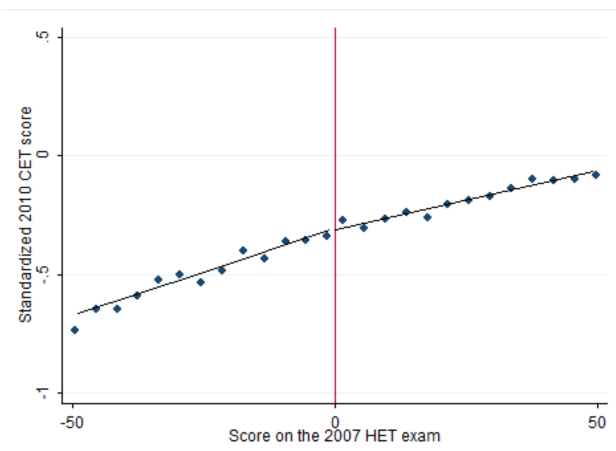
Figure 2: Discontinuities across all admission thresholds



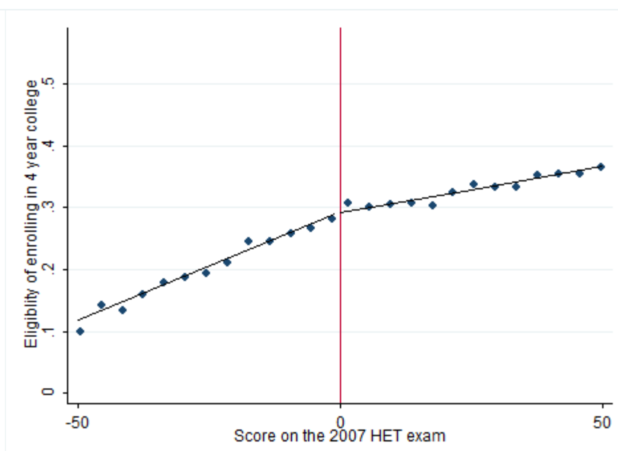
(a) Probability of attending better school



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



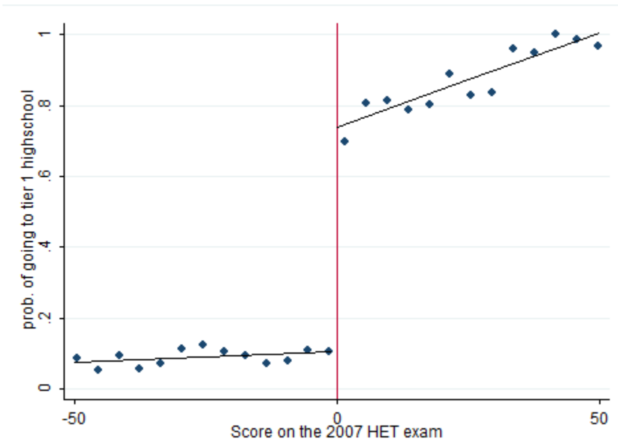
(c) College entrance exam test scores



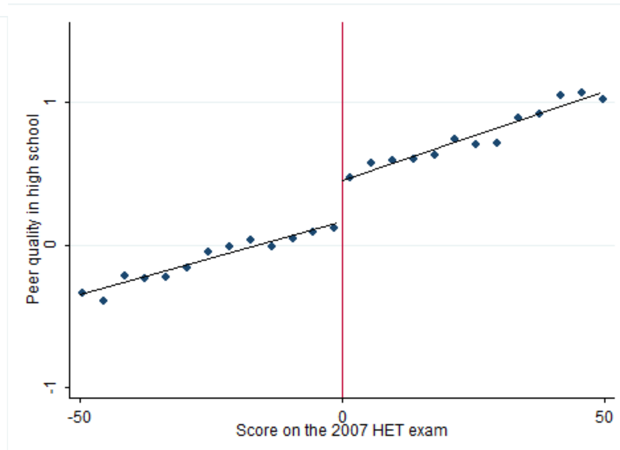
(d) Eligibility to attend a four year college

Notes: Sample includes students who took the HET exam in the year 2007 and the CET exam in the year 2010.

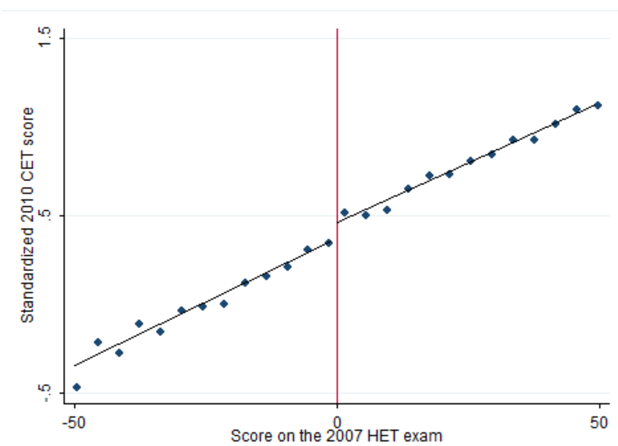
Figure 3: Discontinuities across the Tier I admission threshold



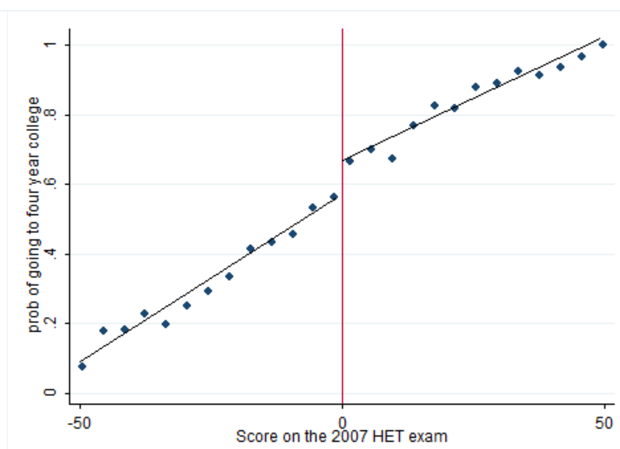
(a) Probability of attending Tier I high school



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



(c) College entrance exam scores

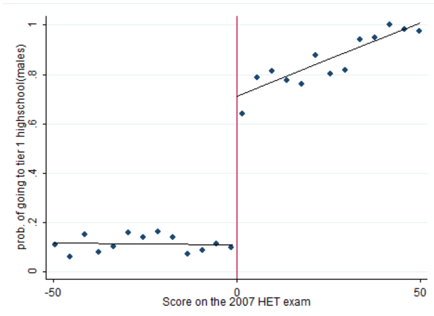


(d) Eligibility to attend a four year college

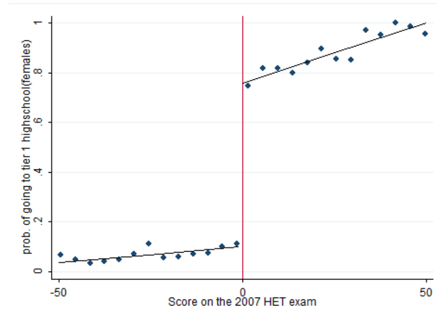
Notes: Sample includes students who took the HET exam in the year 2007 and the CET exam in the year 2010.

