Gender Differences in Reactions to Failure in High-Stakes Competition: Evidence from the National College Entrance Exam Retakes[*]

Le Kang

China Institute for Educational Finance Research, Peking University

lekang@pku.edu.cn

Ziteng Lei

Department of Economics, University of California, Santa Barbara

ziteng_lei@umail.ucsb.edu

Yang Song

Department of Economics, Colgate University

ysong@colgate.edu

Peng Zhang

School of Management and Economics, The Chinese University of Hong Kong, Shenzhen

zhangpeng@cuhk.edu.cn

This version: August 2021

Abstract

We document gender differences in reactions to failure in the National College Entrance Exam, an extremely high-stakes exam that solely determines college admission outcomes for almost all teenagers in China. Using unique administrative data in Ningxia Province and a regression-discontinuity design, we find that students who score just below the tier-2 university cutoff have an eight percentage point higher probability of retaking the exam in the next year, and that retaking improves exam performance substantially. However, the increase in retake probability when confronting the failure of scoring just below the cutoff is more pronounced for men than for women (11 percentage points vs. 5.5 percentage points). The gender disparity in the tendency to retake has important implications for exam performance, college enrollment, and labor market outcomes.

Keywords: Education, gender, competition, retake, National College Entrance Exam, China

JEL Classification: I20, I21, I23, I24, J16, J24

---

[*] We thank Ying Bai, Serena Canaan, Clément de Chaisemartin, Naijia Guo, Peter Kuhn, Shelly Lundberg, Pierre Mouganie, Gonzalo Vazquez-Bare, Huihua Xie, and seminar and conference participants at UCSB Applied Research Group, Jinan University, and CUHK for helpful comments and suggestions.

1 Introduction

Gender disparities in educational outcomes and labor market outcomes have attracted increasing attention. Previous studies have documented that gender differences in non-cognitive traits and attitudes, such as willingness to compete, pressure tolerance, risk aversion, and confidence, may explain an important part of the gender gaps in educational choices and labor market outcomes (see a review article by Delaney and Devereux, 2021). However, less is known about the gender difference in reactions to failure and its mechanisms and implications, especially in settings of high-stakes competitions. As people confront competitions throughout their career for college admission, jobs and promotions, failures and setbacks in these competitions are not uncommon for most people. Different responses to failure, such as whether to try again in subsequent competitions or give up, may lead to very different educational achievements and career paths. Therefore, understanding the gender differences in responses to failure is crucial for understanding gender gaps in educational and labor market outcomes.

To the best of our knowledge, this is the first study that documents gender differences in responses to failure in an admission-relevant exam for colleges and for a broad group of individuals. Specifically, we study how men and women respond differently to failures in the National College Entrance Exam (hereinafter referred to as NCEE), an annual exam that solely determines the admission of almost all students into higher education institutions in China. The setting is important for at least two reasons. First, many countries use high-stakes standardized tests to rank students for college admission, and retaking such exams when confronting failures are not uncommon. Studying gender disparities in response to failures in these admission-relevant exams and the related consequences is helpful to understand gender gaps in college enrollment and labor market outcomes. Second, since almost everyone needs to take the NCEE to get into colleges in China, our setting alleviates concerns over sample selection in the sense that individuals who do not like competition may choose not to participate in the competitions in the first place.

Estimating gender differences in responses to failure in the NCEE is challenging, as failures are typically subjective and not randomly assigned. To overcome this challenge, we exploit a unique feature of the NCEE, which is the exogenously determined cutoff for different tiers of universities. The over 2,000 universities in China are classified into four tiers, with NCEE score cutoffs determining the eligibility of application for universities in each tier.[1] We show evidence that these cutoffs are exogenously determined, and students cannot self-select around the cutoff. Around 10 million students take the NCEE to compete for admissions

[1] The higher education institutions in China are classified into tier-1 key universities, tier-2 regular universities, tier-3 universities, and tertiary technical colleges, by the central government. Only students with the NCEE scores above the tier cutoff can apply for universities in that specific tier. See Section 2 for more discussions.

to the highly selective universities each year, with only around 25% of students receiving scores that make them eligible to apply for high-quality universities in the top two tiers.

