The Long-Run Educational Benefits of High-Achieving Classrooms*

Serena Canaan†
Simon Fraser University

Pierre Mouganie‡
Simon Fraser University

Peng Zhang§
The Chinese University of Hong Kong, Shenzhen

February 6, 2022

Abstract

Despite the prevalence of school tracking, evidence on whether it improves student success is mixed. This paper studies how tracking within high school impacts high-achieving students' short- and longer-term academic outcomes. Our setting is a large and selective Chinese high school, where first-year students are separated into high-achieving and regular classrooms based on their performance on a standardized exam. Classrooms differ in terms of peer ability, teacher quality, class size, as well as level and pace of instruction. Using newly collected administrative data and a regression discontinuity design, we show that high-achieving classrooms improve math test scores by 23 percent of a standard deviation, with effects persisting throughout the three years of high school. Effects on performance in Chinese and English language subjects are more muted. Importantly, we find that high-achieving classrooms raise enrollment in elite universities by 17 percentage points, as they substantially increase scores on the national college entrance exam—the sole determinant of university admission in China.

JEL Classification: I21, I24, I26, J24
Keywords: Classroom Tracking, Peer Quality, Teacher Quality, Regression Discontinuity, China

---

*We thank seminar participants at Shenzhen University and Jinan University. We are also grateful to Joshua Goodman and Mark Hoekstra for helpful comments and suggestions. All errors are our own.

†Department of Economics, Simon Fraser University, and IZA, e-mail: scanaan@sfu.ca

‡Department of Economics, Simon Fraser University, and IZA, e-mail: pierre_mouganie@sfu.ca

§School of Management and Economics, The Chinese University of Hong Kong, Shenzhen, e-mail: zhangpeng@cuhk.edu.cn

1 Introduction

Tracking, the practice of grouping students into classes based on prior achievement, is common in many countries such as the United States, Canada and China. While within-school tracking is widespread, it remains exceedingly controversial.[1] Tracking allows teachers and schools to tailor their instruction and resources to students' abilities and needs, which may boost their educational attainment. On the other hand, it may widen educational gaps between high- and low-achieving students, by putting the latter at a learning disadvantage (Betts, 2011). In the public policy arena, this issue is hotly debated as an increasing number of policymakers are questioning the benefits of placing high-achieving students in separate classes. As a result, many school districts are now moving towards eliminating tracking. For example, in October 2021, Mayor de Blasio announced a highly controversial plan to phase out New York City's Gifted and Talented program by 2022 (The New York Times, 2021).[2]

Despite considerable policy relevance, evidence on whether tracking improves high-achieving students' academic performance is mixed, and less is known about how it affects their long-term outcomes. Additionally, very few studies look at the impact of separating students in achievement-based classrooms at the high school level. Understanding the implications of tracking at the high school level is important for two reasons. First, tracking within high schools is widespread. For example, in 2013, 75% of U.S. high school districts tracked students into different classes through offering honors classes, advanced placement (AP) courses and other gifted programs (National Research Center on the Gifted and Talented, 2013). In China, virtually all high schools separate students into classrooms based on their prior academic performance. Second, while previous studies have focused on tracking at the primary or middle school level, it is unclear whether their results can be extended to older students. Indeed, cross-country evidence suggests that the age at which students are tracked is a strong determinant of their future outcomes, and that tracking at a later age may be more beneficial for students' educational trajectories (Hanushek and Wößmann, 2006; Brunello and Checchi, 2007; Schütz, Ursprung and Wößmann, 2008).

This study is the first to examine whether tracking in high school benefits high-achieving students in terms of academic achievement, as well as college attendance and college selectivity. Our context is China where virtually all high schools have adopted classroom tracking. Indeed, results from a survey we conducted among Chinese university students indicates

[1]The definition of tracking can vary substantially across educational systems. Based on their prior achievement, students may be tracked into (i) different schools, (ii) different classrooms within the same schools or, (iii) vocational and general education. We use the terms tracking or within-school tracking to refer to the practice of separating students into achievement-based classrooms within the same school.

[2]NYC's Gifted and Talented program places students identified as "gifted", based on their performance on a standardized test, in separate classrooms from their peers.

that 93.3% of them attended a high school that divides students into achievement-based classrooms. We investigate the effects of this in one setting: Qingyang First High School, a large and selective high school located in the low-income province of Gansu, which is ideally suited for answering our question. At the beginning of their first year at Qingyang First High School, students have to take a classroom placement exam (CPE). The top performers on this exam are assigned to high-achieving classrooms, while other students are randomly allocated to regular classrooms. The majority of students stay in the same classrooms for all three years of high school. The classroom allocation mechanism used by the high school creates a CPE score cutoff, whereby students scoring above the cutoff are assigned to high-achieving classrooms and those scoring below are placed in regular classrooms. We can therefore estimate the causal effect of being assigned to a high-achieving classroom, by using a regression discontinuity design which compares students who score barely above to those who score barely below the CPE cutoff. We collected rich administrative data on all students who enroll in their first year at this large and selective Chinese high school from the years 2015 to 2017. An advantage of our data is that we can track students' educational outcomes both in the short- and longer-run, as we have information on their performance on common exams taken throughout the three years of high school, scores on the high-stakes national college entrance exam, and the name of the university they enroll in. Our data also allow us to test for differences in classroom educational inputs and hence, provide evidence on the mechanisms driving our effects.

Our results indicate that students substantially benefit from being assigned to high-achieving classrooms in their first year of high school. Eligibility to enroll in a high-achieving classroom leads to a 23 percent of a standard deviation increase in performance on mathematics exams taken during the first year of high school. These benefits persist in the second and third years of high school, as we document significant and comparable improvements in math performance. On the other hand, the impacts of high-achieving classroom assignment on students' performance in Chinese and English language subjects are more muted. To investigate longer-term effects, we look at students' performance on the high-stakes college entrance exam and the type of universities that they enroll in. Indeed, at the end of their last year of high school, students in China take a common national exam which is the sole determinant of admissions into 4-year universities. We show that high-achieving classroom eligibility increases students' performance on the college entrance exam by around 0.28 standard deviations. Importantly, we find that students are 17 percentage points more likely to enroll in highly-selective elite colleges. Since these elite colleges have been previously shown to yield large earnings gains (Jia and Li, 2021), our findings suggest that the benefits of high-achieving classrooms likely persist in the labor market.

An advantage of our data is that it allows us to quantify the mechanisms underlying our effects. We show that students assigned to high-achieving classrooms are naturally exposed to higher-achieving peers compared to those placed in regular classrooms. We further find that students in high-achieving classrooms benefit from having higher quality teachers—measured by teachers' official rank, pay scale and work experience—and smaller class sizes. Finally, our discussions with school administrators indicate that while high-achieving and regular classrooms follow a similar curriculum, instructors in high-achieving classrooms teach at a faster pace and delve deeper into the material.

Our results contribute to a large body of work which examines whether tracking improves student achievement. Early U.S. studies compare schools which track students into achievement-based classrooms to schools that do not track, and find limited evidence that tracking improves academic outcomes (Betts and Shkolnik, 2000; Figlio and Page, 2002; Zimmer, 2003; Lefgren, 2004). Our paper is more directly related to studies which use a regression discontinuity design to look at whether students benefit from being placed in high-achieving versus regular classrooms. Evidence from this literature is quite mixed. Duflo, Dupas and Kremer (2011) find no significant differences in test scores from being placed into high- versus low-achievement grade 1 classrooms in Kenya. Bui, Craig and Imberman (2014) also show that admission to a Gifted and Talented program in U.S. middle schools does not impact student achievement. Tangvatcharapong (2020) finds similar effects in middle schools in Thailand. On the other hand, Card and Giuliano (2016) document that 4th grade gifted classrooms in the U.S. substantially improve high-achieving minority students' math and reading test scores.

Our paper adds to this literature in two ways. First, we present the first causal evidence on the effects of assigning *high school* students to high-achieving classrooms. Prior work focuses instead on tracking at the elementary and middle school level. The scarcity of evidence on within-high school tracking is striking given that it is a common practice in many countries. For example, in the U.S., high schools routinely track students based on their achievements through offering advanced placement (AP) courses, honors classes and other types of gifted programs (Callahan et al., 2017). Nonetheless, evidence on these specific programs is scant. Some studies show that providing students and teachers with cash incentives to pass high school AP exams (Jackson 2010a; 2014) increase college enrollment and graduation, but they do not look at whether taking AP courses affects performance or college outcomes.[3]

[3] Two additional studies look at high school tracking programs in substantially different settings. Welsch and Zimmer (2018) use a sibling fixed effects model to show that participation in U.S. high schfool gifted programs has no significant effect on later-life outcomes. An advantage of our setting is that we can use a regression discontinuity design which rests on a minimal set of assumptions and allows us to better account

A second advantage of our setting is that we can document the *longer-term* educational impacts of high-achieving classrooms. Most previous studies look at academic performance for up to at most two years after students are tracked (Duflo, Dupas and Kremer, 2011; Bui, Craig and Imberman, 2014; Card and Giuliano, 2016). An exception is the study by Cohodes (2020) who evaluates Boston Public Schools' Advanced Work Class (AWC). The program groups high-achieving 4th to 6th grade students in the same classroom and offers advanced literacy curricula as well as accelerated math in later grades. While AWC had positive but insignificant impacts on short-term test scores, it increased high school graduation and college enrollment. Our paper complements this study as we show that assigning high school—instead of elementary and middle school—students to high-achieving classrooms substantially improves their short- and longer-term test scores, as well enrollment in elite colleges.

Additionally, our results are the first to show that within-school tracking is an important determinant of high-achieving students' access to elite universities. Our paper thus relates to recent studies which highlight that many high-achieving low-income students do not enroll in the best colleges available to them (Hoxby and Avery, 2012; Dillon and Smith, 2017), and hence miss out on substantial earnings returns to high-quality colleges (Hoekstra, 2009; Canaan and Mouganie, 2018; Black, Denning and Rothstein, 2020). However, little is known about what determines these students' access to selective colleges. Informational interventions, counseling, financial aid and family networks have all been shown to impact college quality (Hoxby and Turner, 2013; Cohodes and Goodman, 2014; Pallais, 2015; Castleman and Goodman, 2018; Altmejd et al., 2021; Dynarski et al., 2021). Our findings complement these studies, as they indicate that providing top-performing students residing in lower income areas with opportunities to enroll in high-achieving classrooms may be an effective way to boost their enrollment in selective colleges.

Finally, our results are consistent with recent studies showing that students realize substantial achievement gains from accessing selective high schools (Berkowitz and Hoekstra, 2011; Clark and Del Bono, 2016; Jackson, 2010b; 2013; Pop-Eleches and Urquiola, 2013; Dee and Lan, 2015; Beuermann and Jackson, 2018; Hoekstra, Mouganie and Wang, 2018) and U.S. charter schools (Angrist et al., 2010; Abdulkadiroğlu et al., 2011; Angrist, Pathak and Walters, 2013; Dobbie and Fryer, 2015; Cohodes, Setren and Walters, 2021). Similar to high-achieving classrooms, these schools typically provide students with a bundle of im-

for endogeneity of placement into achievement-based classrooms. Vardardottir (2013) estimates the impact of being placed in high-ability classrooms in Icelandic high schools on short-term test scores. The author emphasizes that the only difference between high and low-ability classrooms in Iceland is peer ability. In contrast, as in most tracking systems, classrooms in our setting differ in terms of peer ability, teachers, pedagogy, and class size.

proved educational inputs such as higher peer ability, better teacher quality and tailored pedagogy. Our findings highlight that variation in inputs *within* and not just *across* schools can drive differences in long-term academic success.

The rest of this paper is organized as follows. Section 2 describes the educational system and within-school tracking in China. Section 3 details the data we use. Section 4 outlines the identification strategy. Section 5 presents the main results and robustness checks. In section 6, we discuss the possible mechanisms behind our findings and we conclude in section 7.

2 Institutional Setting

2.1 Overview of the Education System in China

Students in China enroll in elementary school at the age of six or seven. They spend six years in elementary school, followed by another three years in middle school. Elementary and middle schools provide compulsory general education that is common to all students. At the end of middle school, students can pursue either vocational or general secondary education. The general education path allows students to eventually enroll in academically-focused universities, while the vocational path prepares them for specific occupations and restricts access to traditional higher education.

After middle school, students in the general education path pursue three years of high school (grades 10 to 12). High school admission is typically based on students' performance on a city-level high school entrance exam or Zhongkao, taken at the end of middle school. Admission to selective high schools in China is highly-competitive. Students submit a form indicating their ordered preference of high schools. They are then assigned to different high schools using an algorithm that takes into account students' preferences and high school entrance exam scores. In their first year of high school, all students pursue a common curriculum. At the end of their first year, they choose between two academic concentrations: Arts or Sciences. This choice is consequential for their postsecondary studies, as some majors only admit students from one of the concentrations. Students decide on their concentration based on their personal preferences and abilities. However, high-achieving students typically enroll in the Sciences concentration, as it allows them to access a wider set of college majors.

Students in China are granted admission into different 4-year colleges through a centralized admissions process. At the end of the three years of high school, all students wishing to attend 4-year colleges, are required to take a common college entrance exam or Gaokao.[4]

[4]The college entrance exam is graded out of a possible 750 points and is common for all students in

Similar to high schools, college admissions are almost entirely based on students' performance on this exam. The Chinese central government officially divides universities into tiers based on their quality and selectivity, with Tier I universities being the most selective. After the college entrance exam is graded, provinces set and announce minimum admission score cutoffs for each university tier.[5] Students submit a list of preferred colleges and majors after receiving their college entrance exam scores and seeing the minimum admissions cutoffs. Tier I universities then start admitting students based on their listed preferences and college entrance exam scores, followed by Tier 2 universities. Students whose college entrance exam scores exceed the provincial admission cutoffs are not guaranteed a spot at their preferred college. This is because each university can set its own admissions cutoff as long as it exceeds the minimum cutoff set by the province for its corresponding tier.

Among Tier I universities, there is also a great deal of variability in their degree of selectivity. The most selective and prestigious universities are part of two national projects which aim to transform them into world-leading institutions. Specifically, Project 211 (or "Top 100 in the 21st century" Project) and Project 985 (or "World First Class University" Project), which were launched by the Chinese Ministry of Education in 1995 and 1999 respectively, allocate extra funds to top universities in an effort to improve their research standards.[6] Around 112 Tier I institutions are listed as part of Project 211, and 39 of these also constitute Project 985 universities.[7] Project 211 and Project 985 institutions are considered to be the top 100 and top 40 universities in China, respectively. Only around 5% of college students are enrolled in Project 211 universities every year. These universities are not just highly-selective but they also lead to substantial gains in the labor market. Indeed, Jia and Li (2021) show that enrolling in Project 211 universities increases students' average monthly wages from their first job by 28 to 45%. In section 5.3, we estimate the impact of high-achieving classrooms on students' performance on the high-stakes college entrance

the same province, year and academic concentration. Specifically, students with an Arts concentration take tests in English language, Chinese language, Mathematics for Arts concentration and a comprehensive test consisting of Politics, History, and Geology. Science concentration students take tests in English language, Chinese language, Mathematics for Science concentration and a comprehensive test that includes Physics, Chemistry, and Biology.

[5]The cutoffs are set after taking into account the distribution of college entrance exam scores and the Ministry of Education's quotas for the province.

[6]Between 1996 and 2000, the government allocated around $2 billion dollars to universities on the Project 211 list.

[7]These projects have been successful in achieving their goals. For example, Project 211 universities take on the responsibility of training four-fifths of doctoral students, two-thirds of graduate students, half of students from abroad and one-third of undergraduates in China. They hold 96% of key laboratories in China, and consume 70% of scientific research funding. Additionally, most of these universities are ranked among the top 1000 worldwide universities according to the Academic Ranking of World Universities and the Times Higher Education World University Rankings.

exam, as well as their likelihood of enrolling in (i) a Tier I university, (ii) a Project 211 university (henceforth, top 100 university) and, (iii) a Project 985 university (henceforth, top 40 university). Focusing on these consequential outcomes allows us to gauge the longer-term educational benefits of high-achieving classrooms.

2.2 Within-High School Tracking

The practice of separating high school students into achievement-based classrooms is prevalent in China. While there is no official data on the proportion of high schools that track students into high-achieving classrooms in China, virtually all high schools separate students into classrooms based on their prior academic performance in practice. To corroborate this, we conducted a large-scale online survey targeted at students attending 79 different universities throughout China. We received responses from 701 students spanning 30 provinces (all provinces except Tibet) who attended 520 different high schools. 654 respondents (93.3%) indicated that their high schools placed top-performing students into separate classrooms.