Figure 4: Discontinuities across the Tier I admission threshold, by gender

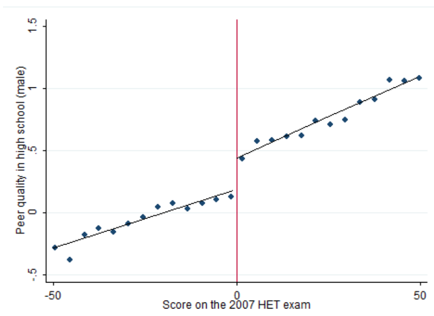
Males | Females



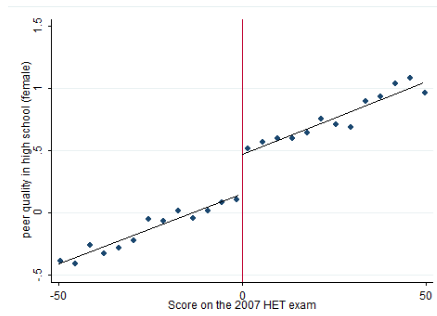
(a) Probability of attending Tier I high school



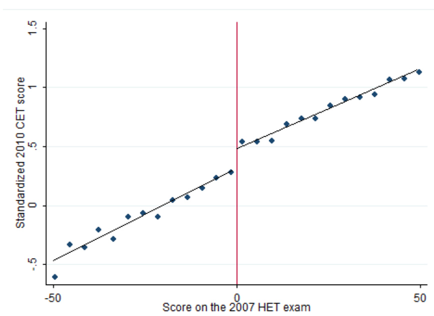
(b) Probability of attending Tier I high school



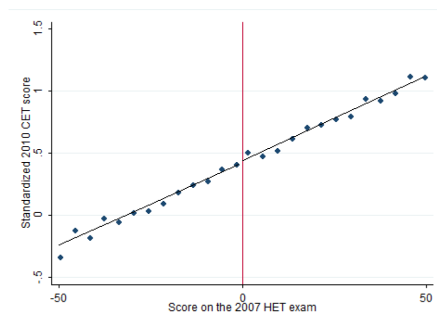
(c) Peer quality in high school



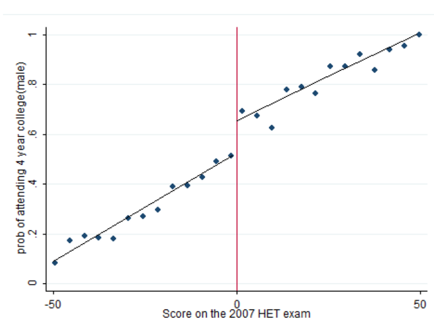
(d) Peer quality in high school



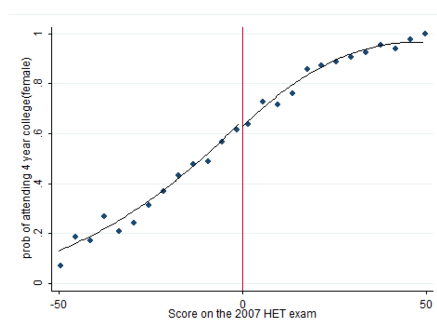
(e) CET exam scores



(f) CET exam scores



(g) Likelihood of enrolling in 4-year college

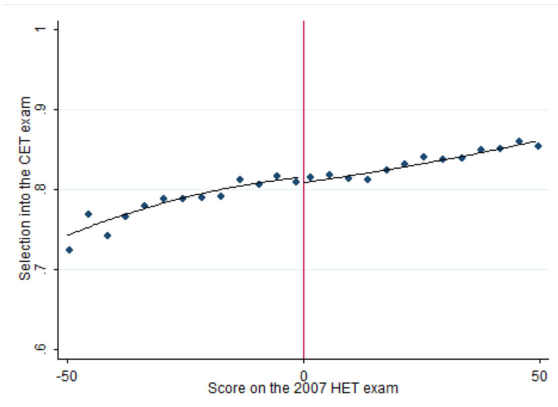


(h) Likelihood of enrolling in 4-year college

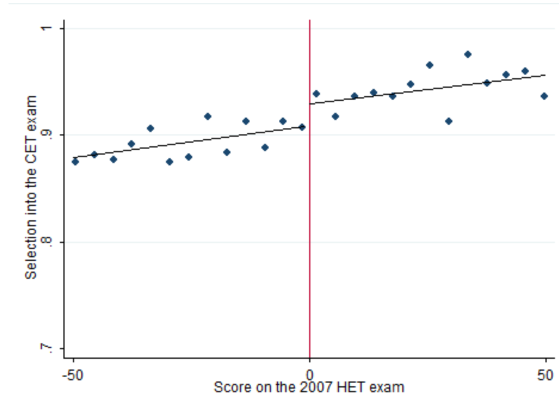
Notes: Sample includes students who took the HET exam in the year 2007 and the CET exam in the year 2010.

Figure 5: Selection into the college entrance exam

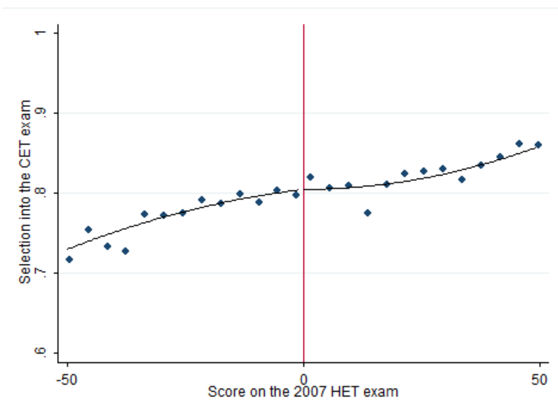
Attending Better Schools (All Cutoffs) | Attending a Tier I School



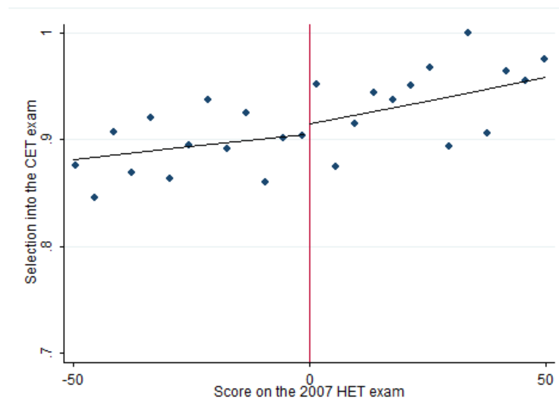
(a) Selection into the CET exam



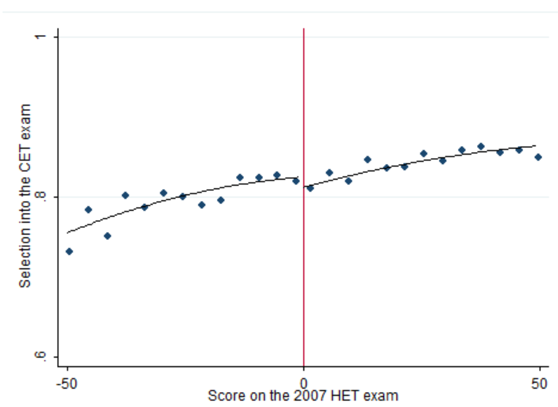
(b) Selection into the CET exam



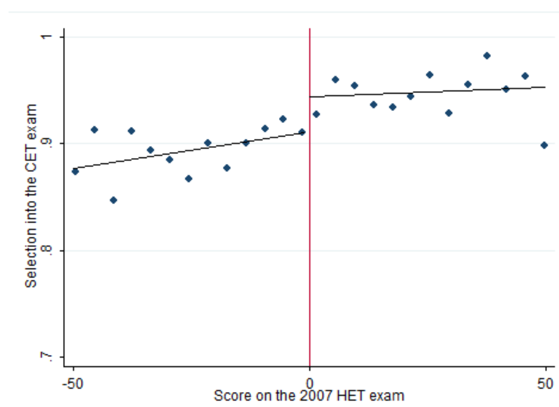
(c) Selection into the CET exam (Males only)



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



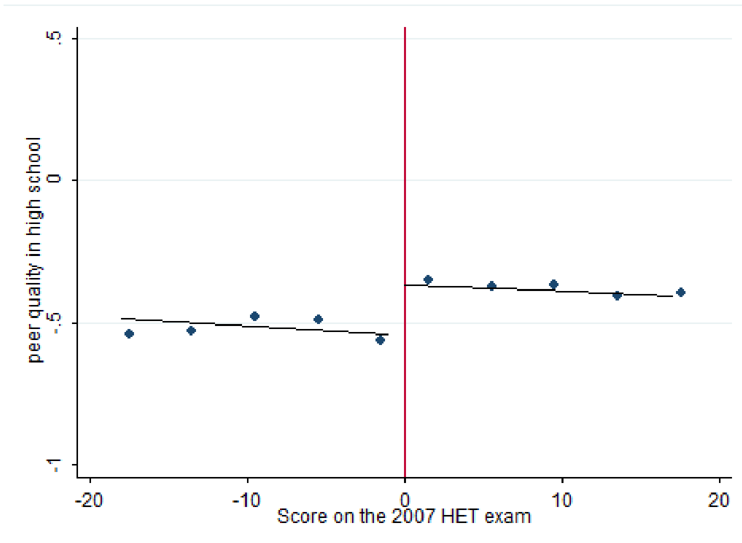
(e) Selection into the CET exam (Females only)