Our empirical strategy thus is to use the gender differences in the discontinuity in retake probability around the tier-2 cutoff to causally identify gender differences in responses to the arrival of a plausibly exogenous failure.[2] To do so, we obtain a unique dataset that covers the universe of NCEE takers in Ningxia Province during 2002-2010. Before we focus on gender differences, we first show that the tier-2 cutoff indeed generates a large discontinuity in the probability of retaking the NCEE regardless of gender. Specifically, students who score just below the tier-2 cutoff, a signal of entering good universities and educational success, have an eight percentage point higher probability of retaking the NCEE in the next year, almost doubling that of those who score just above the cutoff. In addition, we show that retaking the NCEE generates large returns in terms of exam performance and educational success, since it increases the test scores by 0.47 standard deviations and the relative ranking among competitors by 11 percentage points. These results indicate that the response to failure, specifically whether choosing to retake the exam or not, has crucial consequences for college admission and possibly future labor market prospects.

We then focus on gender differences in reactions to the failure of scoring below the tier-2 cutoff. We find consistent evidence that the cutoff-induced retakes from the regression discontinuity design, which reflect the desire to participate in the competition again inspired by the failure of scoring below the cutoff, are much more pronounced for men than for women. Specifically, the increase in retake probability when falling just below the tier-2 cutoff for males is twice as large as for females (11 vs. 5.5 percentage points, respectively), and the gender differences are statistically significant and robust across various specifications.

We detect several important mechanisms that can help explain why women are less likely to retake the NCEE than men when scoring just below the cutoff. We start by testing whether the returns to retake differ across gender. Surprisingly, we find that the causal returns to retake in terms of exam outcomes for women are similar to or sometimes even higher than those for men. In addition, such gender differences in returns cannot be explained by students rationally self-selecting into retake based on returns. These results suggest that gender differences in returns to retake are unlikely to explain the gender differences in the propensity to retake.

[2] We focus on the tier-2 cutoff because for students in Ningxia, admission into a tier-2 university is generally regarded as an educational success compared with tier-3 universities or technical colleges. By contrast, falling below the tier-1 cutoff, which indicates that the student is still eligible for admission into tier-2 universities, is much less viewed as a failure in the NCEE. Consequently, the decline in retake probability at the cutoff is dramatic for the tier-2 cutoff, but much less pronounced for the tier-1 cutoff. See Section 3 for more discussions.

By contrast, our results suggest such gender differences may be explained by gender differences in non-cognitive traits, such as causal attribution and confidence. For example, the psychological literature suggests that men tend to attribute success to internal factors such as talent, and failure to external factors such as luck, whereas women tend to do the opposite (Dweck et al., 1978; Ryckman and Peckham, 1987; Beyer, 1998). Females who score below the cutoff may be more likely to attribute the failure to own ability and less confident about the prospect of the retakes, and thus be less motivated to retake than males. We find that the gender differences are much smaller for repeated takers, who have more experience and likely better judgment on their own ability, than for first-time takers, which is consistent with this explanation.

In addition, we find that the gender differences are large and of similar magnitude for individuals from urban and rural households, of different ethnicity, from high-quality and low-quality high schools, from rich and poor counties, and from places with high and low levels of sex ratio. These results show that the gender differences in reactions to failure are not driven by certain groups, but are pronounced for all types of individuals. They also suggest that gender differences in benefits and costs of retake, as well as in social norms and family support, are unlikely to fully explain our results.

Our paper contributes to three strands of the literature. First, we contribute to the broad literature on gender differences in educational choices and competitions (Niederle and Vesterlund, 2007; Buser et al., 2014; Flory et al., 2015; Berlin and Dargnies, 2016; Buser et al., 2017; Reuben et al., 2017; Astorne-Figari and Speer, 2019; Cai et al., 2019)[3], and specifically on the growing literature that focuses on gender differences in the dynamic evolution of willingness to compete in response to winning and losing (Ellison and Swanson, 2018; Buser and Yuan, 2019; Landaud and Maurin, 2020; Wasserman, 2020; Fang et al., 2021). These studies have documented that when confronting failures in competitions, women are less likely to choose competition again than men in lab experiments and in low-stakes high school math competitions in the Netherlands and the U.S. (Ellison and Swanson, 2018; Buser and Yuan, 2019), in low-stakes Rubik's Cube competitions (Fang et al., 2021), in the entrance exam of highly selected elite science graduate programs in France (Landaud and Maurin, 2020), and in local elections in California (Wasserman, 2020).

Our paper adds to this strand of literature in three important ways. First, we focus on high-stakes admission-relevant exams, which most countries use to select students for college admission. Thus, our findings can directly speak to gender gaps in college enrollment and possibly future labor market. Second, previous studies focus on a selected group in the sense that individuals who do not like competition may choose not to participate in the competition in the first place. Our setting, however, can greatly alleviate the concern of sample selection because almost everyone needs to take the NCEE to get into colleges in China.