Our student level administrative data are collected directly from Qingyang First High School, the most selective high school in the city of Qingyang. Qingyang is a prefecture-level city located in the province of Gansu, with an estimated geographical area of 27,117 km^2 and a population of 2.23 million individuals. In 2019, its GDP per capita was around \$5,130, well below the national GDP per capita of \$10,216.[8]

In Qingyang, all high schools track first-year students into achievement-based classrooms. The high school we focus on, Qingyang First High School, started this practice in 2015. In each academic year, high school administrators aim to place around 80 to 120 students in two high-achieving (HA) classrooms, while all other students are randomly allocated to regular classrooms. To determine classroom placement, students have to take a common exam at the beginning of their first year in high school. The exam comprises 5 subjects: Mathematics, Chinese Language, English Language, Physics and Chemistry. Students' total score, graded out of a possible 650 points, is calculated by taking the sum of their scores on these subjects.[9] The classroom placement exam (CPE) is administered by Qingyang First High School. Its content and grading scale are different than the high school entrance exam, which is administered at the city level.[10] However, they both cover similar topics so students

[8]Source: https://research.hktdc.com/en/data-and-profiles/mcpc/provinces/gansu/qingyang

[9]Students can earn a maximum of 150 points each on Mathematics, Chinese Language and English Language, and a maximum of 100 points each on Physics and Chemistry.

[10]The high school entrance exam includes 10 subjects and is graded out of 1,000 points. The subjects are: Mathematics (150 points), Chinese Language (150 points), English Language (150 points), Physics (100 points), Chemistry (100 points), History (80 points), Geology (80 points), Biology (70 points), Politics (70

do not have to study different material to prepare for each exam.

The top performers on the classroom placement exam are assigned to high-achieving classrooms. The combination of spots available in high-achieving classrooms each year and students' performance on the CPE creates a distinct CPE score cutoff for each cohort in our sample, whereby students scoring above their cohort's cutoff are assigned to high-achieving classrooms and those scoring below are placed in regular classrooms. Students' performance on the CPE is generally the only criterion taken into consideration when determining classroom placement.

Students take all subjects in the classroom they were assigned to, and generally stay in the same classroom until their last year of high school.[11] Students in high-achieving and regular classrooms follow the same national curriculum, and are evaluated using similar exams on all subjects. However, high-achieving classrooms are taught the same material at a faster pace, which gives teachers more time to delve deeper into each topic. For example, students in high-achieving classrooms may be responsible for knowing how to prove a certain mathematical theorem, while those in regular classrooms go over the theorem without the proof. Students in high-achieving classrooms are further given additional and more advanced in-class and at-home exercises. This type of differentiated instruction is a common feature of within-school tracking programs. For example, the gifted classes studied in Bui, Craig and Imberman (2014) and Card and Giuliano (2016) also cover the same curriculum as regular classes, but delve deeper into the material and provide a faster pace of instruction.

Our discussions with high school administrators indicate that there are several additional differences between classrooms. Teachers in high-achieving classrooms are on average higher-ranked than those in regular classrooms. A unique feature of the Chinese educational system is that teachers are assigned one of three official ranks, which influence their salaries. Teachers typically start at the lowest rank when they are first employed, and become eligible to apply for higher ranks after accumulating a few years of experience. However, promotion to higher ranks is not automatic and the evaluation process is quite rigorous. A committee selected by city officials evaluates each eligible teacher's file and takes into consideration his/her teaching performance, education level, publications, awards, and performance on an oral exam. Higher-ranked teachers are perceived to be of higher-quality in China, and previous studies suggest that they improve student outcomes. Indeed, Hoekstra, Mouganie

points), Physical Education (50 points).

[11]The type of classroom that students are assigned to in their second and third years of high school depends on the academic concentration chosen at the end of the first year. In general, most high-achieving students choose a Science concentration. Therefore, students who are placed in a high-achieving classroom in their first year and choose a science concentration, stay together in the same high-achieving classroom until their last year of high school. Those who pick an Arts concentration are instead reassigned to a regular Arts classroom in their second and third years of high school.

and Wang (2018) find that attending the most selective Chinese high schools substantially improves students' performance on the college entrance exam, and that this effect is driven by increased exposure to high-ranked teachers. Teachers do not receive additional training and do not need to acquire extra credentials to teach high-achieving classrooms.[12] In section 6, we use teacher rank as a proxy for teacher quality to provide evidence on the mechanisms driving our effects. We also look at teacher salaries and years of experience as alternative measures of quality.

Another difference between classrooms is that high-achieving classrooms are smaller in size than regular classrooms. Finally, by design, students in high-achieving classrooms are exposed to higher-ability peers compared to those in regular classrooms.

3 Data and Descriptive Statistics

We collected student-level administrative data directly from school administrators at Qingyang First High School, a large and selective high school in China. Our data comprise three student cohorts who first enrolled at the high school in the academic years 2015 to 2017.[13] Our data contain information on students' high school entrance exam (Zhongkao) scores in addition to scores on the separate classroom placement exam (CPE) administered by the high school to enrolled students. Importantly, starting with the 2015 entering cohort, Qingyang First High School began using results from the latter exam to track the highest scoring students into two high-achieving classrooms while randomly assigning all remaining students to other sections. We also have information on students' classroom section, gender, test scores in all subjects taken throughout their three years of high school, scores on the Chinese college entrance exam (Gaokao) and the name of the university students eventually attend. Finally, to quantify the differences between high-achieving and regular classrooms, we collected detailed information on teachers directly from the high school—namely their salary scale, years of experience and official rank.

Column (1) of Table 1 reports descriptive statistics for our full sample, i.e. for the 2,273 students who first enrolled in Qingyang First High School from 2015 to 2017. In Column (2), we also provide summary statistics for students in our marginal sample, i.e. the 1,788 students who scored within 75 points on either sides of the classroom placement exam

[12] Additionally, when within-school tracking was first introduced in Qingyang First High School in 2015, high-achieving classroom teachers were drawn from their regular teaching staff and no new teachers were hired for this purpose.

[13] We do not have data from previous years as tracking in Qingyang First High School was first implemented in 2015. Additionally, we do not collect data for cohorts who enrolled at the high school after 2017, as they are still too young for us to observe their postsecondary outcomes.

cutoff. Panel A shows means and standard deviations for students' baseline characteristics and outcomes. The proportion of male students in the overall and marginal samples are fairly similar at roughly 53 percent. The average high school entrance exam score for students in the overall sample is 790.5 out of a possible 1,000 points for the years 2015 to 2017 with a standard deviation of 88 points. The average is slightly higher for our marginal sample at 796.9 points. For the classroom placement exam, the average score for the overall sample is 413.4 out of a possible 650 points, and is also slightly higher for the marginal sample (429.9 points). The scores on this exam determine whether students are eligible to enroll in a high-achieving classroom, and hence will be used as our running variable (see section 4.1). Only 13.7 percent of students in the overall sample and 17.3 percent of those in the marginal sample are enrolled in high-achieving classrooms.

To examine the impact of high-achieving classrooms on short-term academic performance, we use students' scores on all exams taken in each of their three years in high school as outcomes. Students assigned to high-achieving and regular classrooms take the same exams in each grade. Accordingly, we take the average of all test scores in these exams for each of the three grades of high school. We then standardize average yearly performance for each grade by year of entry (i.e. by cohort). Average performance during the first three years of high school is higher for students in the marginal sample compared to the overall sample. This is expected given that students from the marginal sample are taken from a higher initial test score distribution. In particular, students in the marginal sample outscore those from the overall sample by 0.4, 0.35 and 0.329 standard deviations in grades 1 through 3 of high school, respectively. Additionally, around 87 percent of students in the selective high school we focus on choose a science academic concentration in the second year of high school. This proportion stands at roughly 90 percent for students in the marginal sample.

At the end of high school, students with a science academic concentration score an average of 510.77 points on the Gaokao college entrance exam with a standard deviation of 64.64. The minority of students in the arts concentration score an average of 527.46 points on their version of the college entrance exam. Students in the marginal sample perform even better attaining an average of 522 and 541 points in the science and arts Gaokao exam respectively, which is in line with their better performance during the first three years of high school. Overall, students enrolled in Qingyang first high school perform much better than most students in their province on the national college entrance exam. Indeed, the average science student in our high school scores in the top 11 to 12 percent of all Gaokao test takers in the province of Gansu, depending on the year. Additionally, the average student in the arts concentration scores in the top 5 to 7 percent of the province. Approximately the same proportion of students in both samples end up opting out of the college entrance exam (4.8

and 4.5 percent). The national college entrance exams are extremely high-stakes as they are the sole determinant of university access and quality in China. Unsurprisingly, the majority of students in our overall (90%) and marginal (92.5%) samples end up enrolling in university. This indicates that college access is not the margin of concern for this student population.

Indeed, the main reason that students compete to get into top-ranked high schools is because they increase access to selective universities by better preparing students for the college entrance exam. Officially, universities in China are broken down into tiers, with tier I universities being the most selective. As a result of their higher than national average performance on the college entrance exam, around 65 percent of students in Qingyang First High School attend a tier-I university. This number is even higher for the marginal sample (72 percent). However, top-performing students covet access to a narrower and more selective set of universities within the tier-I designation, the top 100 and top 40 national universities in China. The proportion of students in our high school that attend the coveted top 100 universities stands at about 25 percent for our overall sample and 30 percent for the marginal sample. Additionally, the proportion attending the most prestigious and coveted top 40 universities is 11.6 and 14.4 percent for our overall and marginal samples, respectively.

Students in our three cohorts are distributed across 43 distinct classrooms. We observe two high-achieving classrooms per entering cohort for a total of six high-achieving classrooms across all three cohorts. Panel B of Table 1 presents summary statistics for classroom-level characteristics. The average class size for students in the overall and marginal samples stands at around 58 students per classroom. Teachers' salaries are officially broken down into steps ranging from 7 to 40, with a higher number corresponding to a higher salary scale. The average salary scale for teachers in Qinyang First High School is 22.16 for the overall sample and 22.31 for the marginal sample. Additionally, teachers have an average of 16.8 years of experience. Finally, teachers are assigned one of three official ranks, with three being the highest and one the lowest. The proportion of top teachers, i.e. those in the highest rank category, stands at 25.8 and 26.4 percent for the overall and marginal samples, respectively.

4 Identification Strategy

4.1 Regression Discontinuity Design

The practice of tracking high school students into high-achieving classrooms is prevalent in China. In particular, all high schools in the Gansu province and most elite high schools in China have this form of tracking. The high school we focus on tracks top-performing first-year students into two high-achieving classrooms per year. Assignment to high-achieving classrooms is based solely on students' scores on the classroom placement exam. Accordingly, we use a regression discontinuity design (RD) to estimate the causal impact of high-achieving classroom attendance on academic performance and college outcomes (Imbens and Lemieux, 2008; Lee and Lemieux, 2010). The key identifying assumption underlying an RD design is that all determinants of future outcomes vary smoothly across the high-achieving classroom admissions threshold. This is likely to hold, as precisely manipulating scores on the classroom placement exam would be extremely difficult, if not impossible. This is because the cutoff scores are only determined after the exams are administered and graded. These cutoffs are determined based on percentile ranks, which are only calculated after the tests are graded. As a result, students and graders do not know the admission threshold for each academic year. Additionally, graders do not observe any identifying information on students.

All students in our data attend Qingyang First High School. Within this high school, two classrooms, per academic year, are consistently reserved for the highest-achieving students; which is roughly composed of the top-scoring students in the classroom placement examination, beyond a key threshold. In order to summarize the effects of attending high-achieving classrooms, we pool data across three different entering cohorts. Formally, we estimate the following reduced-form equation:

$$Y_{it} = \alpha + f(S_{it}) + \tau D_{it} + \delta X_i + \epsilon_{it}, \tag{1}$$

Where Y_{it} is the outcome of interest for student i in cohort t. D_{it} is a dummy variable indicating whether student i crosses the year-specific score threshold for attending a high-achieving classroom.[14] S represents students' classroom placement exam scores in the years 2015, 2016 and 2017 measured in points relative to the cutoff score for each respective academic year. Formally, $S_{it} = grade_{it} - \overline{grade_y}$ for all individuals within a year facing a common threshold y. The function $f(.)$ captures the underlying relationship between the running variable S_{it} and the dependent variable Y_{it}. We also allow the slopes of the fitted lines to differ on either side of the admissions threshold by interacting $f(.)$ with the treatment dummy D. X_i is a vector of students' predetermined 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. ϵ_{it} represents the error term. Finally, the parameter τ gives us the causal effect of being eligible to enroll in a high-achieving classroom—i.e., the reduced form estimate.

In our analysis, we specify $f(.)$ to be a linear function of S and estimate the equation over a narrow range of data, using local linear regressions with triangular and uniform kernels.

[14]We have three thresholds in total, each corresponding to a different academic year.

This approach generates estimates that are more local to the threshold without imposing any strong functional assumptions on the data. The preferred specifications in this paper are drawn from local linear regressions with optimal bandwidths chosen by the CCT robust data driven procedure as outlined in Calonico, Cattaneo and Titiunik (2014). Specifically, we use two separate MSE-optimal CCT bandwidth selectors—one for observations below the cutoff and one for those above. We do so because we have significantly more observations to the left of the cutoff as compared to the right given how selective the threshold is. Additionally, because the CCT bandwidth selector predicts different bandwidths depending on outcome, the number of observations in each regression may vary from one outcome to another. However, we also present results from a variety of different common bandwidths for all outcomes as a robustness check. Finally, given the discrete nature of our running variable, we report robust standard errors throughout (Kolesár and Rothe, 2018).

While we generally focus on reduced form estimates from specifications like (1), we also present coefficients from an instrumental variables type specification. This allows us to infer the average effect of attending a high-achieving classroom as opposed to the intent-to-treat (ITT) effect only. Formally, we estimate:

$$Y_{it} = \theta + h(S_{it}) + \beta E(C_{it}|S_{it}) + \gamma X_i + \mu_{it}, \tag{2}$$

$$E(C_{it}|S_{it}) = \nu + g(S_{it}) + \lambda D_{it} + \theta X_i + \zeta_{it}, \tag{3}$$

where (3) is the "first stage" or the compliance ratio C_{it}. It denotes the share of students who actually enroll in a high-achieving classroom. μ_{it} and ζ_{it} are error terms. β from equation (2) gives us the local average treatment estimate (LATE) of attending a high-achieving classroom in a 2SLS framework. This is equivalent to the Wald estimate and can be informally computed by dividing the ITT estimate $\hat{\tau}$ in equation (1) by the first stage estimate $\hat{\lambda}$ from equation (3).

4.2 Tests of Identification

Given the nature of how students are assigned to high-achieving classrooms, we believe it is very unlikely that students are able to precisely manipulate their scores relative to the cutoff. Nonetheless, we provide two formal empirical tests to alleviate concerns over manipulation of the running variable.

We first assess whether there is evidence of bunching around the high-achieving classroom admission threshold. Indeed, if students or graders could manipulate exam scores relative to the cutoff, we would expect to see too few students just short of the cutoff coupled with too many students just exceeding the cutoff. Results from this exercise are summarized in Figure 1, which shows the density function representing the share of students scoring 50 points below and above the classroom placement exam cutoff. Specifically, we find no evidence of a discontinuity (bunching) in the density function using the local polynomial density estimation testing procedure proposed in Catteneo, Janson and Ma (2020). Formally, we estimate a p-value of 0.495 and reject the hypothesis that the density function varies discontinuously at the cutoff.

We also test whether observed determinants of achievement are smooth across the threshold. Indeed, if our identifying assumption holds, we would expect predetermined characteristics of student achievement to vary smoothly across the admissions threshold. Conversely, if students or graders are manipulating scores around the threshold, then we would expect to see students with different characteristics on either sides of the cutoff. Predetermined student characteristics in our data are limited and include gender and test scores on the high school entrance exam—taken just prior to the classroom placement exam. We test whether there is evidence that these two covariates vary discontinuously at the cutoff. Figures 2a and 2b plot the relationship between each of these covariates and the running variable. The figures take the same form as those after them in that circles represent local averages of the outcome over a 5 points score range. The running variable is defined as distance of students' scores from the classroom placement exam cutoff. The cutoff is represented by a 0 on the x-axis. We show results using a bandwidth of 75 points on either sides of the cutoff using a linear fit. Visual evidence suggests that high school entrance exam scores (Figure 2a) and the likelihood that a student is male (Figure 2b) vary smoothly at the cutoff, in line with our identifying assumption.

We present corresponding regression discontinuity estimates taken from equation (1) in Table 2. We report coefficients from local linear regressions using as an outcome: the likelihood a student is male in Columns (1) and (2) and high school entrance exam scores in Columns (3) and (4). All regressions use a bandwidth predicted by the CCT optimal bandwidth selector, as detailed in section 4.1. Columns (1) and (3) show estimates using a triangular kernel function that gives more weight to points close to the cutoff. We also show estimates using a uniform kernel, that give equal weight to all points, in columns (2) and (4). Consistent with the visual evidence, we are unable to detect any significant discontinuities at the cutoff in terms of student gender or high school entrance exam scores. These results hold regardless of kernel choice. We also show that estimates are robust to varying bandwidth choices in Appendix Table A1. Specifically, we are unable to detect any significant discontinuity in gender or high school entrance exam scores using bandwidths of 50, 75 and 100 score points on either side of the high-achieving classroom admissions threshold.