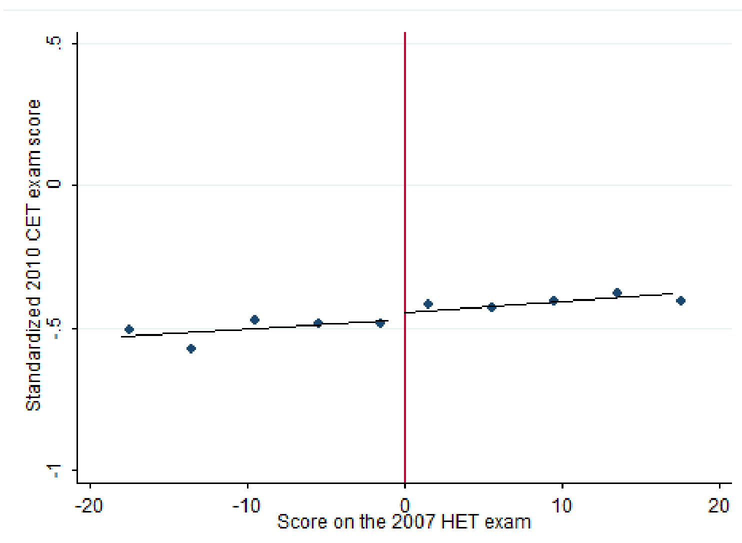
(f) Selection into the CET exam (Females only)

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

Figure 6: Discontinuities across all admission thresholds, excluding the Tier I cutoff



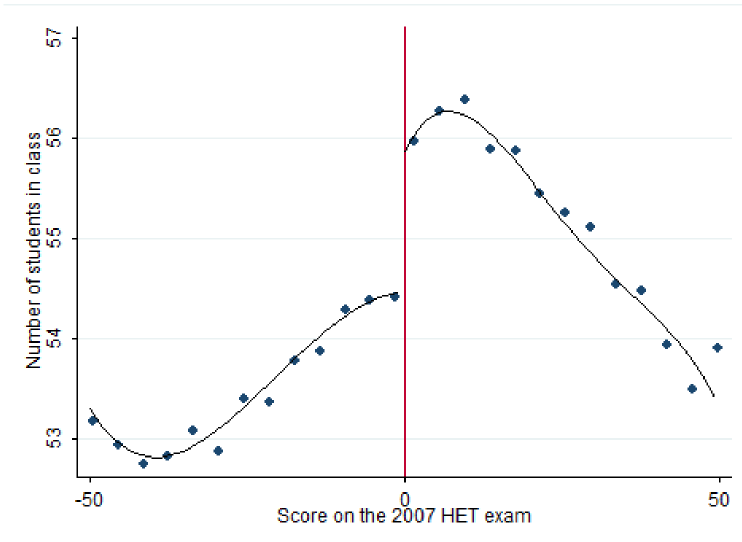
(a) High School Peer Quality



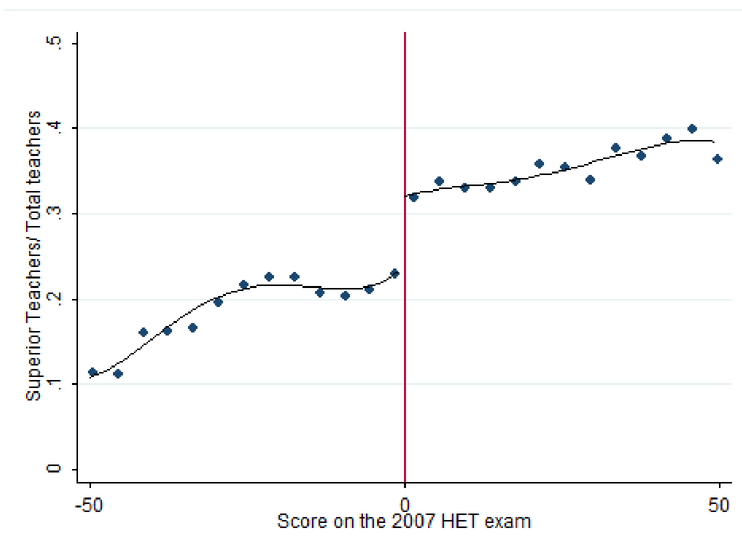
(b) Standardized CET scores

Notes: Sample includes students who took the HET exam in the year 2007 and the CET exam in the year 2010.
In order to exclude the Tier I admission threshold, we can use at most a bandwidth of 18 points on either side of the cutoff.

Figure 7: Discontinuities in class size and teacher quality across the Tier I admission threshold



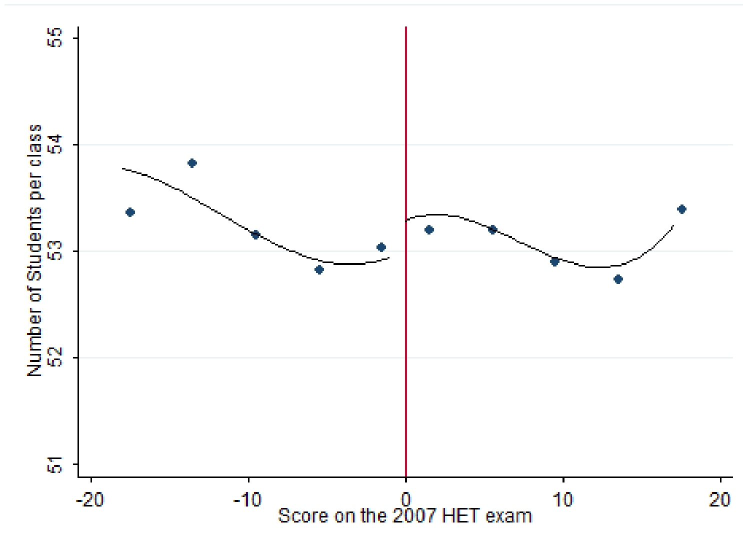
(a) Number of students per class



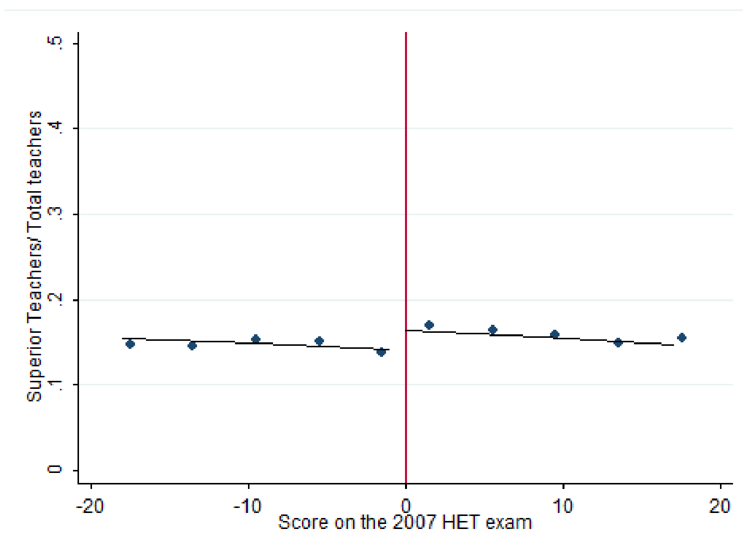
(b) Proportion superior teachers

Notes: Sample includes students who took the HET exam in the year 2007 and the CET exam in the year 2010 (School level data).

Figure 8: Discontinuities in class size and teacher quality across all admission thresholds, excluding the Tier I cutoff



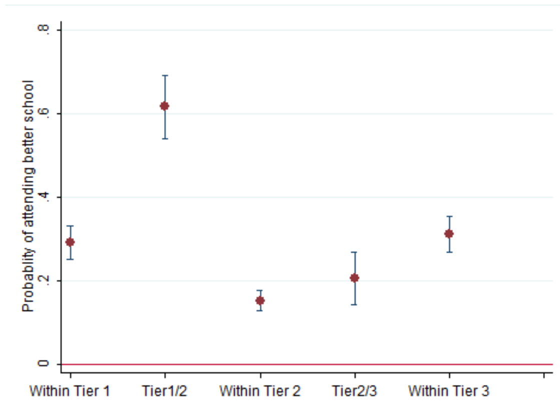
(a) Number of students per class



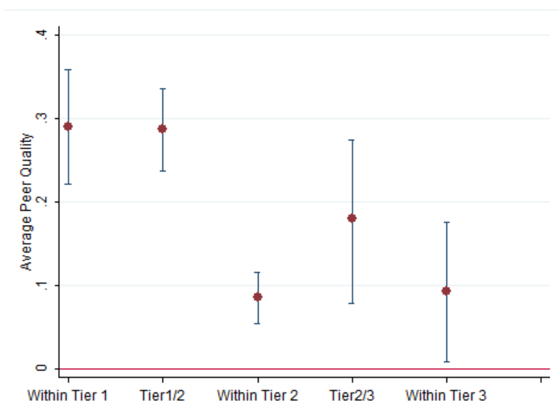
(b) Proportion superior teachers

Notes: Sample includes students who took the HET exam in the year 2007 and the CET exam in the year 2010 (School level data).
In order to exclude the Tier I admission threshold, we can use at most a maximum bandwidth of 18 points.