[3] Cai et al. (2019) look at how males and females perform differently between a mock exam and the actual NCEE in Anxi County, China. We look at how males and females respond differently when they confront a failure in the NCEE.

Our results can thus enhance the external validity of prior findings substantially. Lastly, our results on the gender differences in returns to retake suggest that differential return is unlikely to be an important driver of the gender differences in the tendency to retake, and our rich tests on various heterogeneous groups further improve our understanding of the potential mechanisms of the gender differences in reactions to failure.

Second, we contribute to a growing body of literature that emphasizes the importance of noncognitive skills, such as patience, self-control, and grit on human capital accumulation (Heckman et al., 2006; Borghans et al., 2008; Moffit et al., 2011; Golsteyn et al., 2013; Sutter et al., 2013; Alan et al., 2019). The NCEE admission cutoff provides a quasi-experimental variation in failure, and thus allows us to study how people react to failure in an extremely high-stakes setting. Our study suggests that grit may play an important role in human capital accumulation, which is consistent with Alan et al. (2019), who demonstrate the importance of grit in a randomized educational intervention program. In addition, our study shows that grit may be different between males and females in a high-stakes environment, which may fundamentally change the educational and career paths for all teenagers in China, and possibly for teenagers in countries that heavily rely on standardized tests for college admission.

Lastly, we contribute to the literature on the causal effects of exam retakes, particularly in the high-stakes settings that are admission-relevant (Krishna et al., 2018; Zhang et al., 2019; Goodman et al., 2020). We find that retaking the NCEE can generate substantial returns despite its high opportunity cost of waiting for another year. More interestingly, although females are less likely to retake the NCEE than males, the returns to retake for females are similar to or sometimes even higher than males.

The remainder of our paper is organized as follows. Section 2 describes the institutional background of the NCEE in China and our data. Section 3 presents the results on the cutoff-induced discontinuity in retake probability and its causal effects. Section 4 presents the results on the gender differences in the NCEE retake behavior. Finally, Section 5 concludes.

2 Institutional Background and Data

2.1 Institutional Background

The NCEE, which is also commonly known as *gaokao*, is an annual examination held on June 7th and 8th that determines the admission of almost all students into higher education institutions in China.[4] The NCEE

[4] Some provinces such as Shandong also have exams on June 9th.

is highly competitive and often described as the "toughest exam in the world." Around 10 million students compete for the admission slots of the highly selective universities each year.[5] More than 2,000 universities in China are classified into four tiers, with NCEE score cutoffs determining the eligibility of application for each tier. It is estimated that less than 10% of candidates enroll in top-tier universities, and only less than 0.2% of exam takers will be admitted into China's top five universities (Cai et al., 2019). In addition, success in the NCEE has been taught to be the central goal for most students throughout the 12 years of schooling, and has been shown to lead to substantial improvement in labor market outcomes (Jia and Li, 2021). Therefore, the NCEE is a high-stakes competition for almost the universe of students in China.

Students choose either the science or art (social science) track after the 10th grade, and they take the NCEE in their corresponding track. The most commonly adopted examination system across the provinces is the 3+X system: "3" represents the three compulsory subjects of Chinese, Mathematics, and English, each accounting for 150 of 750 of the total score. "X" represents the combined science subjects (Physics, Chemistry, and Biology) for science track or combined arts subjects (History, Geography, and Politics) for the art track, accounting for 300 of 750 of the total score. The exams are written and graded at the province level, and the test scores are only comparable within the province-year-track. In other words, students only compete with peers within the same province-year-track.

The admission process after the NCEE is hierarchical. The central government designates all higher education institutions into various tiers: tier-1 key universities, tier-2 regular universities, tier-3 universities, and tertiary technical colleges, according to the level of prestige. Tier-1 universities are the most selective universities with the best reputation in China, followed by tier-2 universities, and most tier-1 and tier-2 universities are public universities that are of high quality and charge minimal tuition (Jia and Li, 2021). By contrast, tier-3 universities are mostly private universities that are of lower quality and charge high tuition. All tier-1 to tier-3 universities are four-year universities that grant bachelor's degrees, whereas tertiary technical colleges mostly offer programs lasting two to three years. Admission into tier-1 and tier-2 universities is generally considered as an educational success, while admission into tier-3 universities or tertiary technical colleges is often considered as less desirable and a failure in college admission (Zhang et al., 2019).