5 Results

5.1 First Stage—Likelihood of Enrolling in High-Achieving Classrooms

We begin by presenting evidence that the classroom placement assignment rule was binding in practice. To do so, we show visual evidence that students are discontinuously admitted to high-achieving classrooms based on their scores in the classroom placement exam. Figure 3 summarizes results from this exercise where bins represent local averages over a 5 point score range. We use a linear fit on either side of the cutoff to approximate the discontinuity. The figure reveals a large and positive discontinuity in the likelihood that students enroll in a high-achieving classroom at the admissions cutoff. Corresponding regression discontinuity estimates presented in Table 3 indicate a high compliance rate with discontinuity coefficients ranging from 77.5 to 81.8 percentage points depending on kernel choice and controls. In Table A2 of the Appendix, we show that these results are robust to various bandwidth choices. We conclude that scoring just above the classroom placement exam threshold increases students' likelihood of being in a high-achieving classroom by approximately 80 percentage points.

5.2 Performance in High School

We next examine the short-run effects of attending high-achieving classrooms. In particular, we focus on students' academic performance during their three years of high school. Students sit for numerous common exams in various subjects throughout the year, which are meant to measure their progress in a given grade. Importantly, these exams are common across all classrooms in a given grade and year. This enables us examine whether high-achieving classrooms give students an advantage in terms of high school performance, over students in regular classrooms. We look at performance on three subjects, which students are consistently tested on throughout the three years of high school: Mathematics, English and Chinese.[15] Specifically, we focus on average test scores in these three subjects for each year of high school. To ease cross-cohort comparisons, we standardize scores by cohort and grade.[16] We then look at whether these standardized test scores, measured in each year of high school, discontinuously change at the high-achieving classroom admissions cutoff.

[15]On average, students take 6 sets of exams in various subjects in a given grade and year. We use scores on Mathematics, English and Chinese, as these are the only subjects that are included in all sets of exams.

[16]Since students are divided into science and arts tracks in the 2nd and 3rd year of high school, we also standardize scores within tracks for those years.

For all three subjects, Figures 4 to 6 graphically show results in the first through third years of high school, respectively. Figures 4a, 5a and 6a respectively show clear and positive increases in Math performance at the threshold in years 1, 2 and 3 of high school. Figures 4b, 5b and 6b suggest that there are also some improvements in Chinese test scores at the cutoff, but the discontinuities are visually less compelling than those for Math. On the other hand, we find no evidence of a jump at the threshold when looking at performance in English, regardless of high school year (Figures 4c, 5c and 6c).

Formal regression discontinuity estimates from equations as in (1) are presented in Table 4. Panel A shows reduced form local linear regression estimates on first year high school performance. Consistent with the visual evidence, we find that scoring just above the classroom placement exam threshold increases Mathematics test scores by 23 to 27 percent of a standard deviation using a triangular or uniform kernel (Columns (1) and (2)). However, we find no evidence that threshold crossing significantly impacts first-year performance in Chinese or English subjects in Columns (3) through (6).

We present reduced form effects on similar outcomes during the second year of high school in Table 4. We find strong evidence that threshold crossing increases performance in Mathematics by 27 to 31 percent of a standard deviation. We also find some evidence that performance in Chinese is also increased, though this result is not robust to kernel choice. Additionally, we find no evidence that test scores in English are improved in the second year of high school. Estimates in Panel C of Table 4 summarize effects during the final year of high school. Similar to the first two years, we find that high-achieving classroom eligibility increases performance by 24 to 26 percent of a standard deviation in Mathematics. However, we find less compelling evidence of a significant change in performance on Chinese or English test scores.

Finally, we check whether these results are robust to bandwidth choice in Appendix Tables A3 through A5. We find consistent and robust evidence that threshold crossing impacts Mathematics test scores in all three years, but had no impact on English test scores. In terms of performance in Chinese, our findings are less clear as we detect significant increases with larger bandwidths. Taken together, our results indicate that placement in a high-achieving classroom in the first year of high school substantially improves students' contemporaneous performance in math, and these benefits do not fade out as they persist until the last year of high school. On the other hand, high-achieving classrooms' effects on performance in Chinese or English are more muted.

5.3 Performance on College Entrance Exam and College Outcomes

We now turn to longer-term outcomes that directly impact students' university choices. We begin by looking at student performance on the high-stakes national college entrance exams. These exams are conducted at the end of high school and are the sole determinant of college eligibility in China. Figure 7a plots students' standardized college entrance exam scores as a function of the running variable.[17] We see a sizable increase in college entrance exam scores at the classroom placement exam threshold. We present formal regression results in Columns (1) and (2) of Table 5. Specifically, reduced form local linear estimates indicate that threshold-crossing increases scores on the college entrance exam by around 27 to 28 percent of a standard deviation. We also report local average treatment effects of attending high-achieving classrooms by re-scaling the intent-to-treat estimates in the second row by the previously estimated discontinuity in the likelihood of attending a high-achieving classroom. Results are shown in the third row of Table 5 and indicate that enrolling in a high-achieving classroom increases college entrance exam test scores by 35 percent of a standard deviation.

We also present effects on college entrance exam performance by subject in Appendix Figure A1 and Table A6. In particular, we focus on scores in the four main components of the exam: Mathematics, English, Chinese and a "Main Subject". The Main Subject is Sciences (i.e., an exam covering Physics, Chemistry and Biology) for students in the science concentration and Arts (i.e., an exam covering History, Politics and Geography) for those in the arts concentration. Visual evidence presented in Figure A1 indicates that high-achieving classrooms significantly improve students' scores in the Mathematics and Main Subjects components of the college entrance exam. We find weaker evidence of improvements in Chinese scores and no evidence of changes in performance on the English portion of the exam at the threshold. Regression estimates presented in Table A6 are in line with the visual evidence. We find large, robust and significant gains in the Mathematics and "Main Subjects" portion of the exam on the order of 30 to 35 percent of a standard deviation. Conversely, we find no significant impacts on Chinese and English performance.

Next, we look at crucial university choices that are directly affected by students' exam scores on the college entrance exam. In particular, we focus on four outcomes: enrolling in any Chinese university, a first-tier university, a top-100 university and a top-40 university. We present graphical RD results for these four outcomes in Panels (b) through (e) of Figure 7.

[17] We standardize college entrance exam scores by year and high school concentration. In Section 5.4, we show that the likelihood of choosing a science or arts concentration is smooth at the cutoff, mostly because virtually all students around the cutoff select a science concentration. As a result, the choice of standardizing college entrance exam scores within concentrations has no substantial effect on results.

Unsurprisingly, we find no visual evidence of a discontinuity in the likelihood that students attend any Chinese university (Figure 7b). This is because the high school we analyze is highly selective and the margin of interest for enrolled students is most likely college quality as opposed to just access. Indeed, almost all students around the cutoff end up enrolled in a university as shown in Figure 7b.[18] As a result, we next look at effects on college quality, the more likely affected margin for students in our sample. We find no compelling visual evidence of a change in the likelihood that students attend a first-tier university (Figure 7c). This is most likely because a significant portion of students around the cutoff end up attending a first-tier university. We therefore use a narrower definition of college quality as our outcome: enrollment in the more selective and prestigious top-100 national universities. Indeed, we find a compelling increase in the likelihood of attending top-100 national universities (Figure 7d) at the threshold. Additionally, while a linear fit suggests a potential discontinuity in the chances of attending top-40 universities (Figure 7e), this seems to be largely driven by noise, most likely because this is a rare outcome.

We turn to formal regression estimates to get a sense of the magnitude of these results. Local linear estimates presented in Columns (3) through (6) of Table 5 indicate no statistical link between being in a high-achieving classroom and the likelihood of attending any university or a first-tier university. On the other hand, we find a substantial increase in the likelihood of attending a top-100 university, with intent-to-treat estimates ranging from a 16.5 to 18.9 percentage points in columns (7) and (8) of Table 5. This translates into LATE estimates of 22.5 to 24.1 percentage points indicating that attending a high-achieving classroom increases students' chances of enrolling in a top-100 university by roughly 50 percent. In line with the visual evidence, estimates in Columns (9) and (10) are positive but fairly imprecise, precluding us from making any strong conclusions regarding the causal link between high-achieving classroom attendance and top-40 university enrollment.[19] Finally, while we do not have data on students' labor market outcomes, findings from this section suggest that attending a high-achieving classroom in high school may have significant impacts on later lifetime outcomes. Indeed, Jia and Li (2021) show that the wage premium to attending a top-100 university in China ranges from 28 to 45 percent.

[18] We are unable to observe if students not attending university in China are instead enrolled in a university abroad. However, anecdotal evidence from our conversations with high school officials reveal that students rarely end up attending a university outside of China. This is expected given that the city and province we look at are both relatively poor.

[19] Estimates reported in Appendix Table A7 indicate that results for college exam performance and university choice are mostly robust to various bandwidths. The major exception is that local linear estimates on top-40 university enrollment are statistically significant for larger bandwidths.

5.4 Threats to Identification

A potential threat to identification is the possibility that students endogenously select into taking the college entrance exam.[20] Indeed, if students just above the threshold are more likely to sit for the college entrance exam, then this would complicate the interpretation of our longer-term effects. Appendix Figure A2a shows that the likelihood of opting out of the college entrance exam does not vary discontinuously at the cutoff. Formal regression estimates in Columns (1) and (2) of Table 6 also show no statistically significant link between high-achieving classroom enrollment and selection out of the college entrance exam.

An equally worrying threat to identification is if enrolling in a high-achieving classroom influences students' academic concentration choice in the second year of high school, i.e. whether they enroll in a science or arts concentration. For instance, if being in a high-achieving classroom causes students to enroll in the science concentration at higher rates, then that difference, rather than a broader sense of improved classroom quality, could drive our results on longer term outcomes. To alleviate such concerns, Figure A2b plots the likelihood of choosing a science versus arts concentration as a function of the running variable. We see no evidence of a discontinuity at the cutoff. Additionally, corresponding local linear estimates presented in Columns (3) and (4) indicate that high-achieving classroom enrollment does not influence academic concentration choice the following year. This is not surprising given the very high rates of science concentration enrollment for students on either side of the cutoff.[21]

6 Mechanisms

We now turn to the question of why there are sizable returns to being in a high-achieving classroom. As detailed in section 2.2, our discussions with high school administrators suggest that high-achieving and regular classrooms differ in terms of several important inputs into education: peer quality, class size, and teacher quality. One advantage of our data is that we can document whether classrooms actually differ along these dimensions and the magnitudes of those differences, allowing us to better understand the mechanisms behind our effects. We

[20] A student may opt out of taking the college entrance exam either because they have dropped out of the education sector altogether or because they want to independently sit for it the following year. Taking the college entrance exam is unrelated to grade repetition. Grade repetition is extremely rare in our context because students are generally not allowed to repeat any year of high school without special permission from school officials.

[21] In Appendix Table A8, we show that findings from this section are robust to bandwidth choice. Specifically, we are unable to detect any significant effects on college entrance exam take-up or academic concentration choice from local linear regressions using bandwidths of 50, 75 or 100 points either side of the cutoff.

look at these three inputs in the first year of high school, i.e. during the year students are initially tracked into different classrooms.

First, since students are assigned to high-achieving classrooms based on their academic performance, classrooms naturally differ in terms of peer quality. For each student, we construct a leave-one-out classroom-level peer quality measure (i.e., excluding the student themself) using peers' standardized scores on the high school entrance exam. Students take this exam prior to enrolling in high school and interacting with their high school peers. Figure 8a plots average peer exam scores, as a function of distance of students' scores from the classroom placement exam cutoff. The figure shows a large increase in classroom peer quality at the threshold. In Table 7, the first two rows of Columns (1) and (2) reveal that the reduced form effect on peer exam scores ranges from 1.031 to 1.078 standard deviations. The corresponding LATE estimates (third row) indicate that attending a high-achieving classroom is linked with having classroom peers who are, on average, 1.35 standard deviations higher ability than those found in regular classrooms.

These estimates show that students in our setting have a large increase in peer quality when they are placed in high-achieving classrooms. Nonetheless, it is unclear to what extent peer quality is driving our main effects. On one hand, a large body of work documents that an increase in mean peer ability improves students' academic success (Sacerdote, 2011). On the other hand, several recent studies—which also use a regression discontinuity design—find that tracking does not necessarily improve high-achieving students' outcomes despite exposing them to higher-quality peers. Indeed, Duflo, Dupas and Kremer (2011) and Bui, Craig and Imberman (2014) show that students marginally placed in high-achieving and gifted classes are exposed to higher-ability peers, but do not have higher test scores than those who are marginally assigned to regular classes. These findings are consistent with studies by Abdulkadiroğlu, Angrist and Pathak (2014) and Dobbie and Fryer (2014), who show that marginally gaining admission to elite high schools in Boston and New York, which enroll high-achieving peers, does not result in improved test scores.

Second, we show that students just above the classroom placement admissions cutoff are, on average, in smaller classes. Visual evidence in Figure 8b reveals a substantial drop in class size at the cutoff. Reduced form RD estimates in columns (3) and (4) of Table 7 show that scoring just above the cutoff reduces average class size by 2.94 to 3.92 students during the first year of high school. The corresponding LATE estimate indicates that students in high-achieving classrooms have an average of 4.5 less students in their classroom. Looking at the previous literature on the returns to class size, it is also ambiguous whether smaller classes can explain our main results. Seminal studies show that small classes increase test scores, years of education, college attendance and graduation of kindergarten and primary school students (Angrist and Lavy, 1999; Chetty et al., 2011; Fredriksson, Öckert, and Oosterbeek, 2013). However, other studies indicate that older students, in middle school and high school, may not see improvements in test scores or completed schooling due to small classes (Leuven, Oosterbeek and Rønning, 2008; Leuven and Løkken, 2020).

To get a sense of the importance of class size in our context, we perform back-of-the-envelope calculations. Fredriksson, Öckert, and Oosterbeek (2013) estimate that placement in a classroom with one less student during grades 4 to 6 (from ages 10 to 13) improves test scores at age 13 by 0.03 standard deviations.[22] In our setting, we assume that students placed in high-achieving classrooms stay in these classrooms until the end of high school (i.e., for 3 years), and hence benefit from having 4.5 less students in their classroom for all 3 years. This implies that we should see an increase in test scores due to smaller classes of at most 0.135 standard deviations (=0.03×4.5). However, this upper bound is substantially less than the documented LATE estimate of 0.349 standard deviations we find in Table 5 from attending high-achieving classrooms. This indicates that while reduced class size may explain part of our main effects, there are other important channels that drive our results.

The final input we examine is teacher quality, which has been shown to be an important predictor of student performance in many other settings (Chetty, Friedman and Rockoff, 2014; Jackson, 2018). We do so by exploiting a unique feature of the Chinese education system which designates official ranks to teachers. Indeed, in our context, teachers are awarded a rank of 1 through 3 with the higher number indicating a better ranked teacher. As detailed in Section 2.2, promotion to a higher rank is difficult to attain and teachers wishing to do so have to go through a rigorous evaluation process. Higher teacher rank has been previously shown to improve students' test scores in China (Hoekstra, Mouganie and Wang, 2018). Graphical evidence in Figure 8c indicates that students just above the cutoff are exposed to teachers with a higher rank during the first year of high school. We provide formal evidence from regressions as in equation (1) using standardized teacher rank as an outcome. Estimates from the final two columns of Table 7 show that these effects are statistically significant and that students who are eligible to enroll in high-achieving classrooms are matched with teachers who are 0.36 to 0.40 standard deviations higher-ranked, on average.[23]

We further decompose teacher rank by subject in Figure 9 and Appendix Table A10. We find that our overall teacher effects are driven by significantly better teachers in Mathematics, followed by English. We find a much smaller, but statistically significant, increase in Chinese

[22]The magnitudes of class size effects in Fredriksson, Öckert, and Oosterbeek (2013) are comparable to those found in Angrist and Lavy (1999) and Chetty et al. (2011).

[23]In Appendix Table A9, we further show that estimates of the impact of high-achieving classrooms on teacher rank, peer quality and class size are all robust to bandwidth choice.

teacher rank at the cutoff. Specifically, high-achieving classroom placement increases math teachers' rank by 1.9 standard deviations, English teachers' rank by 0.886 standard deviations, and Chinese teachers' rank by a marginally significant 0.2 standard deviations. At first glance, the fact that we observe an increase in English teacher rank but no improvement in English test scores at the cutoff, suggests that teacher rank may not explain our main effects. However, previous studies find that while teacher quality is a strong predictor of math achievement, it has a much weaker effect on performance in English. This is because mathematics is believed to be mainly learned in the classroom, while English skills are often acquired outside of school (Jackson, Rockoff and Staiger, 2014).