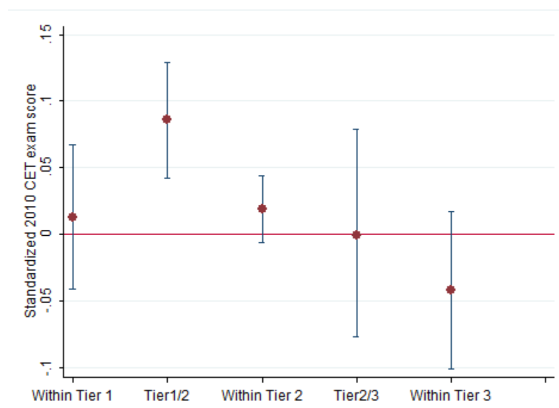
Figure 9: Regression discontinuity estimates across the selectivity distribution of schools



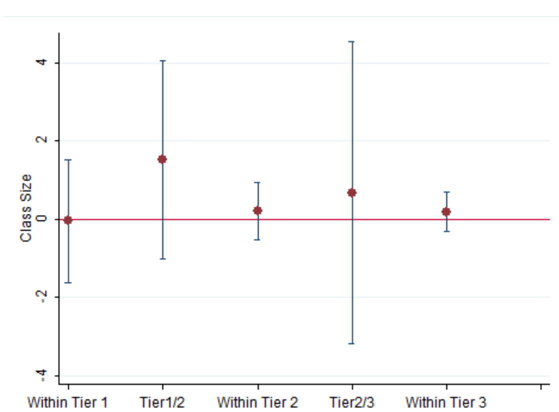
(a) Probability of attending school



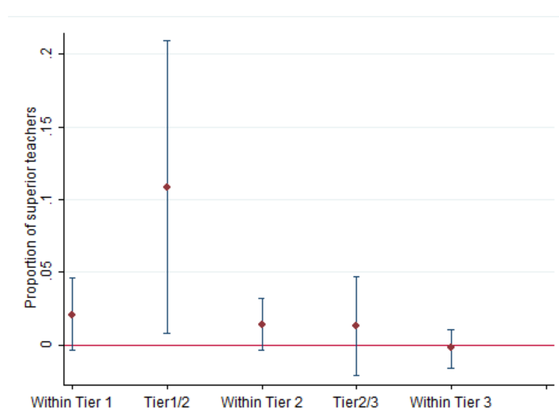
(b) Average Peer Quality



(c) Standardized 2010 CET exam scores



(d) Class size



(e) Proportion of superior teachers

Note: Sample includes students who took the HET exam in 2007 and the CET exam in the year 2010. Estimates and 95% confidence intervals reported in bars. All estimates are from RD regressions using a bandwidth of 50 HET points on either side of the cutoff.

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	669.02	602.07	537.47
	(59.52)	(31.91)	(39.62)	(43.17)
College entrance exam scores	487.57	567.69	464.40	386.14
	(99.69)	(60.57)	(83.37)	(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	145.97	123.24	99.14
	(42.26)	(47.71)	(34.70)	(27.90)
Number of students per class	53.62	55.02	53.24	51.54
	(5.14)	(2.55)	(2.26)	(11.62)
Ratio of superior teachers	0.22	0.38	0.16	0.07
	(0.16)	(0.08)	(0.13)	(0.05)
Number of schools	25	4	12	9
Number of Students	12,259	4,306	5,900	2,044

Notes: *1 Chinese Mu = 7176 sq feet.
Data are for students taking the high school entrance exam in 2007 and the college entrance exam in 2010. Standard errors (for non-binary variables) in parentheses.

Table 2: Regression discontinuity estimates across all admission thresholds

Bandwidth	2.5 CCT	2 CCT	1.5 CCT	1.25 CCT	CCT	0.75 CCT
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Probability of						
attending better school	0.329***	0.314***	0.313***	0.319***	0.336***	0.344***
	(0.007)	(0.007)	(0.008)	(0.009)	(0.011)	(0.013)
With Controls	0.331***	0.314***	0.312***	0.318***	0.337***	0.344***
	(0.007)	(0.007)	(0.008)	(0.009)	(0.011)	(0.012)
Observations	44181	35650	26890	22493	18001	13610
Panel B: Discontinuity in						
high school peer quality	0.108***	0.137***	0.149***	0.188***	0.187***	0.205***
	(0.014)	(0.017)	(0.021)	(0.024)	(0.027)	(0.031)
With Controls	0.106***	0.132***	0.148***	0.188***	0.187***	0.204***
	(0.013)	(0.017)	(0.020)	(0.023)	(0.026)	(0.030)
Observations	33,414	26,806	20,221	17,369	13,570	9,619
Panel C: Discontinuity in						
CET exam scores	–0.014	–0.016	–0.017	–0.009	–0.007	0.006
	(0.011)	(0.012)	(0.013)	(0.016)	(0.020)	(0.025)
With Controls	–0.012	–0.016	–0.019	–0.014	–0.014	–0.004
	(0.011)	(0.012)	(0.013)	(0.016)	(0.019)	(0.024)
Observations	66,530	53,930	41,599	34,334	27,777	21,135
Panel D: Discontinuity in						
likelihood of enrolling in						
4-year college	–0.002	–0.001	–0.003	–0.001	–0.001	0.009
	(0.005)	(0.005)	(0.006)	(0.007)	(0.009)	(0.012)
With Controls	–0.005	–0.004	–0.004	–0.001	–0.003	0.005
	(0.005)	(0.005)	(0.006)	(0.007)	(0.009)	(0.011)
Observations	70,493	57,330	42,524	37,029	29,682	22,065

Notes: Controls include age, gender, district fixed effects and middle school fixed effects. Optimal bandwidth selected using the CCT bandwidth selector for local linear estimation proposed in Calonico et al. (2015). Because optimal bandwidth differs across outcomes within a given column, the number of observations differs as well. Optimal bandwidth equals 20 for likelihood of attending a better school, 14 for peer quality, 29 for CET scores, and 31 for the likelihood of enrolling in four year colleges. Standard errors are clustered at the student ID level due to repeated observations.
*** p <0.01 ** p <0.05 * p <0.1

Table 3: Regression discontinuity estimates across the Tier I admission threshold

Bandwidth	2.5 CCT	2 CCT	1.5 CCT	1.25 CCT	CCT	0.75 CCT
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Discontinuity in probability of attending Tier I school	0.631***	0.632***	0.671***	0.649***	0.642***	0.628***
	(0.031)	(0.036)	(0.046)	(0.052)	(0.064)	(0.079)
With Controls	0.637***	0.637***	0.677***	0.656***	0.654***	0.637***
	(0.031)	(0.037)	(0.047)	(0.053)	(0.064)	(0.078)
Observations	7,167	6,046	4,654	3,901	3,133	2,389
Panel B: Discontinuity in high school peer quality	0.301***	0.314***	0.372***	0.357***	0.358***	0.345***
	(0.025)	(0.028)	(0.032)	(0.035)	(0.043)	(0.051)
With Controls	0.315***	0.329***	0.395***	0.378***	0.380***	0.359***
	(0.025)	(0.029)	(0.033)	(0.037)	(0.044)	(0.051)
Observations	6,680	5,578	4,278	3,642	2,886	2,237
Panel C: Discontinuity in CET exam scores	0.082***	0.089***	0.083***	0.085***	0.073***	0.069**
	(0.017)	(0.019)	(0.022)	(0.023)	(0.025)	(0.028)
With Controls	0.081***	0.086***	0.084***	0.090***	0.073***	0.065**
	(0.016)	(0.017)	(0.021)	(0.021)	(0.022)	(0.027)
Observations	9,909	8,870	7,352	6,454	5,298	4,112
Panel D: Discontinuity in likelihood of enrolling in 4-year college	0.135***	0.097***	0.063**	0.049	0.051	0.056
	(0.023)	(0.024)	(0.028)	(0.030)	(0.035)	(0.043)
With Controls	0.136***	0.101***	0.066**	0.048	0.048	0.053
	(0.023)	(0.024)	(0.028)	(0.030)	(0.034)	(0.041)
Observations	8,478	7,236	5,824	4,859	3,941	3,030