We show that our findings on teachers are robust to various definitions of teacher quality. We first provide estimates using a binary definition of teacher quality. Specifically, we use as an outcome a dummy variable that equals 1 if a teacher has the highest rank (or is a "top teacher"), and 0 if he/she is of lower rank. Using this definition, Appendix Figure A3a shows that threshold-crossing leads to approximately a 10 percentage point increase in the likelihood that students match with top teachers. We further use teacher salaries and years of experience as alternative measures of teacher quality. Teachers in China are paid according to a salary scale ranging from 7 to 40, with a higher number indicating a higher pay scale. Appendix Figure A3b reveals a large and significant increase in teachers' salary scale at the cutoff. Finally, Appendix Figure A3c indicates that students who are eligible to attend high-achieving classrooms are matched to teachers who have 2 additional years of experience, on average. These findings are in line with those using our initial definition of teacher quality and indicate that students who score above the classroom placement exam threshold are matched with significantly higher quality teachers.

Taken together, these results indicate that improved teacher quality is likely to explain at least part of our high-achieving classroom effects. This is corroborated by previous evidence from China on the importance of high-ranked teachers in student learning. Using a regression discontinuity design, Hoekstra, Mouganie and Wang (2018) show that being marginally admitted to the most selective Chinese high schools substantially improves students' performance on the college entrance exam. Selective high schools enroll higher-achieving peers and employ higher-ranked teachers than other schools. The authors however find that the academic benefits of selective high schools in China are driven by teacher quality and not peer ability.

One final potential channel, which we cannot quantify, is that students in high-achieving classrooms are taught at a faster pace and delve deeper into topics, despite following the same curriculum as regular classrooms. Duflo, Dupas and Kremer (2011) further highlight that tracking benefits students if it provides them to have a level of instruction that matches their abilities. In our setting, this implies that students benefit from high-achieving classrooms if their level of instruction is better suited for their abilities than regular classrooms.

In summary, findings from this section reveal that students who are marginally placed in high-achieving classrooms are exposed to higher-quality peers and teachers, as well as smaller class sizes. We cannot rule out that the short and longer-term benefits of high-achieving classrooms are due to peer quality and class size. However, the large effects we document on teacher quality coupled with an extensive literature documenting the importance of teacher quality as an input into education suggests that teachers may be the largest driver of our documented findings.

7 Conclusion

This paper provides new evidence on the impacts of within-high school tracking on high-achieving students' long-term academic success. We collect rich and unique data from a large and selective high school in China, where first-year students are allocated into high-achieving and regular classrooms solely based on their performance on a common exam. Using a regression discontinuity design, we show that placement in a high-achieving classroom largely improves performance in math in all three years of high school. The benefits of high-achieving classrooms persist even after students graduate from high school. Indeed, being in a high-achieving classroom increases students' scores on the high-stakes national college entrance exam by 0.28 standard deviations. Additionally, we find that while high-achieving classrooms do not impact access to college, they do increase students' enrollment in the most prestigious and selective Chinese universities (i.e., in the top 100 or Project 211 universities) by 50 percent. Since attending these universities has been previously shown to substantially increase future wages (Jia and Li, 2021), our results suggest that enrolling in high-achieving classrooms can have large labor market returns.

Our data allow us to explore the mechanisms driving the benefits of high-achieving classrooms. We show that students assigned to high-achieving classrooms are exposed to higher-ability peers and smaller class sizes, compared to those placed in regular classrooms. Additionally, students in high-achieving classrooms are exposed to teachers who are higher-ranked, earn higher salaries and have more years of teaching experience. Finally, students may benefit from receiving an instruction that is tailored to their abilities, as instructors in high-achieving classrooms delve deeper into topics and teach at a faster pace.

Our finding that students substantially benefit from high-achieving classrooms has important implications for current policy debates on the costs and benefits of school tracking. Indeed, separating students into achievement-based classrooms is highly controversial as opponents argue that it may exacerbate socioeconomic inequalities and question its potential benefits. These arguments have pushed several school districts in the United States and Canada to consider eliminating this type of tracking. While our results cannot speak to whether tracking exacerbates socioeconomic inequalities, they do indicate that high-achieving students may miss out on substantial benefits if they lose the opportunity to attend high-achieving classrooms.

References

Abdulkadiroğlu, Atila, Joshua Angrist, and Parag Pathak. 2014. The elite illusion: Achievement effects at Boston and New York exam schools. *Econometrica* 82 (1): 137-196.

Abdulkadiroğlu, Atila, Joshua D. Angrist, Susan M. Dynarski, Thomas J. Kane, and Parag A. Pathak. 2011. Accountability and flexibility in public schools: Evidence from Boston's charters and pilots. *The Quarterly Journal of Economics* 126 (2): 699-748.

Altmejd, Adam, Andrés Barrios-Fernández, Marin Drlje, Joshua Goodman, Michael Hurwitz, Dejan Kovac, Christine Mulhern, Christopher Neilson, and Jonathan Smith. 2021. O brother, where start thou? Sibling spillovers on college and major choice in four countries. *The Quarterly Journal of Economics*. Forthcoming.

Angrist, Joshua D. and Victor Lavy. 1999. Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics* 114(2): 533-575.

Angrist, Joshua D., Susan M. Dynarski, Thomas J. Kane, Parag A. Pathak, and Christopher R. Walters. 2010. Inputs and impacts in charter schools: KIPP Lynn. *American Economic Review* 100 (2): 239-243.

Angrist, Joshua D., Parag A. Pathak, and Christopher R. Walters. 2013. Explaining charter school effectiveness. *American Economic Journal: Applied Economics* 5 (4): 1-27.

Berkowitz, Daniel, and Mark Hoekstra. 2011. Does high school quality matter? Evidence from admissions data. *Economics of Education Review* 30 (2): 280-288.

Betts, Julian R. 2011. The economics of tracking in education. *In Handbook of the Economics of Education* Vol. 3, eds. Eric A. Hanushek, Stephen Machin, and Ludger Woessmann: 341–381. Elsevier.

Betts, Julian R., and Jamie L. Shkolnik. 2000. The effects of ability grouping on student achievement and resource allocation in secondary schools. *Economics of Education Review* 19 (1): 1-15.

Beuermann, Diether W., and C. Kirabo Jackson. 2018. The short and long-run effects of attending the schools that parents prefer. *NBER Working Paper* no. 24920.

Black, Sandra E., Jeffrey T. Denning, and Jesse Rothstein. 2020. Winners and losers? the effect of gaining and losing access to selective colleges on education and labor market outcomes. *National Bureau of Economic Research Working Paper* no. 26821.

Brunello, Giorgio, and Daniele Checchi. 2007. Does school tracking affect equality of opportunity? New international evidence. *Economic Policy* 22 (52): 782-861.

Bui, Sa A., Steven G. Craig, and Scott A. Imberman. 2014. Is gifted education a bright idea? Assessing the impact of gifted and talented programs on students. *American Economic Journal: Economic Policy* 6 (3): 30-62.

Callahan, Carolyn M., Tonya R. Moon, and Sarah Oh. 2013. Status of high school gifted programs. National Research Center on the Gifted and Talented. University of Virginia.

Callahan, Carolyn M., Tonya R. Moon, and Sarah Oh. 2017. Describing the status of programs for the gifted: A call for action. *Journal for the Education of the Gifted* 40 (1): 20-49.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2014. Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica* 82 (6): 2295-2326.

Canaan, Serena, and Pierre Mouganie. 2018. Returns to education quality for low-skilled students: Evidence from a discontinuity. *Journal of Labor Economics* 36 (2): 395-436.

Card, David, and Laura Giuliano. 2016. Can tracking raise the test scores of high-ability minority students?. *American Economic Review* 106 (10): 2783-2816.

Castleman, Benjamin, and Joshua Goodman. 2018. Intensive college counseling and the enrollment and persistence of low-income students. *Education Finance and Policy* 13 (1): 19-41.

Cattaneo, Matias D., Michael Jansson, and Xinwei Ma. 2020. Simple local polynomial density estimators. *Journal of the American Statistical Association* 115 (531): 1449-1455.

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. 2011. How does your kindergarten classroom affect your earnings? Evidence from Project STAR. *The Quarterly Journal of Economics* 126 (4): 1593-1660.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American economic review* 104 (9): 2633-2679.

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 (1): 150-176.

Cohodes, Sarah R. 2020. The long-run impacts of specialized programming for high-achieving students. *American Economic Journal: Economic Policy* 12 (1): 127-166.

Cohodes, Sarah R., and Joshua S. Goodman. 2014. Merit aid, college quality, and college completion: Massachusetts' Adams scholarship as an in-kind subsidy. *American Economic Journal: Applied Economics* 6 (4): 251-285.

Cohodes, Sarah R., Elizabeth M. Setren, and Christopher R. Walters. 2021. Can successful schools replicate? Scaling up Boston's charter school sector. *American Economic Journal: Economic Policy* 13 (1): 138-67.

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: 128-142.

Dillon, Eleanor Wiske, and Jeffrey Andrew Smith. 2017. Determinants of the match between student ability and college quality. *Journal of Labor Economics* 35 (1): 45-66.

Dobbie, Will, and Roland G. Fryer Jr. 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 (3): 58-75.

Dobbie, Will, and Roland G. Fryer Jr. 2015. The medium-term impacts of high-achieving charter schools. *Journal of Political Economy* 123 (5): 985-1037.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2011. Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *American Economic Review* 101 (5): 1739-1774.

Dynarski, Susan, C. J. Libassi, Katherine Michelmore, and Stephanie Owen. 2021. Closing the gap: The effect of reducing complexity and uncertainty in college pricing on the choices of low-income students. *American Economic Review* 111 (6): 1721-1756.

Figlio, David N., and Marianne E. Page. 2002. School choice and the distributional effects of ability tracking: does separation increase inequality?. *Journal of Urban Economics* 51 (3): 497-514.

Fredriksson, Peter, Björn Öckert, and Hessel Oosterbeek. 2013. Long-term effects of class size. *The Quarterly Journal of Economics* 128 (1): 249-285.

Hanushek, Eric A., and Ludger Wößmann. 2006. Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *The Economic Journal* 116 (510): C63-C76.

Hoekstra, Mark. 2009. The effect of attending the flagship state university on earnings: A discontinuity-based approach. *The Review of Economics and Statistics* 91 (4): 717-724.

Hoekstra, Mark, Pierre Mouganie, and Yaojing Wang. 2018. Peer quality and the academic benefits to attending better schools. *Journal of Labor Economics* 36 (4): 841-884.

Hoxby, Caroline M., and Christopher Avery. 2012. The missing "one-offs": The hidden supply of high-achieving, low income students. *National Bureau of Economic Research Working Paper* no. 18586.

Hoxby, Caroline M., and Sarah Turner. 2013. *Informing students about their college options: A proposal for broadening the expanding college opportunities project.* The Hamilton Project. The Brookings Institution.

Imbens, Guido W., and Thomas Lemieux. 2008. Regression discontinuity designs: A guide to practice. *Journal of Econometrics* 142 (2): 615-635.

Jackson, C. Kirabo. 2010a. A little now for a lot later a look at a Texas advanced placement incentive program. *Journal of Human Resources* 45 (3): 591-639.

Jackson, C. Kirabo. 2010b. Do students benefit from attending better schools? Evidence from rule-based student assignments in Trinidad and Tobago. *The Economic Journal* 120 (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.

Jackson, C. Kirabo. 2014. Do college-preparatory programs improve long-term outcomes?. *Economic Inquiry* 52 (1): 72-99.

Jackson, C. Kirabo. 2018. What do test scores miss? The importance of teacher effects on non-test score outcomes. *Journal of Political Economy* 126 (5): 2072-2107.

Jackson, C. Kirabo, Jonah E. Rockoff, and Douglas O. Staiger. 2014. Teacher effects and teacher-related policies. *Annual Review of Economics* 6 (1): 801-825.

Jia, Ruixue, and Hongbin Li. 2021. Just above the exam cutoff score: Elite college admission and wages in China. *Journal of Public Economics* 196: 104371.

Kolesár, Michal, and Christoph Rothe. 2018. Inference in regression discontinuity designs with a discrete running variable. *American Economic Review* 108 (8): 2277-2304.

Lee, David S., and Thomas Lemieux. 2010. Regression discontinuity designs in economics. *Journal of Economic Literature* 48 (2): 281-355.

Lefgren, Lars. 2004. Educational peer effects and the Chicago public schools. *Journal of urban Economics* 56 (2): 169-191.

Leuven, Edwin, Hessel Oosterbeek, and Marte Rønning. 2008. Quasi-experimental estimates of the effect of class size on achievement in Norway. *The Scandinavian Journal of Economics* 110 (4): 663-693.

Leuven, Edwin, and Sturla A. Løkken. 2020. Long-term impacts of class size in compulsory school. *Journal of Human Resources* 55 (1): 309-348.

Pallais, Amanda. 2015. Small differences that matter: Mistakes in applying to college. *Journal of Labor Economics* 33 (2): 493-520.

Pop-Eleches, Cristian, and Miguel Urquiola. 2013. Going to a better school: Effects and behavioral responses. *American Economic Review* 103 (4): 1289-1324.

Sacerdote, Bruce. 2011. Peer effects in education: How might they work, how big are they and how much do we know thus far?. *In Handbook of the Economics of Education*, vol. 3, pp. 249-277. Elsevier.

Schütz, Gabriela, Heinrich W. Ursprung, and Ludger Wößmann. 2008. Education policy and equality of opportunity. *Kyklos* 61 (2): 279-308.

Tangvatcharapong, Meradee. 2020. The impact of school tracking and peer quality on student achievement: Regression discontinuity evidence from Thailand. *Unpublished Manuscript*.

The New York Times. 2021. *De Blasio to phase out N.Y.C. gifted and talented program*. URL: https://www.nytimes.com/2021/10/08/nyregion/gifted-talented-nyc-schools.html. Accessed: 2021-11-12.

Vardardottir, Arna. 2013. Peer effects and academic achievement: A regression discontinuity approach. *Economics of Education Review* 36: 108-121.

Welsch, David M., and David M. Zimmer. 2018. Do high school gifted programs lead to later-in-life success? *Journal of Labor Research* 39 (2): 201-218.

Zimmer, Ron. 2003. A new twist in the educational tracking debate. *Economics of Education Review* 22 (3): 307-315.