Notes: Controls include age, gender, district fixed effects and middle school fixed effects. Optimal bandwidth selected using the CCT bandwidth selector for local linear estimation proposed in Calonico et al. (2015). Because optimal bandwidth differs across outcomes within a given column, the number of observations differs as well. Optimal bandwidth equals 20 for the probability of attending a Tier I school, 18 for peer quality, 34 for CET scores, and 25 for the likelihood of enrolling in four year colleges. Standard errors are 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	0.632***	—-	0.609***	—-	0.652***	—-
	(0.036)		(0.045)		(0.033)	
Panel B: Discontinuity in school inputs						
High school peer quality	0.303***	0.486***	0.290***	0.484***	0.317***	0.489***
	(0.026)	(0.023)	(0.030)	(0.029)	(0.032)	(0.035)
Panel C: Discontinuity in outcomes						
College entrance exam test scores	0.094***	0.155***	0.174***	0.306***	0.015	0.020
	(0.024)	(0.040)	(0.034)	(0.061)	(0.027)	(0.042)
Eligibility to attend a 4-year college	0.071***	0.119***	0.103***	0.191***	–0.042	–0.060
	(0.027)	(0.043)	(0.036)	(0.059)	(0.042)	(0.063)
Observations	6046	6046	2813	2813	3233	3233

Notes: All regressions are estimated with controls, which include gender, age, district fixed effects and middle 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	2 CCT	1.5 CCT	1.25 CCT	CCT	0.75 CCT
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: (Going to a better school)						
Selecting into the CET entrance exam (All)	–0.011	–0.007	–0.001	0.004	–0.000	0.010
	(0.008)	(0.008)	(0.011)	(0.012)	(0.013)	(0.014)
Females only	–0.015	–0.017	–0.012	–0.008	–0.001	–0.003
	(0.009)	(0.011)	(0.014)	(0.016)	(0.017)	(0.020)
Males only	–0.004	0.004	0.010	0.017	0.001	0.025
	(0.010)	(0.012)	(0.016)	(0.017)	(0.019)	(0.021)
Observations (females)	18,226	14,615	10,944	9,076	7,461	5,552
Observations (males)	16,522	13,177	9,840	8,246	6,768	5,082
Panel B: (Going to a top school)						
Selecting into the CET entrance exam (All)	0.007	0.020**	0.020**	0.017	0.015	0.025*
	(0.008)	(0.009)	(0.010)	(0.011)	(0.012)	(0.014)
Females only	0.017	0.037***	0.021	0.019	0.011	0.025
	(0.013)	(0.014)	(0.015)	(0.017)	(0.019)	(0.022)
Males only	–0.005	0.001	0.019	0.015	0.019	0.023
	(0.015)	(0.017)	(0.019)	(0.021)	(0.024)	(0.027)
Observations (females)	5,386	4,663	3,791	3,214	2,607	1,983
Observations (males)	4,800	4,166	3,375	2,862	2,363	1,822

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 for local linear estimation proposed in Calonico et al. (2015). Because optimal bandwidth differs across outcomes within a given column, the number of observations differs as well. Optimal bandwidth equals 12 for the likelihood of taking the CET exam across all cutoffs, and 29 for the likelihood of taking the CET exam across the Tier I cutoff.
*** p <0.01 ** p <0.05 * p <0.1

Table 6: The distributional effects of Tier I schools on CET Scores

Outcome	College eligibility threshold	Top 25%	Top 50%	Top 75%
Panel A: (Original Sample)				
Discontinuity	.072***	.012	.086***	.032***
	(.027)	(.022)	(.022)	(.010)
Discontinuity (With controls)	.074***	.014	.090***	.033***
	(.027)	(.022)	(.020)	(.009)
Observations	6,107	6,107	6,107	6,107
Panel B: (Includes non-test-takers)				
Discontinuity	.077***	.014	.092***	.048***
	(.027)	(.021)	(.023)	(.012)
Discontinuity (With controls)	.077***	.013	.095***	.049***
	(.027)	(.021)	(.021)	(.011)
Observations	6,653	6,653	6,653	6,653

Notes: The outcome variables are defined as indicators for exceeding significant CET exam thresholds. Controls include sex, gender, district fixed effects, and middle school fixed effects. All local linear regressions use a bandwidth of 40 points on either side of the Tier I cutoff. College eligibility requires a student be in the 48th and 69th percentile of the CET distribution for science and arts students respectively. Panel B includes students who did not sit for the CET exam, each of whom is assigned a zero for each outcome.
*** p <0.01 ** p <0.05 * p <0.1

Table 7: OLS regression of school-level CET RD estimates on school-level RD estimates of changes in inputs

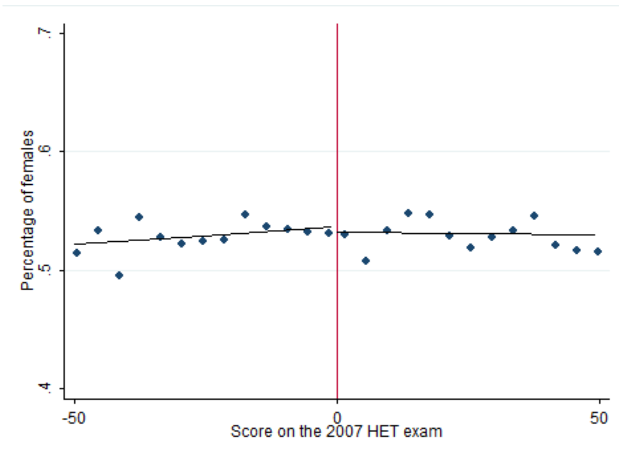
	CET test scores (1)	CET test scores (2)
Peer Quality	0.046	-0.140
	(0.161)	0.174
	[0.95]	[0.435]
Teacher Quality	0.489*	0.677*
	(0.280)	(0.370)
	[0.33]	[0.093]
Class size	0.008	0.001
	(0.010)	(0.012)
	[0.35]	[0.914]
Number of Schools	23	31
Number of suburban school districts	Two (Original sample)	Four

Notes: Schools were excluded if their admission thresholds were within 5 HET points of another cutoff or if the RD sample contained fewer than 50 students. We use a bandwidth of 15 HET points to estimate discontinuities. Standard errors are clustered at the school level and are reported in parentheses. We also report p-values from a wild-t cluster bootstrap procedure in brackets.
*** p <0.01 ** p <0.05 * p <0.1

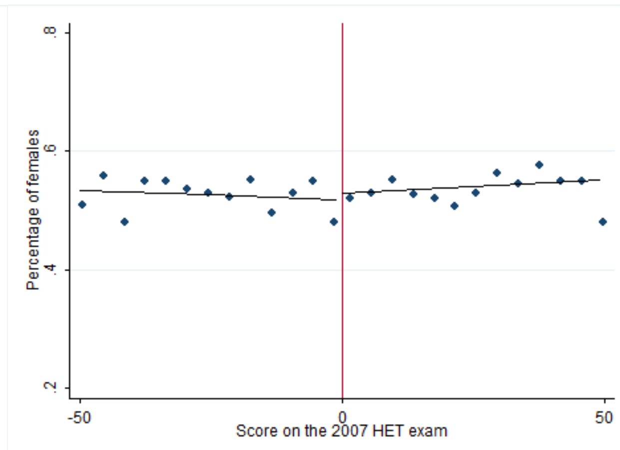
C Appendix Figures (for online publication only)

Figure A1: Smoothness of baseline covariates

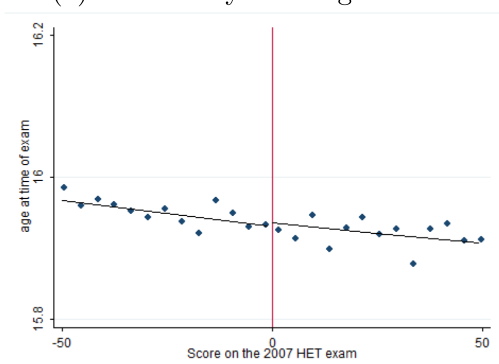
Attending Better Schools (All Cutoffs) | Attending a Tier I school