A Figures

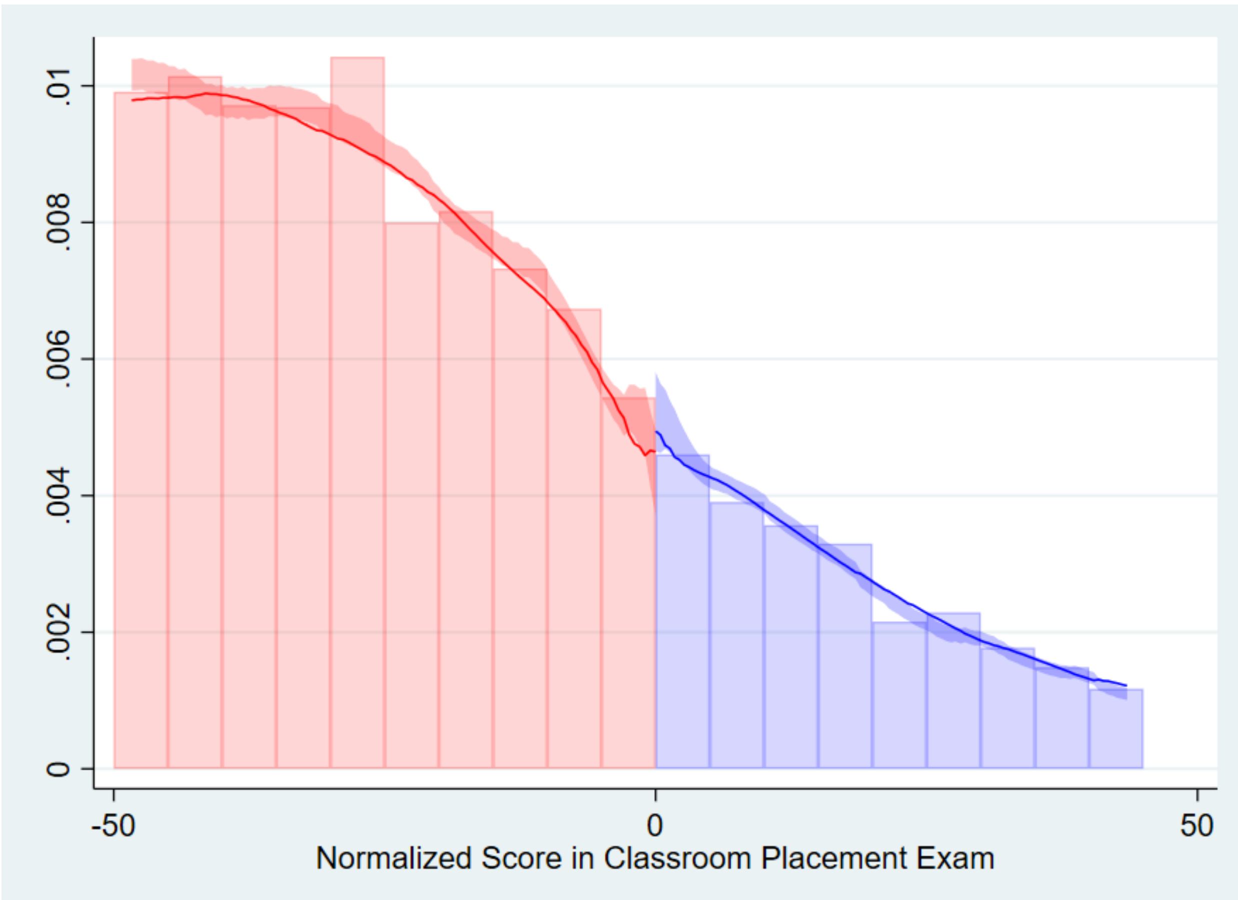
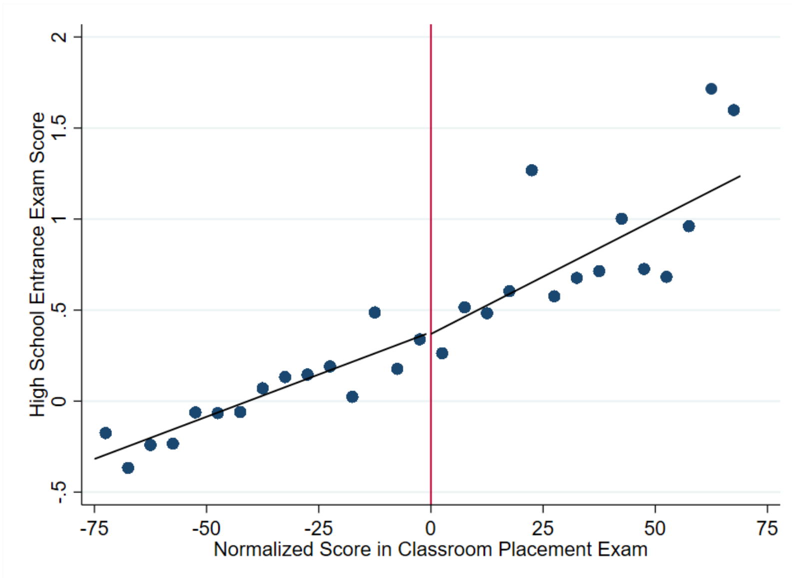
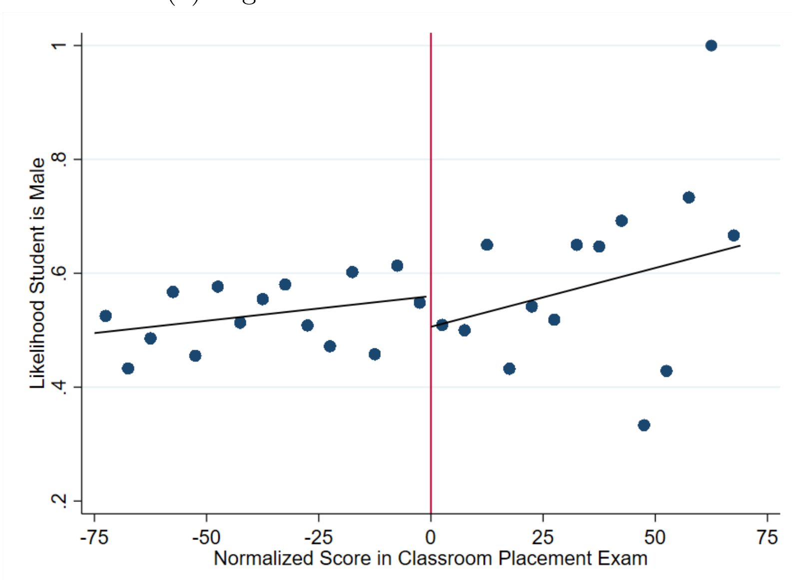


Figure 1: Test of running variable density smoothness around cutoff

Notes: Sample includes students who entered high school from 2015 to 2017. Bars represent frequency distribution over a 5 point score range. The above figure implements manipulation testing procedures using the local polynomial density estimators proposed in Cattaneo, Jansson and Ma (2020). We estimate a p-value of 0.495 and are able to formally reject the existence of a discontinuity in the density function at the cutoff.



(a) High school entrance exam scores



(b) Gender

Figure 2: Test of Smoothness of Baseline Covariates

Notes: Sample includes students who entered high school from 2015 to 2017. Bins represent local averages over a 5 point score range. All figures are drawn using a linear fit on either side of the cutoff.

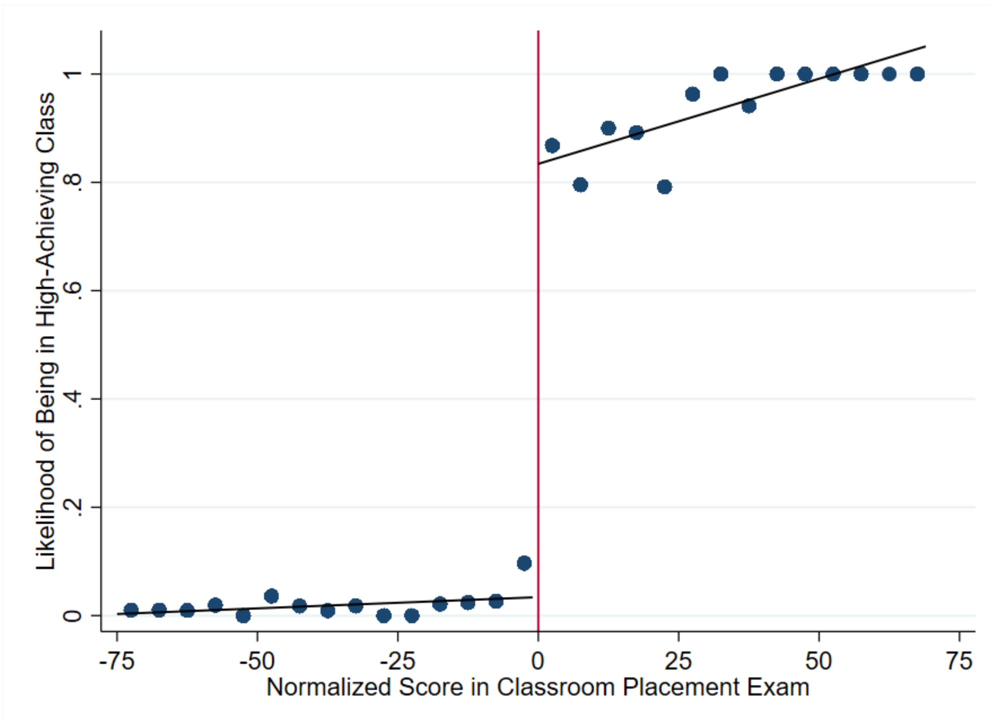
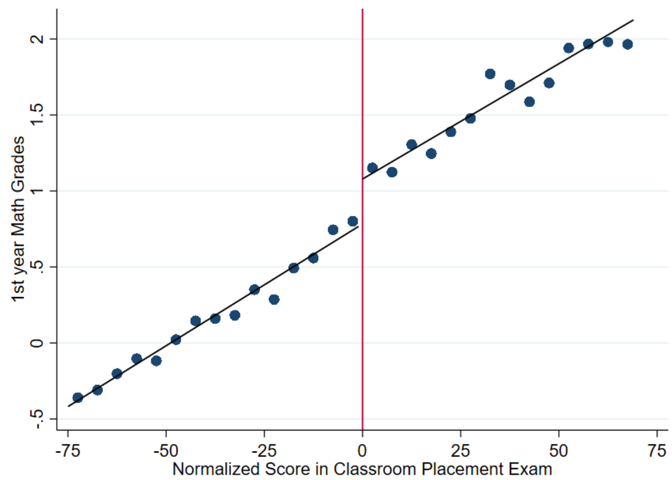
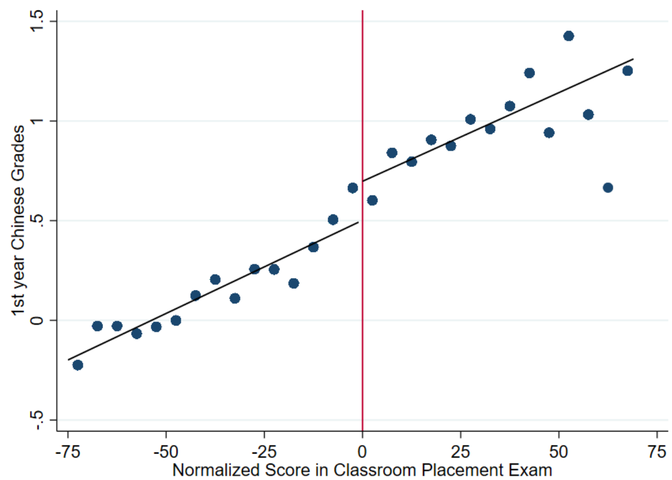


Figure 3: First Stage—Likelihood of Enrolling in a High-Achieving Classroom

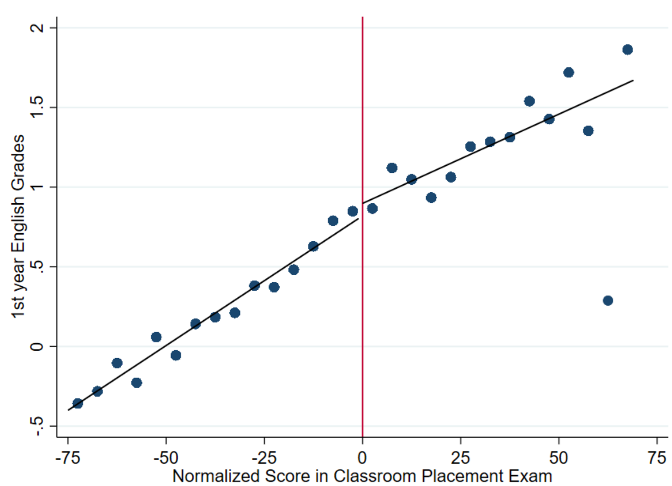
Notes: Sample includes students who entered high school from 2015 to 2017. Bins represent local averages over a 5 point score range. All figures are drawn using a linear fit on either side of the cutoff.



(a) 1st year Math grades



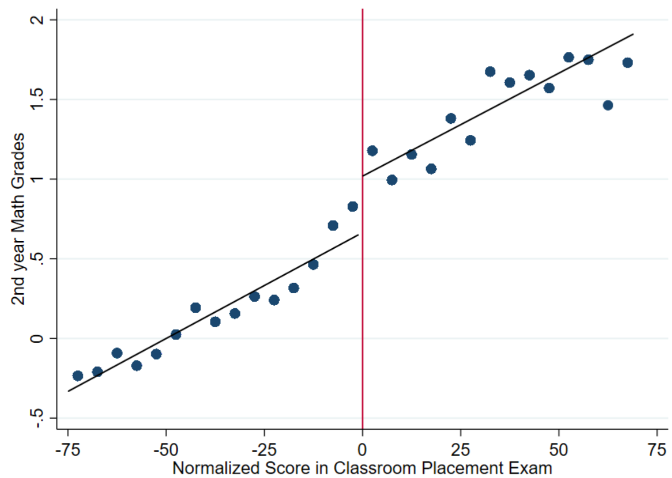
(b) 1st year Chinese grades



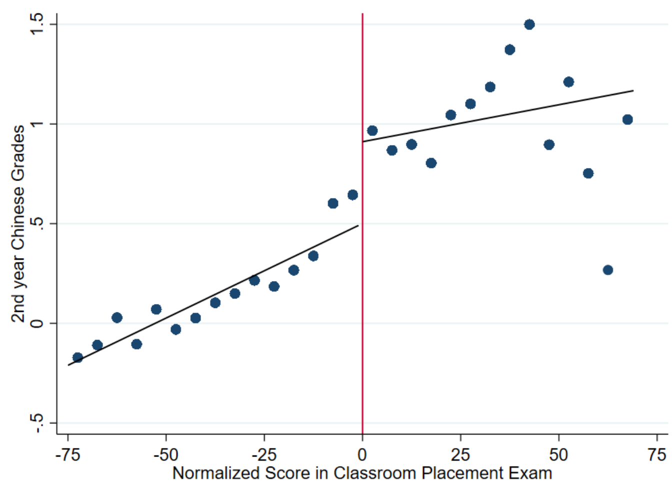
(c) 1st year English grades

Figure 4: First Year High School Grades

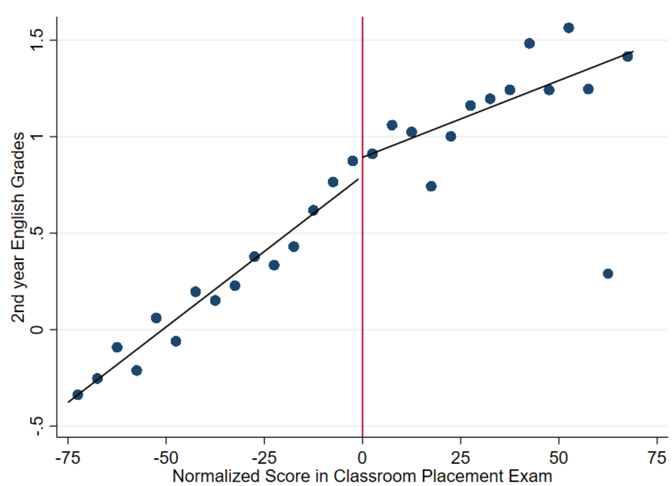
Notes: Sample includes students who entered high school from 2015 to 2017. Bins represent local averages over a 5 point score range. All figures are drawn using a linear fit on either side of the cutoff.



(a) 2nd year Math grades



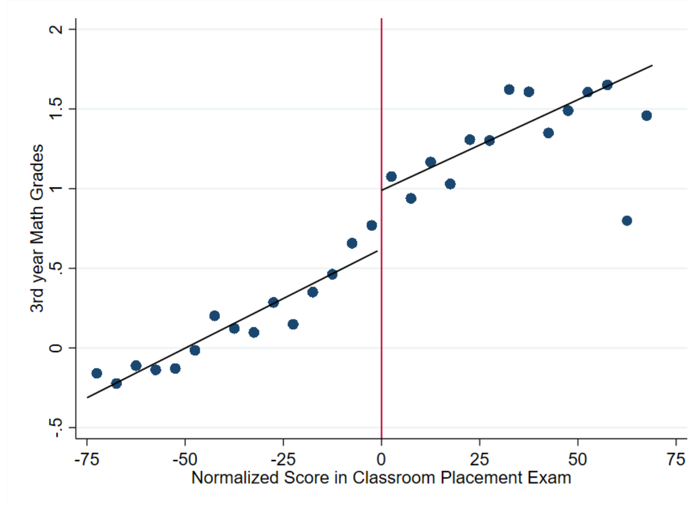
(b) 2nd year Chinese grades



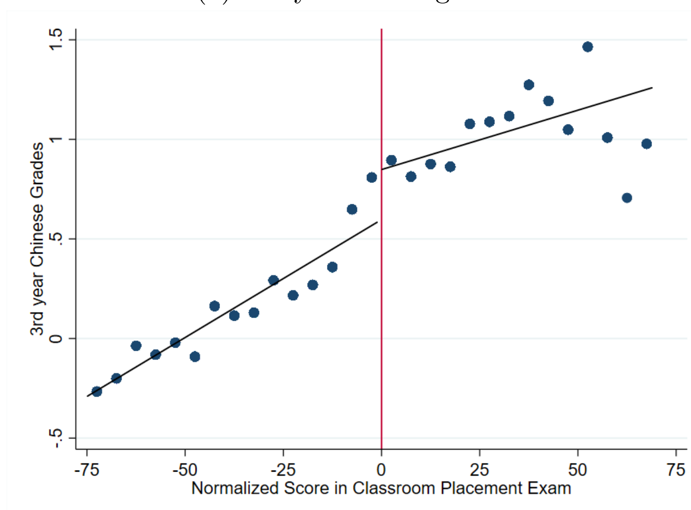
(c) 2nd year English grades

Figure 5: Second Year High School Grades

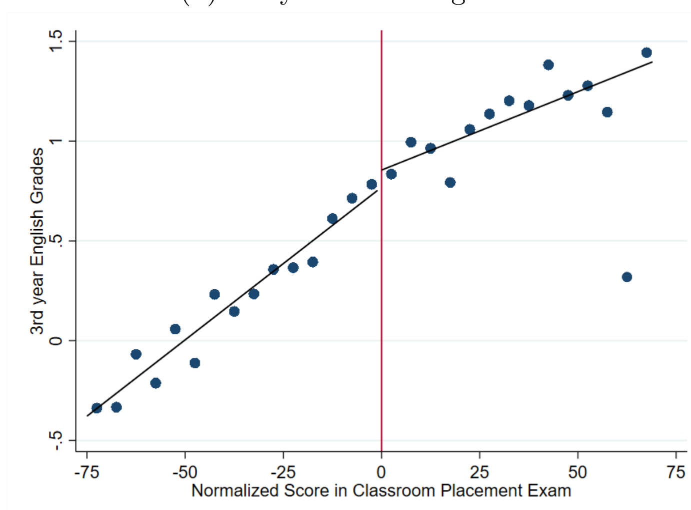
Notes: Sample includes students who entered high school from 2015 to 2017. Bins represent local averages over a 5 point score range. All figures are drawn using a linear fit on either side of the cutoff.



(a) 3rd year Math grades



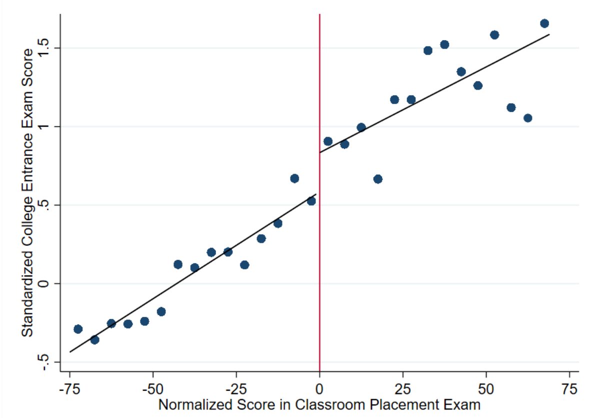
(b) 3rd year Chinese grades



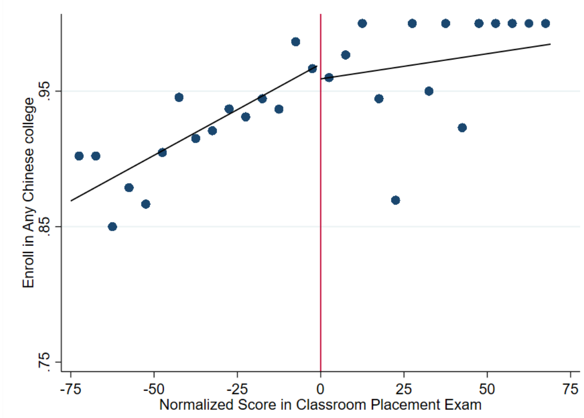
(c) 3rd year English grades

Figure 6: Third Year High School Grades

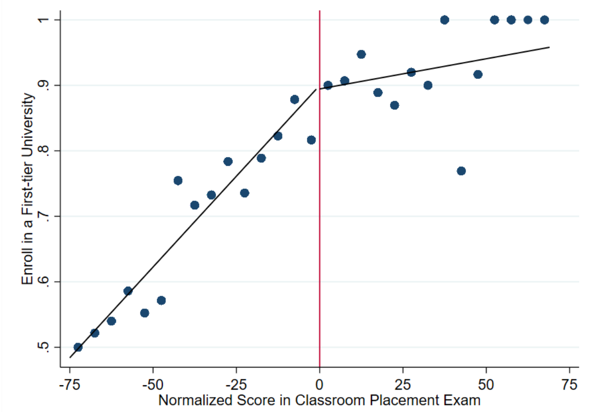
Notes: Sample includes students who entered high school from 2015 to 2017. Bins represent local averages over a 5 point score range. All figures are drawn using a linear fit on either side of the cutoff.



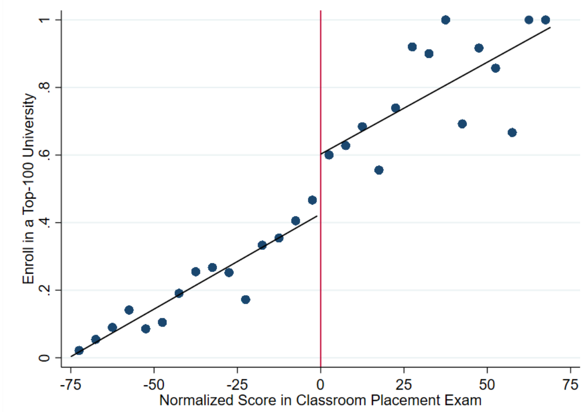
(a) Standardized college entrance exam scores



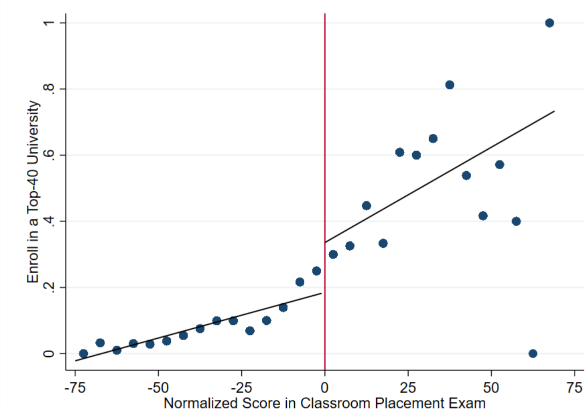
(b) Enroll in any Chinese university



(c) Enroll in a first-tier university



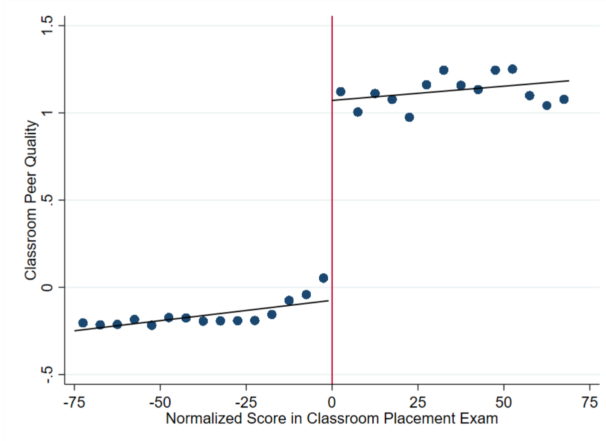
(d) Enroll in a top 100 university (Project 211)



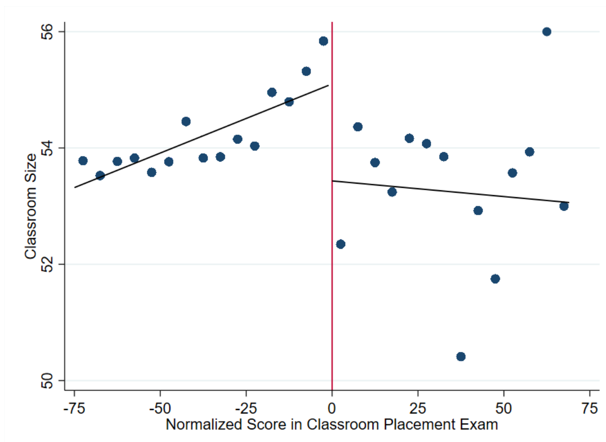
(e) Enroll in a top 40 university (Project 985)

Figure 7: Long-Run Educational Outcomes

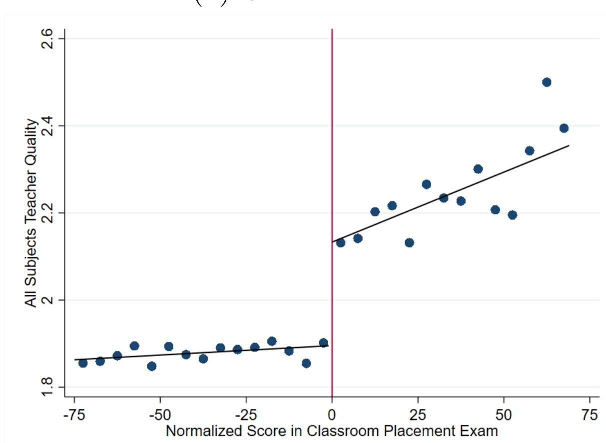
Notes: Sample includes students who entered high school from 2015 to 2017. Bins represent local averages over a 5 point score range. All figures are drawn using a linear fit on either side of the cutoff.



(a) Classroom peer quality



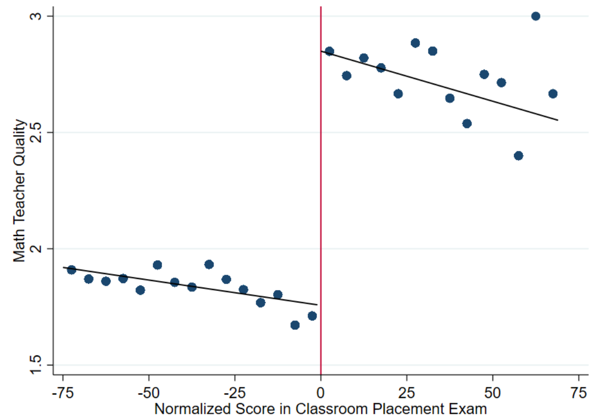
(b) Classroom size



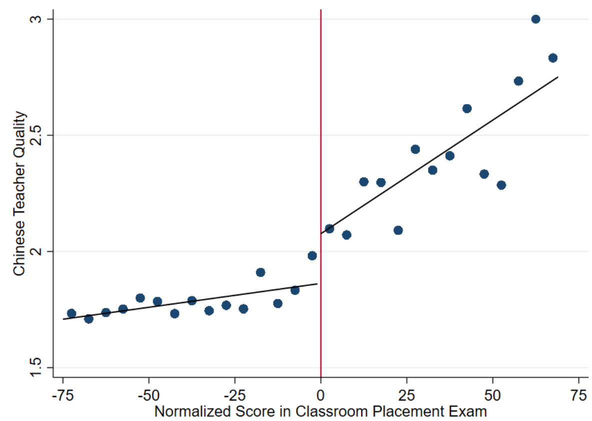
(c) Classroom teacher quality

Figure 8: Mechanisms

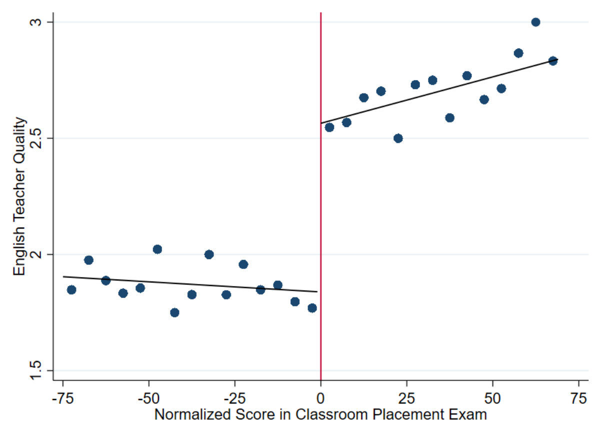
Notes: Sample includes students who entered high school from 2015 to 2017. All figures represent first-year tracking averages. Bins represent local averages over a 5 point score range. All figures are drawn using a linear fit on either side of the cutoff. Classroom teacher quality is based on a teacher's rank which is classified as 3=senior rank, 2=first rank and 1= second rank. Teacher ranks are not automatic and are generally based on teaching performance and publications.



(a) Math Teacher Quality



(b) Chinese Teacher Quality



(c) English Teacher Quality

Figure 9: Average Classroom Teacher Quality By Main Subjects

Notes: Sample includes students who entered high school from 2015 to 2017. All figures represent first-year tracking averages. Bins represent local averages over a 5 point score range. All figures are drawn using a linear fit on either side of the cutoff. Classroom teacher quality is based on a teacher's rank which is classified as 3=senior rank, 2=first rank and 1= second rank. Teacher ranks are not automatic and are generally based on teaching performance and publications.

B Tables

Table 1: Summary statistics

	Overall Sample (1)	Marginal Sample (2)
A) Student Characteristics		
Proportion Male	0.532	0.529
High School Entrance Exam Score	790.5	796.9
	(88.96)	(88.65)
Classroom Placement Exam Score (Running Variable)	413.4	429.9
	(56.49)	(39.25)
Proportion of students in high-achieving classroom	0.137	0.173
Year 1 High School Scores (Standardized)	0.016	0.404
Year 2 High School Scores (Standardized)	0.005	0.353
Year 3 High School Scores (Standardized)	-0.016	0.329
Proportion Selecting Science Track	0.871	0.898
College Entrance Exam Scores (Science Track)	510.77	522.10
	(64.64)	(61.63)
College Entrance Exam Scores (Arts Track)	527.46	541.48
	(54.84)	(54.94)
Proportion Not Sitting for College Entrance Exam	0.048	0.045
Proportion Enrolled in any Chinese University	0.900	0.925
Proportion Enrolled in Tier-1 University	0.644	0.720
Proportion Enrolled in Top-100 University	0.246	0.294
Proportion Enrolled in Top-40 University	0.116	0.144
Number of Students	2,273	1,788
B) Classroom-level Characteristics		
Class Size	58.66	58.67
	(7.07)	(6.97)
Teacher Salary Scale	22.16	22.31
	(7.90)	(7.98)
Teacher Experience (Years)	16.84	16.92
	(10.10)	(10.21)
Proportion of Top-Teachers	0.258	0.264
Number of Classrooms	43	43
Number of Top-Classrooms	6	6

Notes: Sample in Column (1) includes all students who first enrolled in high school in the academic years 2015 to 2017. The marginal sample in Column (2) contains all students scoring within 75 points on either sides of the classroom placement exam cutoff. High school test scores are standardized by year of entry (i.e. by cohort) for each grade. Classroom-level characteristics represent averages across all three years of high school.

Table 2: Baseline covariates balance tests

Outcome	Student is Male		High School Entrance Exam Scores	
	(1)	(2)	(3)	(4)
Estimated Discontinuity	-0.031	-0.019	-0.063	-0.037
	(0.085)	(0.080)	(0.083)	(0.084)
Observations	1,207	1,207	837	837
Bandwidth	CCT	CCT	CCT	CCT
Kernel	Triangular	Uniform	Triangular	Uniform

Notes: Sample includes students who entered high school from 2015 to 2017. All estimates are from local linear regressions using various bandwidths and kernel distributions. The number of observations vary by outcome since the CCT bandwidth selector predicts different bandwidths depending on outcome. All regressions include year fixed effects. High school entrance exam scores are standardized by year. Robust standard errors reported in parentheses. *** p <0.01 ** p <0.05 * p <0.1

Table 3: First stage—Enrollment in a high-achieving classroom

Outcome	Likelihood of Enrolling in High-achieving Classroom	
	(1)	(2)
Estimated Discontinuity	0.811***	0.794***
	(0.055)	(0.051)
With Controls	0.787***	0.775***
	(0.051)	(0.047)
Observations	1,174	1,174
Bandwidth	CCT	CCT
Kernel	Triangular	Uniform

Notes: Sample includes students who entered high school from 2015 to 2017. All estimates are from local linear regressions using various bandwidths and kernel distributions. The number of observations vary by outcome since the CCT bandwidth selector predicts different bandwidths depending on outcome. Controls include: gender, high school entrance exam scores and year fixed effects. Robust standard errors reported in parentheses. *** p <0.01 ** p <0.05 * p <0.1

Table 4: Average test scores during three years of high school

Outcome	Math Grades		Chinese Grades		English Grades	
	(1)	(2)	(3)	(4)	(5)	(6)
A) First Year						
Estimated Discontinuity	0.240**	0.274***	-0.018	0.072	-0.071	-0.018
	(0.103)	(0.090)	(0.128)	(0.114)	(0.125)	(0.113)
With Controls	0.231**	0.270***	-0.031	0.055	-0.095	-0.069
	(0.096)	(0.084)	(0.0116)	(0.104)	(0.111)	(0.101)
Observations	1,111	1,111	969	969	1,165	1,165
B) Second Year						
Estimated Discontinuity	0.270*	0.316**	0.209	0.278**	0.041	0.153
	(0.145)	(0.127)	(0.150)	(0.140)	(0.113)	(0.109)
With Controls	0.267*	0.312**	0.200	0.261**	0.010	0.120
	(0.140)	(0.124)	(0.136)	(0.129)	(0.107)	(0.102)
Observations	814	814	726	726	1,084	1,084
C) Third Year						
Estimated Discontinuity	0.246*	0.253*	0.131	0.173	0.060	0.090
	(0.143)	(0.130)	(0.140)	(0.135)	(0.114)	(0.103)
With Controls	0.242*	0.256**	0.120	0.160	0.043	0.065
	(0.143)	(0.125)	(0.136)	(0.132)	(0.108)	(0.098)
Observations	769	769	742	742	1,162	1,162
Bandwidth	CCT	CCT	CCT	CCT	CCT	CCT
Kernel	Triangular	Uniform	Triangular	Uniform	Triangular	Uniform