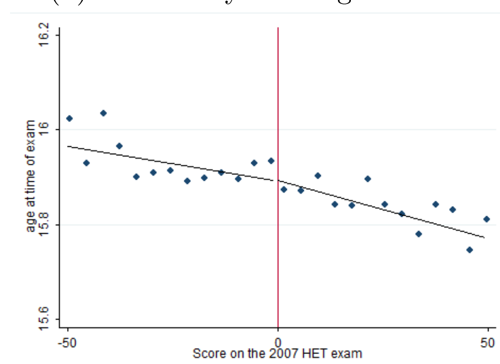
(a) Probability of being female



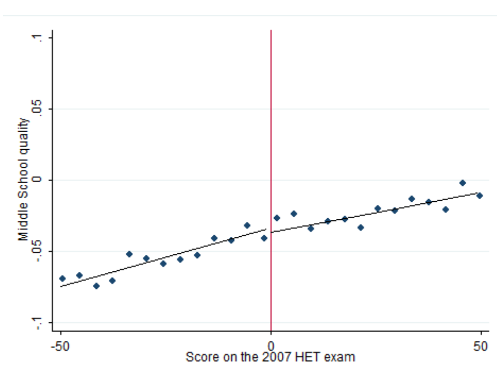
(b) Probability of being female



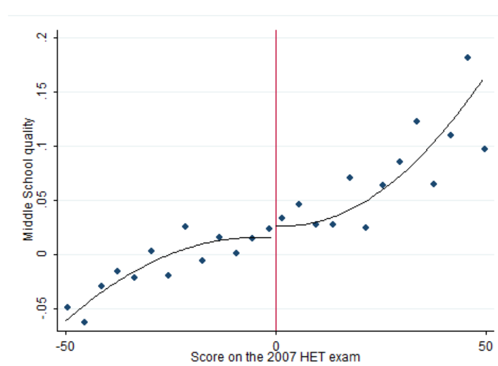
(c) Age at time of high school exam



(d) Age at time of high school exam



(e) Middle School Quality

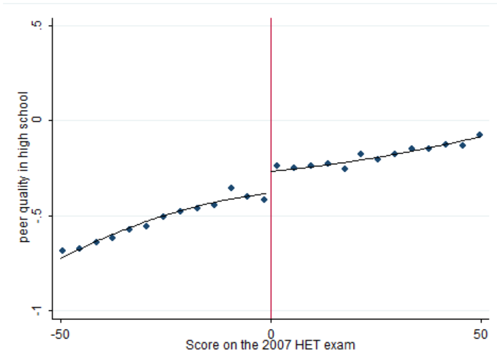


(f) Middle School Quality

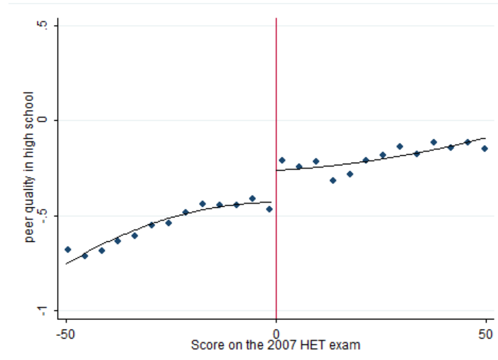
Notes: Sample includes students who took the high school entrance exam in the year 2007 and college entrance exam in 2010.

Figure A2: Discontinuities across all admission thresholds, by gender

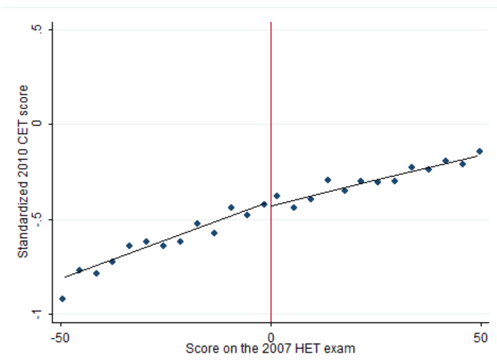
Males	Females



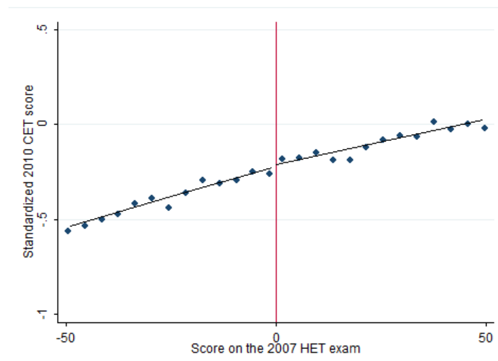
(a) Peer quality in high school



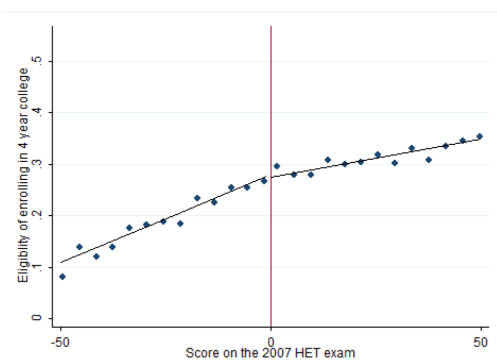
(b) Peer quality in high school



(c) CET exam scores



(d) CET exam scores



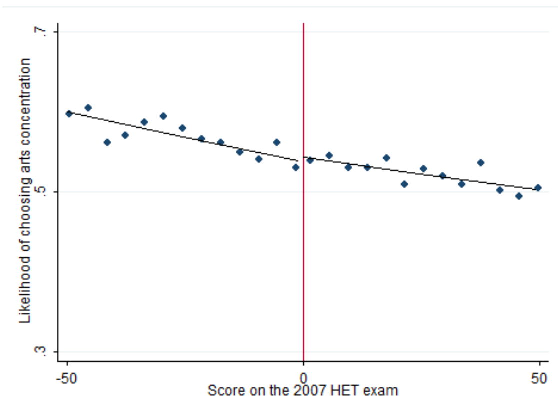
(e) Likelihood of enrolling in 4-year college

(f) Likelihood of enrolling in 4-year college

Notes: Sample includes students who took the HET exam in the year 2007 and the CET exam in the year 2010.

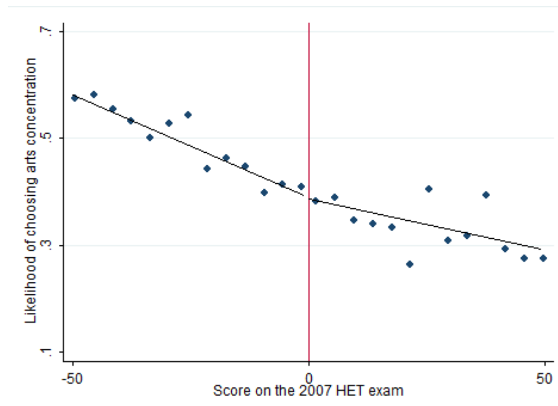
Figure A3: Threats to interpretation

Attending Better Schools

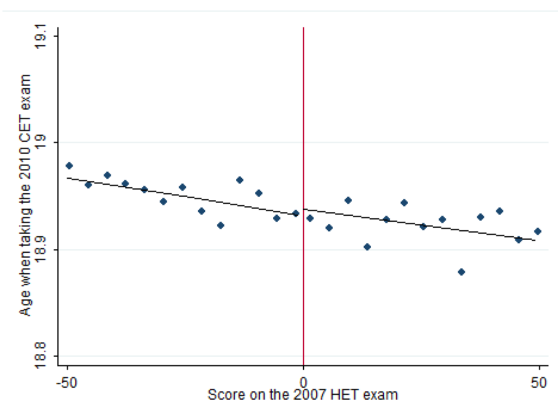


(a) Likelihood of majoring in arts versus sciences in High school.