Notes: Sample includes students who entered high school from 2015 to 2017. All estimates are from local linear regressions using various bandwidths and kernel distributions. The number of observations vary by outcome since the CCT bandwidth selector predicts different bandwidths depending on outcome. Controls include: gender, high school entrance exam scores and year fixed effects. Test scores are standardized by subject-year in year 1 and subject-year-track in years 2 and 3. Robust standard errors reported in parentheses. *** p <0.01 ** p <0.05 * p <0.1

Table 5: Long run educational outcomes

Outcome	College Entrance Exam Scores		Enroll in Any University		Enroll in First-Tier University		Enroll in Top-100 University		Enroll in Top-40 University	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Estimated Discontinuity	0.278**	0.284**	-0.005	0.013	0.018	0.051	0.165*	0.186**	0.062	0.093
	(0.139)	(0.135)	(0.030)	(0.033)	(0.058)	(0.055)	(0.085)	(0.076)	(0.087)	(0.078)
With Controls	0.268**	0.278**	-0.010	-0.001	0.017	0.044	0.172**	0.189**	0.066	0.095
	(0.134)	(0.132)	(0.030)	(0.031)	(0.058)	(0.055)	(0.083)	(0.075)	(0.084)	(0.076)
IV Estimate (With Controls)	0.349*	0.354**	-0.014	-0.001	0.023	0.057	0.225**	0.241**	0.087	0.125
	(0.180)	(0.171)	(0.040)	(0.041)	(0.078)	(0.071)	(0.110)	(0.097)	(0.109)	(0.100)
Observations	1,231	1,231	926	926	815	815	1,146	1,146	973	973
Bandwidth	CCT	CCT	CCT	CCT	CCT	CCT	CCT	CCT	CCT	CCT
Kernel	Triangular	Uniform	Triangular	Uniform	Triangular	Uniform	Triangular	Uniform	Triangular	Uniform

Notes: Sample includes students who entered high school from 2015 to 2017. All estimates are from local linear regressions using various bandwidths and kernel distributions. The number of observations vary by outcome since the CCT bandwidth selector predicts different bandwidths depending on outcome. Controls include: gender, high school entrance exam scores and year fixed effects. College Entrance Exam scores are standardized by year and track. Robust standard errors reported in parentheses. *** p <0.01 ** p <0.05 * p <0.1

Table 6: Robustness check—Selection issues

Outcome	Selection Out of College Entrance Exam		Selection Into Science Track	
	(1)	(2)	(3)	(4)
Estimated Discontinuity	0.043	0.037	0.007	0.014
	(0.039)	(0.035)	(0.043)	(0.038)
Observations	873	873	997	997
Bandwidth	CCT	CCT	CCT	CCT
Kernel	Triangular	Uniform	Triangular	Uniform

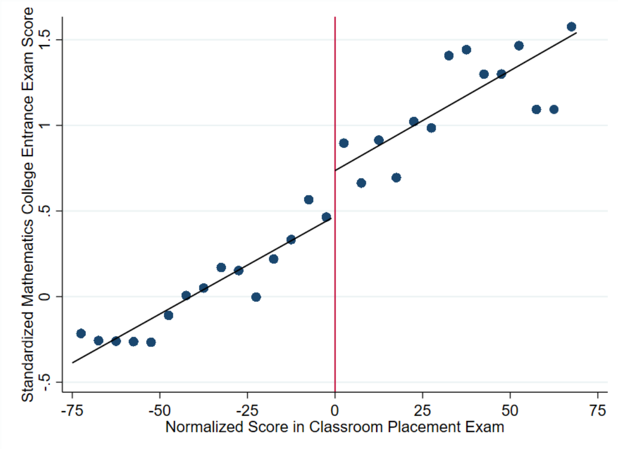
Notes: Sample includes students who entered high school from 2015 to 2017. All estimates are from local linear regressions using various bandwidths and kernel distributions. The number of observations vary by outcome since the CCT bandwidth selector predicts different bandwidths depending on outcome. All regressions include controls for gender, high school entrance exam scores and year fixed effects. Robust standard errors reported in parentheses. *** p <0.01 ** p <0.05 * p <0.1

Table 7: Mechanisms

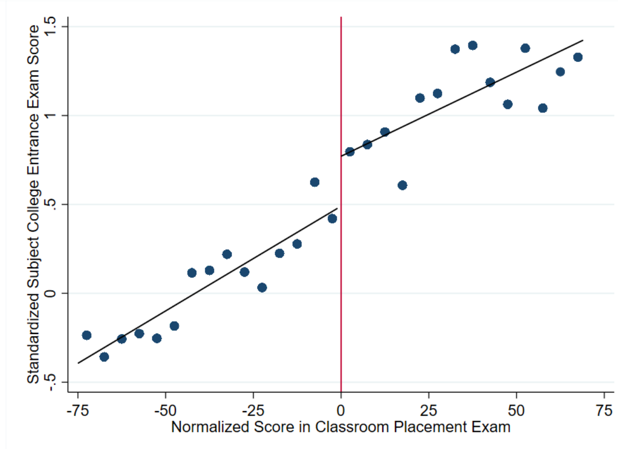
Outcome	Classroom Peer Quality		Classroom Size		Classroom Teacher Quality	
	(1)	(2)	(3)	(4)	(5)	(6)
Estimated Discontinuity	1.078***	1.078***	-3.37**	-2.94**	0.360***	0.368***
	(0.086)	(0.080)	(1.50)	(1.35)	(0.045)	(0.040)
With Controls	1.031***	1.048***	-3.92***	-3.28**	0.404***	0.409***
	(0.073)	(0.069)	(0.971)	(0.890)	(0.035)	(0.032)
IV Estimate (With Controls)	1.345***	1.372***	-5.19***	-4.31***	0.510***	0.512***
	(0.040)	(0.036)	(1.311)	(1.20)	(0.028)	(0.026)
Observations	835	835	737	737	1,325	1,325
Bandwidth	CCT	CCT	CCT	CCT	CCT	CCT
Kernel	Triangular	Uniform	Triangular	Uniform	Triangular	Uniform

Notes: Sample includes students who entered high school from 2015 to 2017. All estimates are from local linear regressions using various bandwidths and kernel distributions. The number of observations vary by outcome since the CCT bandwidth selector predicts different bandwidths depending on outcome. Controls include: gender, high school entrance exam scores and year fixed effects. All class-level mechanisms are estimated in the first year of tracking. Peer quality is standardized and based on students' performance in high school entrance exam. Teacher quality is standardized and based on based on teachers' rank which is classified as 3=senior rank, 2=first rank and 1= second rank. Teacher ranks are not automatic and are generally based on teaching performance and publications. Robust standard errors reported in parentheses. *** p <0.01 ** p <0.05 * p <0.1

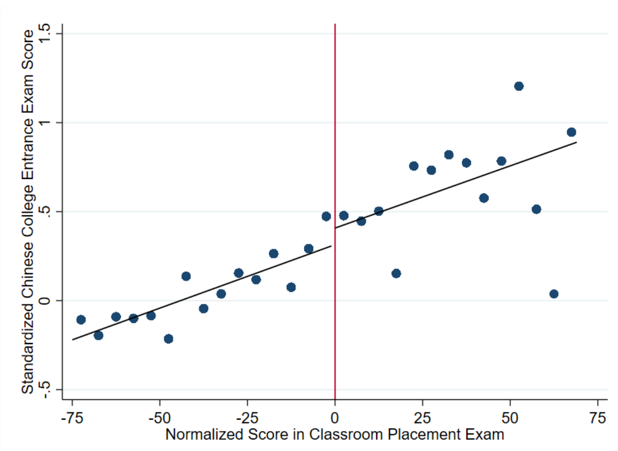
C Appendix Figures



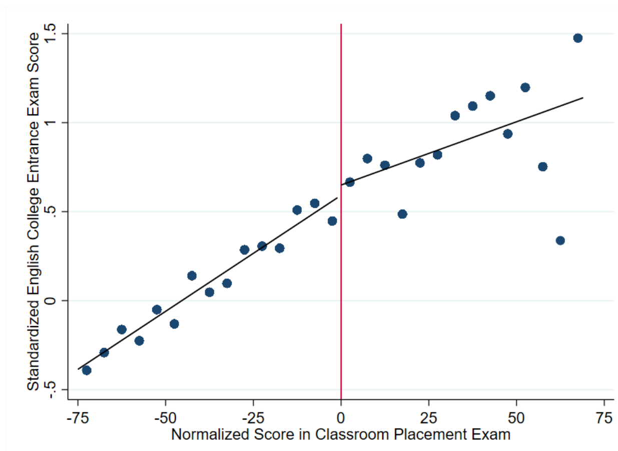
(a) Math college entrance exam score



(b) Main subjects college entrance exam score



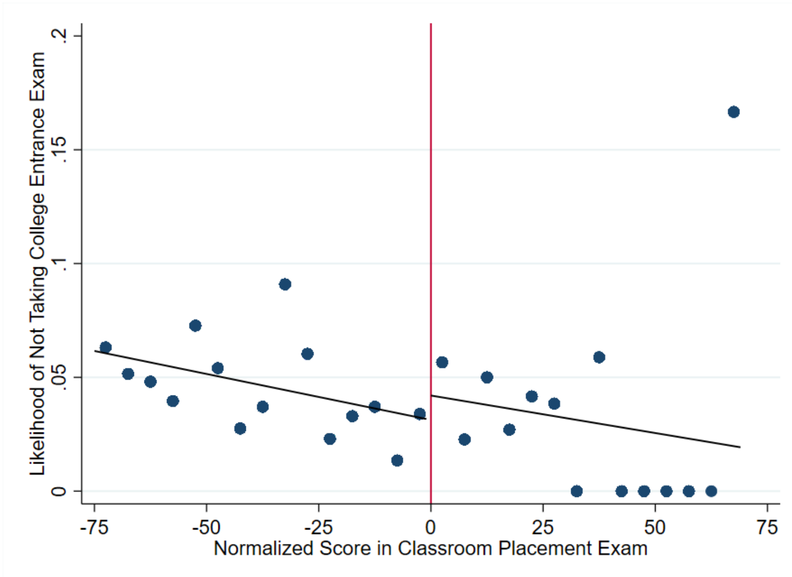
(c) Chinese college entrance exam score



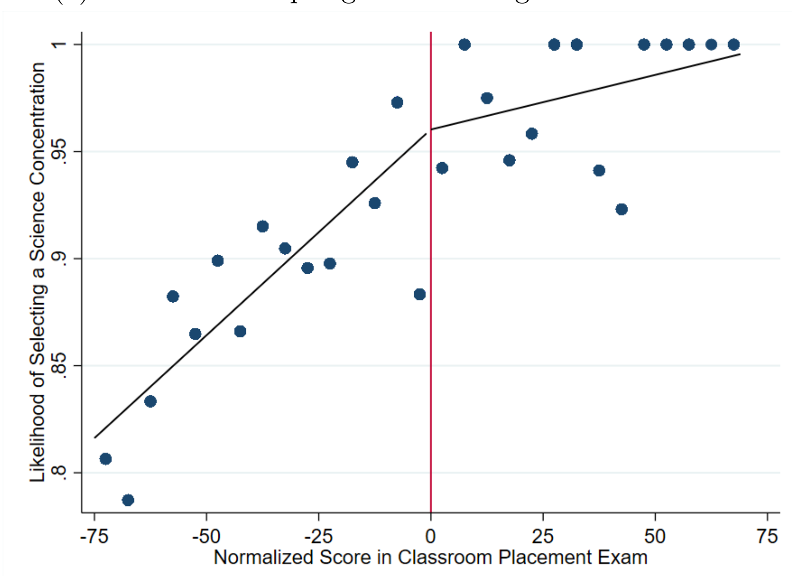
(d) English college entrance exam score

Figure A1: College Entrance Exam Scores by Subject

Notes: Sample includes students who entered high school from 2015 to 2017. Bins represent local averages over a 5 point score range. All figures are drawn using a linear fit on either side of the cutoff. Main subject scores are physics, chemistry, and biology for the science track and history, politics, and geography for the arts track.



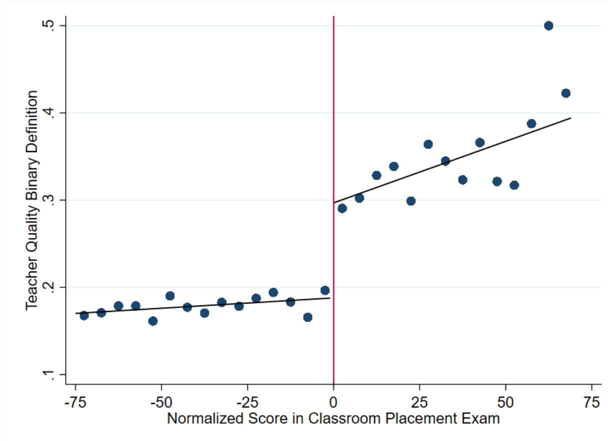
(a) Likelihood of opting out of College Entrance Exam



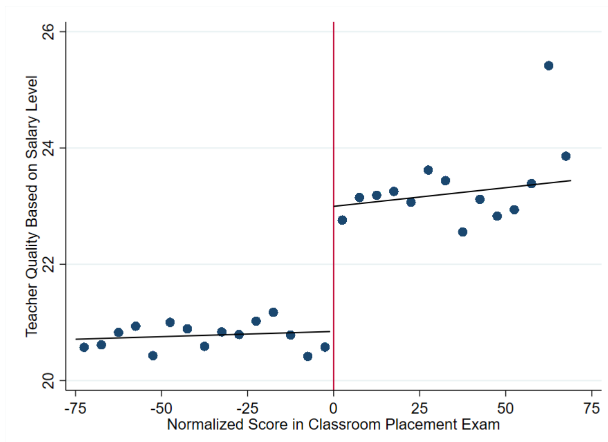
(b) Selection into science concentration in second year of high school

Figure A2: Robustness Check—Selection issues

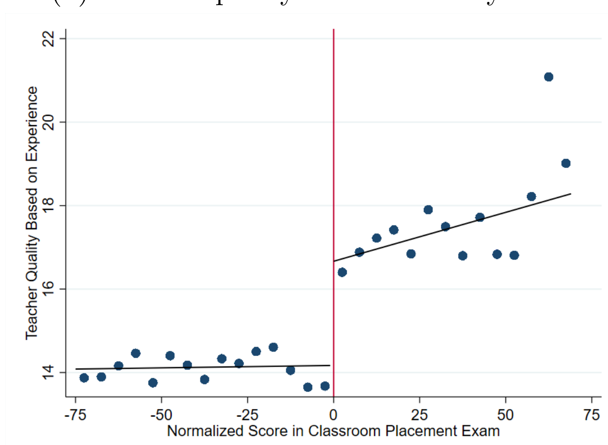
Notes: Sample includes students who entered high school from 2015 to 2017. Bins represent local averages over a 5 point score range. All figures are drawn using a linear fit on either side of the cutoff.



(a) Binary definition of teacher quality



(b) Teacher quality based on salary scale



(c) Teacher quality based on years of experience

Figure A3: Alternative Definitions of Teacher Quality

Notes: Sample includes students who entered high school from 2015 to 2017. All figures represent mechanisms from first year tracking. Bins represent local averages over a 5 point score range. All figures are drawn using a linear fit on either side of the cutoff. A teacher's rank is generally classified as 3=senior rank, 2=first rank and 1= second rank.Our binary definition of teacher quality defines senior rank teachers as top and first and second rank teachers as non-top. The teacher salary scale ranges from 9 to 40 with 40 being the highest.