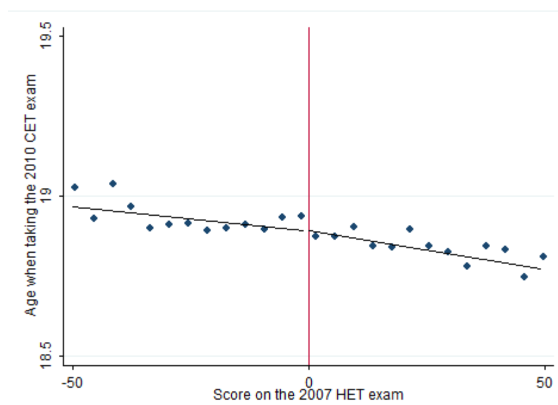
Attending a Tier I school



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



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

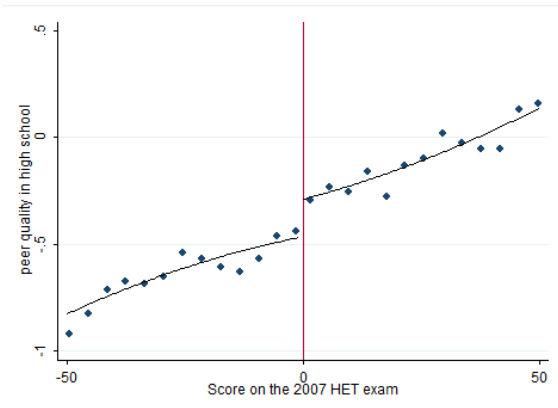


(d) Exact age when taking the 2010 CET exam.

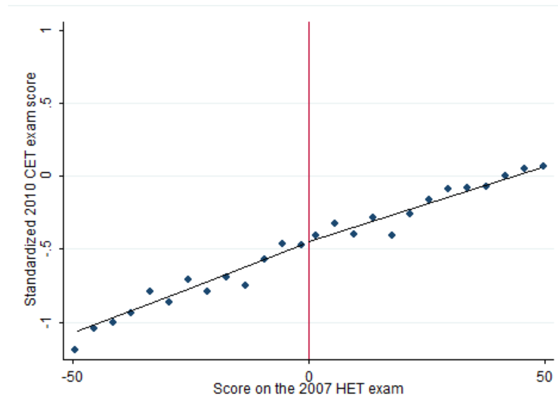
Notes: Sample includes students who took the HET exam in the year 2007 and the CET exam in the year 2010.

Figure A4: Discontinuities in peer quality and CET scores across other admission thresholds

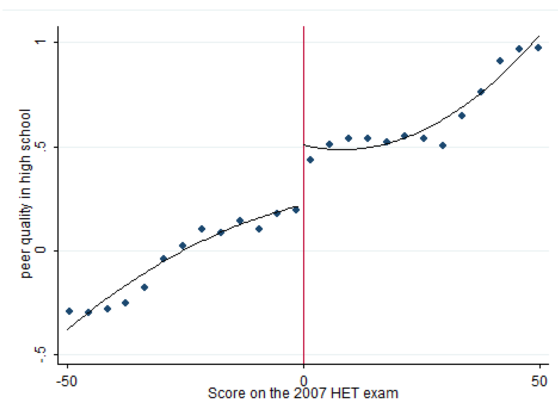
High School Peer Quality | Standardized CET scores



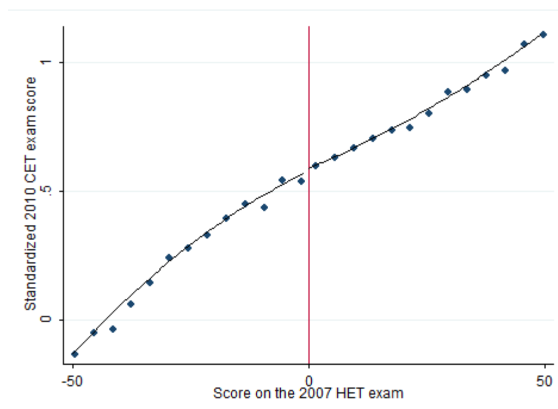
(a) Tier II school cutoff



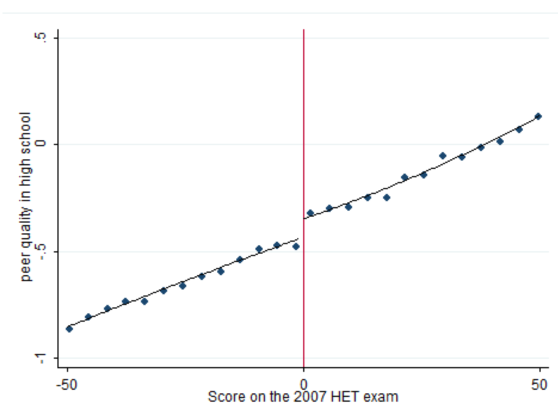
(b) Tier II school cutoff



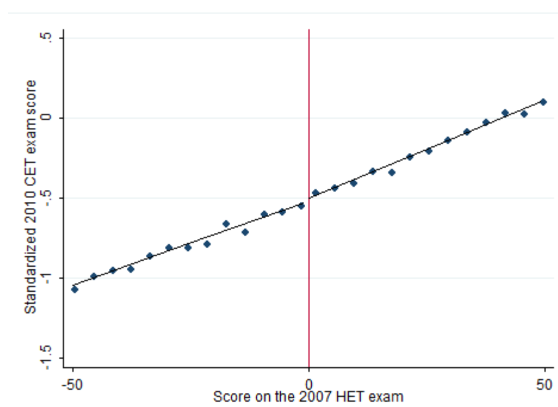
(c) Within Tier I school cutoffs



(d) Within Tier I school cutoffs



(e) Within Tier II school cutoffs

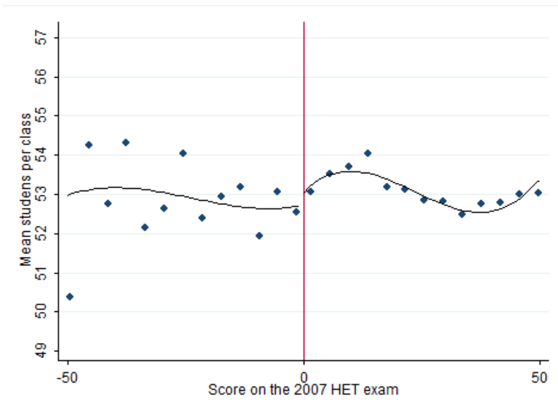


(f) Within Tier II school cutoffs

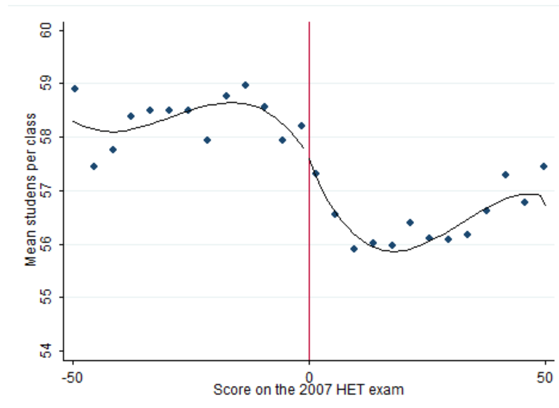
Notes: Sample includes students who took the HET exam in the year 2007 and the CET exam in the year 2010.

Figure A5: Discontinuities in class size and teacher quality across other admission thresholds

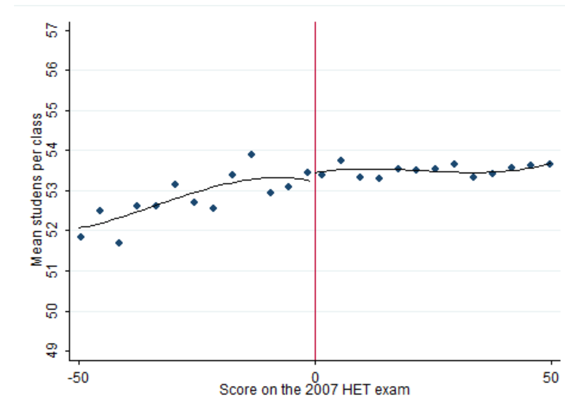
Tier II cutoff | Within Tier I cutoffs | Within Tier II cutoffs