D Appendix Tables

Table A1: Baseline covariates balance tests using different bandwidths

	(1)	(2)	(3)	(4)	(5)	(6)
Student is male	-0.048	-0.026	-0.048	-0.057	-0.053	-0.048
	(0.066)	(0.061)	(0.056)	(0.053)	(0.053)	(0.051)
High School Entrance Exam Scores	0.002	0.026	0.007	-0.001	0.008	0.039
	(0.069)	(0.076)	(0.065)	(0.073)	(0.066)	(0.074)
Observations	1,245	1,245	1,788	1,788	2,092	2,092
Bandwidth	50		75		100	
Kernel	Triangular	Uniform	Triangular	Uniform	Triangular	Uniform

Notes: Sample includes students who entered high school from 2015 to 2017. All estimates are from local linear regressions using various bandwidths and kernel distributions. All regressions include year fixed effects. High school entrance exam scores are standardized by year. Robust standard errors reported in parentheses. *** p <0.01 ** p <0.05 * p <0.1

Table A2: First stage—Likelihood of enrolling in a high-achieving classroom using different bandwidths

	(1)	(2)	(3)	(4)	(5)	(6)
First Stage						
Estimated Discontinuity	0.775***	0.786***	0.788***	0.798***	0.794***	0.800***
	(0.041)	(0.036)	(0.035)	(0.031)	(0.033)	(0.031)
With Controls	0.766***	0.781***	0.782***	0.794***	0.790***	0.796***
	(0.040)	(0.035)	(0.034)	(0.031)	(0.032)	(0.030)
Observations	1,245	1,245	1,774	1,774	2,083	2,083
Bandwidth	50		75		100	
Kernel	Triangular	Uniform	Triangular	Uniform	Triangular	Uniform

Notes: Sample includes students who entered high school from 2015 to 2017. All estimates are from local linear regressions using various bandwidths and kernel distributions. Controls include: gender, high school entrance exam scores and year fixed effects. Robust standard errors reported in parentheses. *** p <0.01 ** p <0.05 * p <0.1

Table A3: Average test scores during 1st year of high school using different bandwidths

	(1)	(2)	(3)	(4)	(5)	(6)
A) Math Grades						
Estimated Discontinuity	0.204***	0.258***	0.251***	0.284***	0.275***	0.322***
	(0.078)	(0.070)	(0.066)	(0.060)	(0.061)	(0.058)
With Controls	0.209***	0.258***	0.254***	0.286***	0.278***	0.319***
	(0.074)	(0.067)	(0.062)	(0.058)	(0.059)	(0.056)
B) Chinese Grades						
Estimated Discontinuity	0.051	0.138	0.131	0.191**	0.161**	0.204***
	(0.101)	(0.092)	(0.085)	(0.078)	(0.080)	(0.076)
With Controls	0.036	0.127	0.108	0.161**	0.135*	0.177**
	(0.093)	(0.086)	(0.079)	(0.074)	(0.074)	(0.072)
C) English Grades						
Estimated Discontinuity	-0.005	0.001	0.031	0.067	0.065	0.102
	(0.089)	(0.080)	(0.075)	(0.069)	(0.070)	(0.066)
With Controls	-0.041	-0.020	-0.003	0.027	0.027	0.073
	(0.083)	(0.075)	(0.069)	(0.065)	(0.064)	(0.062)
Observations	1,225	1,225	1,758	1,758	2,052	2,052
Bandwidth	50		75		100	
Kernel	Triangular	Uniform	Triangular	Uniform	Triangular	Uniform

Notes: Sample includes students who entered high school from 2015 to 2017. All estimates are from local linear regressions using various bandwidths and kernel distributions. Controls include: gender, high school entrance exam scores and year fixed effects. Test scores are standardized by subject-year. Robust standard errors reported in parentheses. *** p <0.01 ** p <0.05 * p <0.1

Table A4: Average test scores during 2nd year of high school using different bandwidths

	(1)	(2)	(3)	(4)	(5)	(6)
A) Math Grades						
Estimated Discontinuity	0.209**	0.274***	0.282***	0.340***	0.320***	0.349***
	(0.098)	(0.087)	(0.082)	(0.075)	(0.076)	(0.072)
With Controls	0.210**	0.270***	0.277***	0.333***	0.314***	0.341***
	(0.095)	(0.085)	(0.080)	(0.074)	(0.075)	(0.071)
B) Chinese Grades						
Estimated Discontinuity	0.197*	0.228**	0.303***	0.402***	0.363***	0.432***
	(0.107)	(0.100)	(0.091)	(0.086)	(0.086)	(0.083)
With Controls	0.177*	0.216**	0.279***	0.373***	0.336***	0.405***
	(0.100)	(0.094)	(0.085)	(0.081)	(0.080)	(0.078)
C) English Grades						
Estimated Discontinuity	0.016	0.011	0.046	0.084	0.085	0.138*
	(0.094)	(0.086)	(0.079)	(0.075)	(0.074)	(0.071)
With Controls	-0.020	-0.010	0.011	0.048	0.050	0.104
	(0.089)	(0.081)	(0.075)	(0.072)	(0.071)	(0.069)
Observations	1,218	1,218	1,745	1,745	2,042	2,042
Bandwidth	50		75		100	
Kernel	Triangular	Uniform	Triangular	Uniform	Triangular	Uniform

Notes: Sample includes students who entered high school from 2015 to 2017. All estimates are from local linear regressions using various bandwidths and kernel distributions. Controls include: gender, high school entrance exam scores and year fixed effects. Test scores are standardized by subject-year and track. Robust standard errors reported in parentheses. *** p <0.01 ** p <0.05 * p <0.1

Table A5: Average test scores during 3rd year of high school using different bandwidths

	(1)	(2)	(3)	(4)	(5)	(6)
A) Math Grades						
Estimated Discontinuity	0.206**	0.283***	0.283***	0.363***	0.330***	0.378***
	(0.099)	(0.088)	(0.083)	(0.077)	(0.078)	(0.074)
With Controls	0.217**	0.280***	0.286***	0.363***	0.333***	0.377***
	(0.095)	(0.085)	(0.080)	(0.075)	(0.075)	(0.073)
B) Chinese Grades						
Estimated Discontinuity	0.088	0.142	0.172**	0.248***	0.228***	0.299***
	(0.101)	(0.094)	(0.088)	(0.084)	(0.083)	(0.081)
With Controls	0.069	0.127	0.149*	0.220***	0.202**	0.273***
	(0.098)	(0.092)	(0.085)	(0.082)	(0.081)	(0.079)
C) English Grades						
Estimated Discontinuity	0.029	0.024	0.057	0.087	0.090	0.144**
	(0.094)	(0.084)	(0.079)	(0.073)	(0.073)	(0.069)
With Controls	0.003	0.007	0.028	0.051	0.057	0.107
	(0.091)	(0.080)	(0.076)	(0.070)	(0.070)	(0.067)
Observations	1,197	1,197	1,711	1,711	2,005	2,005
Bandwidth	50		75		100	
Kernel	Triangular	Uniform	Triangular	Uniform	Triangular	Uniform

Notes: Sample includes students who entered high school from 2015 to 2017. All estimates are from local linear regressions using various bandwidths and kernel distributions. Controls include: gender, high school entrance exam scores and year fixed effects. Test scores are standardized by subject-year and track. Robust standard errors reported in parentheses. *** p <0.01 ** p <0.05 * p <0.1

Table A6: College entrance exam scores by subject

Outcome	Math Score		Main Subject Scores		Chinese Score		English Score	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Estimated Discontinuity	0.300**	0.350***	0.306**	0.345**	0.056	0.073	0.074	0.135
	(0.128)	(0.135)	(0.152)	(0.145)	(0.148)	(0.161)	(0.136)	(0.131)
With Controls	0.307**	0.358***	0.305**	0.341**	0.041	0.057	0.033	0.098
	(0.120)	(0.130)	(0.146)	(0.141)	(0.147)	(0.160)	(0.130)	(0.126)
IV Estimate (With Controls)	0.399**	0.460***	0.396**	0.437**	0.054	0.072	0.044	0.126
	(0.161)	(0.169)	(0.195)	(0.185)	(0.191)	(0.204)	(0.169)	(0.163)
Observations	1,225	1,225	1,248	1,248	1,253	1,253	1,191	1,191
Bandwidth	CCT	CCT	CCT	CCT	CCT	CCT	CCT	CCT
Kernel	Triangular	Uniform	Triangular	Uniform	Triangular	Uniform	Triangular	Uniform

Notes: Sample includes students who entered high school from 2015 to 2017. All estimates are from local linear regressions using various bandwidths and kernel distributions. The number of observations vary by outcome since the CCT bandwidth selector predicts different bandwidths depending on outcome. Controls include: gender, high school entrance exam scores and year fixed effects. All scores are standardized by year and track. Main subject scores are physics, chemistry, and biology for the science track and history, politics, and geography for the arts track. Robust standard errors reported in parentheses. *** p <0.01 ** p <0.05 * p <0.1

Table A7: Long run educational outcomes using different bandwidths

	(1)	(2)	(3)	(4)	(5)	(6)
A) College Entry Exam Scores						
Estimated Discontinuity	0.170	0.168*	0.186**	0.235***	0.216**	0.259***
	(0.109)	(0.096)	(0.091)	(0.085)	(0.085)	(0.082)
With Controls	0.162	0.164*	0.178**	0.228***	0.210**	0.252***
	(0.106)	(0.094)	(0.089)	(0.084)	(0.083)	(0.081)
IV Estimate (With Controls)	0.215	0.212*	0.231**	0.290***	0.269**	0.319***
	(0.144)	(0.123)	(0.117)	(0.108)	(0.108)	(0.103)
B) Enroll in any Chinese University						
Estimated Discontinuity	-0.007	-0.007	-0.014	-0.013	-0.014	-0.022
	(0.024)	(0.023)	(0.021)	(0.020)	(0.020)	(0.020)
With Controls	-0.004	-0.000	-0.009	-0.008	-0.009	-0.015
	(0.024)	(0.022)	(0.020)	(0.020)	(0.019)	(0.020)
IV Estimate (With Controls)	-0.006	-0.000	-0.012	-0.010	-0.012	-0.020
	(0.031)	(0.029)	(0.026)	(0.025)	(0.025)	(0.025)
C) Enroll in First-Tier University						
Estimated Discontinuity	0.035	0.021	0.006	-0.006	-0.004	-0.019
	(0.044)	(0.041)	(0.037)	(0.035)	(0.035)	(0.033)
With Controls	0.040	0.031	0.014	0.002	0.003	-0.010
	(0.044)	(0.041)	(0.037)	(0.035)	(0.035)	(0.033)
IV Estimate (With Controls)	0.053	0.040	0.018	0.003	0.004	-0.013
	(0.059)	(0.053)	(0.048)	(0.044)	(0.044)	(0.042)
D) Enroll in Top 100 University						
Estimated Discontinuity	0.282**	0.325**	0.150***	0.182***	0.176***	0.221***
	(0.067)	(0.059)	(0.056)	(0.050)	(0.052)	(0.049)
With Controls	0.125*	0.132**	0.157***	0.186***	0.182***	0.224***
	(0.066)	(0.058)	(0.055)	(0.050)	(0.052)	(0.048)
IV Estimate (With Controls)	0.166*	0.171**	0.203***	0.237***	0.233***	0.284***
	(0.088)	(0.076)	(0.073)	(0.064)	(0.067)	(0.062)
E) Enroll in Top 40 University						
Estimated Discontinuity	0.047	0.092*	0.108**	0.152***	0.134***	0.172***
	(0.062)	(0.056)	(0.053)	(0.048)	(0.049)	(0.046)
With Controls	0.048	0.092*	0.108**	0.153***	0.135***	0.173***
	(0.061)	(0.055)	(0.052)	(0.047)	(0.049)	(0.046)
IV Estimate (With Controls)	0.064	0.119*	0.140**	0.194***	0.172***	0.219***
	(0.081)	(0.071)	(0.067)	(0.060)	(0.063)	(0.059)
Observations	1,224	1,224	1,189	1,189	1,216	1,216
Bandwidth	50		75		100	
Kernel	Triangular	Uniform	Triangular	Uniform	Triangular	Uniform

Notes: Sample includes students who entered high school from 2015 to 2017. All estimates are from local linear regressions using various bandwidths and kernel distributions. Controls include: gender, high school entrance exam scores and year fixed effects. College Entrance Exam scores are standardized by year and track. Robust standard errors reported in parentheses. *** p <0.01 ** p <0.05 * p <0.1

Table A8: Robustness check—Selection issues using different bandwidths

	(1)	(2)	(3)	(4)	(5)	(6)
Selection Out Of College Entrance Exam	0.026	0.020	0.018	0.010	0.011	0.003
	(0.027)	(0.023)	(0.023)	(0.022)	(0.021)	(0.021)
Selection Into Science Concentration	0.027	0.024	0.018	0.002	0.008	0.005
	(0.030)	(0.027)	(0.025)	(0.023)	(0.022)	(0.021)
Observations	1,224	1,224	1,756	1,756	2,051	2,051
Bandwidth	50		75		100	
Kernel	Triangular	Uniform	Triangular	Uniform	Triangular	Uniform

Notes: Sample includes students who entered high school from 2015 to 2017. All estimates are from local linear regressions using various bandwidths and kernel distributions. All regressions include controls for gender, high school entrance exam scores and year fixed effects. Robust standard errors reported in parentheses. *** p <0.01 ** p <0.05 * p <0.1

Table A9: Mechanisms using different bandwidths

	(1)	(2)	(3)	(4)	(5)	(6)
A) Class Peer Quality						
Estimated Discontinuity	1.063***	1.091***	1.107***	1.146***	1.134***	1.161***
	(0.063)	(0.056)	(0.054)	(0.048)	(0.050)	(0.047)
With Controls	1.045***	1.083***	1.094***	1.138***	1.123***	1.155***
	(0.057)	(0.051)	(0.049)	(0.046)	(0.047)	(0.045)
IV Estimate (With Controls)	1.365***	1.386***	1.399***	1.432***	1.423***	1.450***
	(0.032)	(0.028)	(0.026)	(0.023)	(0.023)	(0.021)
B) Class Size						
Estimated Discontinuity	-2.409**	-1.676*	-1.976**	-1.668**	-1.751**	-1.526*
	(1.016)	(0.926)	(0.885)	(0.814)	(0.849)	(0.811)
With Controls	-2.790***	-2.035***	-2.220***	-1.796***	-1.914***	-1.583***
	(0.716)	(0.660)	(0.624)	(0.585)	(0.600)	(0.585)
IV Estimate (With Controls)	-3.643***	-2.605***	-2.839***	-2.261***	-2.425***	-1.988***
	(0.946)	(0.851)	(0.805)	(0.741)	(0.765)	(0.738)
C) Class Teacher Quality						
Estimated Discontinuity	0.385***	0.396***	0.384***	0.377***	0.380***	0.372***
	(0.041)	(0.036)	(0.033)	(0.030)	(0.030)	(0.029)
With Controls	0.402***	0.405***	0.403***	0.399***	0.402***	0.401***
	(0.030)	(0.027)	(0.025)	(0.023)	(0.023)	(0.021)
IV Estimate (With Controls)	0.519***	0.515***	0.513***	0.501***	0.506***	0.502***
	(0.028)	(0.025)	(0.022)	(0.020)	(0.020)	(0.019)
Observations	1,245	1,245	1,788	1,788	2,092	2,092
Bandwidth	50		75		100	
Kernel	Triangular	Uniform	Triangular	Uniform	Triangular	Uniform

Notes: Sample includes students who entered high school from 2015 to 2017. All estimates are from local linear regressions using various bandwidths and kernel distributions. Controls include: gender, high school entrance exam scores and year fixed effects. All class-level mechanisms are estimated in the first year of tracking. Peer quality is standardized and based on students' performance in high school entrance exam. Teacher quality is standardized and based on based on teachers' rank which is classified as 3=senior rank, 2=first rank and 1= second rank. Teacher ranks are not automatic and are generally based on teaching performance and publications. Robust standard errors reported in parentheses. Robust standard errors reported in parentheses. *** p <0.01 ** p <0.05 * p <0.1

Table A10: Teacher quality by subject

Outcome	Mathematics Teacher Quality		Chinese Teacher Quality		English Teacher Quality	
	(1)	(2)	(3)	(4)	(5)	(6)
Estimated Discontinuity	1.937***	1.809***	0.203*	0.240**	0.886***	0.879***
	(0.139)	(0.123)	(0.116)	(0.111)	(0.125)	(0.114)
With Controls	1.943***	1.814***	0.229**	0.259**	0.974***	0.986***
	(0.138)	(0.124)	(0.113)	(0.109)	(0.103)	(0.101)
IV Estimate (With Controls)	2.467***	2.344***	0.278**	0.310**	1.255***	1.263***
	(0.172)	(0.154)	(0.132)	(0.124)	(0.114)	(0.112)
Observations	801	801	1,100	1,100	1,147	1,147
Bandwidth	CCT	CCT	CCT	CCT	CCT	CCT
Kernel	Triangular	Uniform	Triangular	Uniform	Triangular	Uniform

Notes: Sample includes students who entered high school from 2015 to 2017. All estimates are from local linear regressions using various bandwidths and kernel distributions. The number of observations vary by outcome since the CCT bandwidth selector predicts different bandwidths depending on outcome. Controls include: gender and year fixed effects. Teacher quality is estimated in the first year of tracking. Teacher quality is standardized and based on based on teachers' rank which is classified as 3=senior rank, 2=first rank and 1= second rank. Teacher ranks are not automatic and are generally based on teaching performance and publications. Robust standard errors reported in parentheses. *** p <0.01 ** p <0.05 * p <0.1