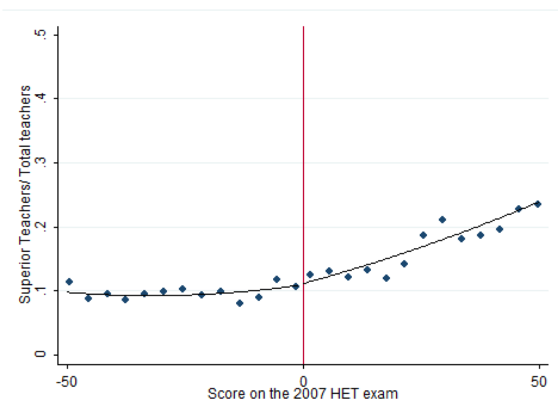
(a) Number of students per class



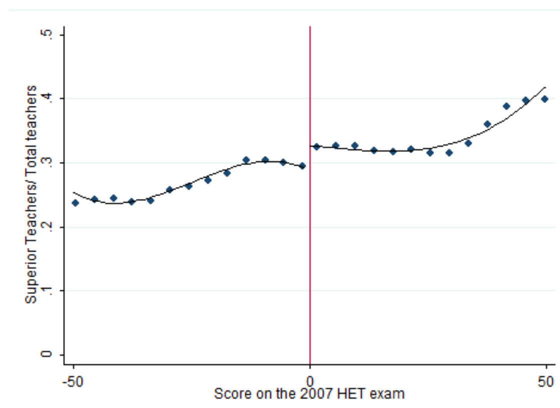
(b) Number of students per class



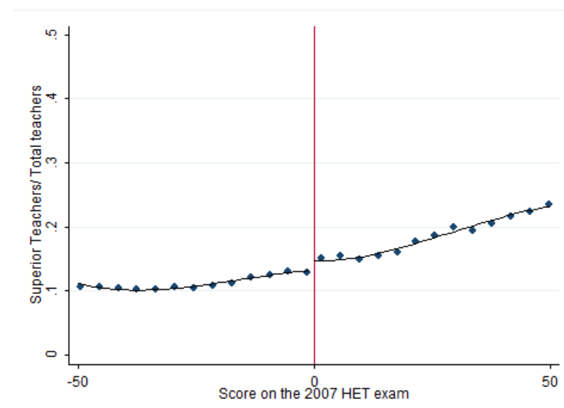
(c) Number of students per class



(d) Proportion superior teachers



(e) Proportion superior teachers



(f) Proportion superior teachers

Notes: Sample includes students who took the HET exam in the year 2007 and the CET exam in the year 2010 (School level data).

D Appendix Tables (for online publication only)

Table A1: Regression discontinuity estimates for baseline covariates

Bandwidth	2.5 CCT	2 CCT	1.5 CCT	1.25 CCT	1 CCT	0.75 CCT
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: (All admission thresholds)						
Predicted CET score	–0.001	0.001	–0.001	0.001	0.003	0.007
	(0.004)	(0.005)	(0.006)	(0.006)	(0.007)	(0.008)
Likelihood of being a female	–0.004	–0.006	–0.007	–0.007	–0.008	–0.001
	(0.006)	(0.007)	(0.008)	(0.008)	(0.009)	(0.010)
Age when taking HET entrance exam	0.003	0.005	0.002	0.006	0.006	–0.004
	(0.008)	(0.009)	(0.010)	(0.011)	(0.013)	(0.014)
Middle School Quality	–0.011	–0.014	0.003	0.003	0.007	0.007
	(0.013)	(0.014)	(0.015)	(0.017)	(0.018)	(0.019)
Observations for predicted score	75,977	62,474	47,880	40,676	32,499	24,931
Panel B: (Tier I admission threshold)						
Predicted CET score	–0.007	–0.005	0.014	0.027	0.017	0.022
	(0.013)	(0.014)	(0.014)	(0.016)	(0.017)	(0.021)
Likelihood of being a female	0.011	0.017	0.019	0.029	0.049**	0.049*
	(0.019)	(0.020)	(0.021)	(0.022)	(0.023)	(0.027)
Age when taking HET entrance exam	–0.011	–0.012	–0.003	–0.006	–0.024	–0.051
	(0.021)	(0.022)	(0.027)	(0.029)	(0.032)	(0.036)
Middle School Quality	0.001	–0.002	0.003	0.003	0.007	0.011
	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)
Observations for predicted score	6,680	5,578	4,278	3,642	2,886	2,237

Notes: Predicted HET score is based on the following controls: age, gender, district fixed effects, and middle school fixed effects. For panel A, optimal bandwidth equals 34 for predicted score, 34 for probability of being a female, 40 for age when taking HET exam, 24 for middle school quality (Going to a better school). For panel B, optimal bandwidth equals 18 for predicted score, 26 for probability of being a female, 34 for age when taking HET exam, and 17 for middle school quality.
*** p <0.01 ** p <0.05 * p <0.1

Table A2: Regression discontinuity estimates for all admission thresholds, excluding the Tier I cutoff

Bandwidth	18 points	16 points	14 points
	(1)	(2)	(3)
Panel A:			
Discontinuity in high school peer quality	0.163***	0.158***	0.180***
	(0.026)	(0.028)	(0.030)
With Controls	0.166***	0.161***	0.183***
	(0.02)	(0.03)	(0.03)
Panel B:			
Discontinuity in CET exam scores	0.020	–0.002	0.014
	(0.032)	(0.035)	(0.037)
With Controls	0.028	0.007	0.018
	(0.03)	(0.03)	(0.04)
Observations	14,624	13,055	11,432

Notes: Controls include age, gender, district fixed effects, and middle school fixed effects. In order to exclude the Tier I admission threshold, we can use at most a bandwidth of 18 points on either side of the cutoff for local linear 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 A3: Robustness to adding third and fourth suburban districts to the sample

Treatment effect	Original Sample	Add districts 3 and 4 (assuming district 4 has five Tier I schools)	Add districts 3 and 4 (assuming district 4 has one Tier I school)
Panel A: All admission thresholds			
High school peer quality	0.170***	0.191***	0.191***
	(0.021)	(0.017)	(0.017)
College exam scores	-0.016	-0.016	-0.016
	(0.014)	(0.013)	(0.013)
Likelihood of enrolling in four year college	–0.002	–0.001	–0.001
	(0.006)	(0.007)	(0.007)
Ratio of superior teachers	0.014	0.016*	0.016*
	(0.010)	(0.008)	(0.008)
Number of students per class	0.083	–0.038	–0.038
	(0.186)	(0.198)	(0.198)
Observations	37,961	57,828	57,828
Panel B: Tier I admission threshold			
First Stage	0.632***	0.477***	0.661***
	(0.036)	(0.025)	(0.035)
High school peer quality	0.303***	0.210***	0.350***
	(0.026)	(0.024)	(0.028)
College exam scores	0.094***	0.083***	0.078***
	(0.024)	(0.022)	(0.024)
Likelihood of enrolling in four year college	0.071***	0.075***	0.054***
	(0.027)	(0.026)	(0.021)
Ratio of superior teachers	0.111**	0.094**	0.109***
	(0.050)	(0.050)	(0.034)
Number of students per class	1.615	2.072	1.215
	(1.280)	(1.394)	(1.738)
Observations	6,046	8,056	8,056

Notes: District 3 has no Tier I schools. All specifications control for gender, age, district fixed effects and junior high school fixed effects. For ease of comparison, all local linear regressions use a bandwidth of 40 points on either side of the cutoff.
*** p <0.01 ** p <0.05 * p <0.1

Table A4: Bounding analysis for the estimated impact of attending Tier I schools

Bandwidth	2.5 CCT	2 CCT	1.5 CCT	1.25 CCT	CCT	0.75 CCT
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Selection into the CET exam	0.003	0.009	0.020**	0.017*	0.013	0.017
	(0.009)	(0.009)	(0.009)	(0.010)	(0.011)	(0.012)
Constant (Control mean)	0.926	0.916	0.903	0.904	0.908	0.902
Proportion to be trimmed	00.32%	00.98%	20.2%	10.88%	10.43%	10.88%
Panel B: College entrance exam test scores (Original regression estimates)	0.082***	0.089***	0.083***	0.085***	0.073***	0.069**
	(0.017)	(0.019)	(0.022)	(0.023)	(0.025)	(0.028)
College entrance exam test scores (Lower bound estimates)	0.078***	0.077***	0.060**	0.063**	0.054**	0.048*
	(0.020)	(0.020)	(0.025)	(0.026)	(0.027)	(0.028)
College entrance exam test scores (Upper bound estimates)	0.084***	0.110***	0.124***	0.120***	0.101***	0.096***
	(0.018)	(0.022)	(0.024)	(0.025)	(0.029)	(0.032)
Observations	9,909	8,870	7,352	6,454	5,298	4,112
Observations after trimming	9,851	8,826	7,268	6,391	5,258	4,068

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 for local linear estimation 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