After the NCEE, provinces announce the track-specific admission cutoff scores for each university tier, based on the score distributions and university quotas assigned by the Ministry of Education. Students then apply to universities by submitting a rank order list.[6] The college assignment is organized sequentially by

[5] https://www.sohu.com/a/434396300_116509 (in Chinese)

[6] Students are aware of the cutoff scores for each tier and their own test scores when they submit their applications in our sample period. See Ha et al. (2020) for more discussions on the timing of the college application submission in China.

tier: tier-1 universities first finish their assignment, then tier-2 universities recruit, followed by tier-3 universities and tertiary technical colleges. Students who score above the cutoff score of a given tier are eligible to apply to the universities in that tier, but without a guarantee of being admitted into a school in that tier. The cutoff for tier-1 universities is set as the minimum score for admission into tier-1 universities, which is often lower than the actual admission cutoff scores for most tier-1 universities. For example, a student scoring just above the tier-1 cutoff who lists only super selective universities may not be admitted into any tier-1 university because her score is lower than the admission cutoffs for the universities on her rank order list. Students scoring below the cutoff score of a given tier will not be eligible to apply for any university in that tier.

If a student is unsatisfied with the exam and admission outcomes, then she can choose not to enroll in the assigned college and retake the NCEE next year, regardless of whether she is currently admitted into a program. As the NCEE is held annually, she must wait a year for the next take. Retakers will be marked so in the administrative records but face no advantages or disadvantages in the competition. There is no official restriction on the number of times to take the NCEE, but taking the NCEE more than two times is rare.

2.2 Data

Our administrative data include the test scores and demographic information for the universe of NCEE takers in the Ningxia Province (or Ningxia Hui Autonomous Region) from 2002 to 2010. Our data only have the total test score and do not contain detailed test scores by subject.[7] Ningxia is a small province in China, with around 7 million population. Recently, there are around 60,000 NCEE takers each year in Ningxia, and the number of NCEE takers is comparable to direct-controlled municipalities such as Beijing and Shanghai.[8] We also hand-collect the year-track cutoff points for the tier-1 and tier-2 universities in Ningxia Province from publicly available records.[9]

In order to identify whether NCEE takers retake the exam in the following year and their exam performance, we match observations in the two consecutive years based on the name identifier (which uniquely identifies a full name), exact date of birth, gender, ethnicity (Han/Hui/other ethnicity) and exam track (science/art). Individuals who are matched with the observations in the next year are defined to have

[7] The test score discussed in this paper is the total score for admission purpose, which is the raw test score plus the "bonus scores" for the students. For example, students of minority ethnicity in Ningxia get "bonus scores" because of their ethnicity. As these "bonus scores" are usually still applicable if they retake the NCEE in the next year, it will not confound the decision to retake.

[8] https://www.163.com/dy/article/FGP06FE50516EN5U.html (in Chinese)

[9] Admission into tier-3 universities is much less competitive that 40% to 50% of students are eligible for a tier-3 or better university (Cai et al., 2019). In addition, we are unable to find complete public records of the cutoff points for the tier-3 universities during the sample period. Therefore, we do not focus on the tier-3 cutoffs in this paper.

retaken the NCEE in the next year.[10] Observations with identical information on the variables listed above within each year are dropped from the sample (approximately 0.1% of the sample) as they cannot be uniquely identified. Our final sample consists of 362,592 observations of NCEE takers from 2002 to 2009 and contains information on their exam performance, whether they retake the NCEE in the next year, and if so, their exam performance for the retake exam.[11]

3 Cutoff-Induced NCEE Retakes and the Effects on Exam Outcomes

3.1 Empirical Strategy

We first investigate the effects of the failure of scoring just below the cutoff and the causal effects of retaking the NCEE on exam outcomes. We focus on gender differences in Section 4. To make exam outcomes comparable across different years, we standardize the test score within each year-track, with a mean of 0 and a standard deviation of 1. We also consider an alternative measure of exam outcomes, the relative ranking of the test score, which measures the proportion of students with lower test scores within the same year-track. This measure is admission-relevant because it is the relative position among all competitors within the same year-track that determines the admission outcomes.

The propensity to retake the NCEE in the next year may be strongly correlated with unobserved student characteristics, such as inherent ability and risk preferences, and these characteristics may also be correlated with exam outcomes. In addition, students who choose to retake the NCEE may be a selective group and very different from the general population. To address endogenous retaking, we exploit the tier cutoffs for university admission and use a regression-discontinuity design to estimate the causal effects of retaking on exam outcomes. An important feature is that the tier cutoffs are exogenously determined by the score distribution and the quota assigned by the Ministry of Education each year. Students are not able to predict the exact cutoff scores, or to manipulate their test scores to be above the cutoffs. We provide evidence in Section 3.2.

Figure 1 plots the probability of retaking the NCEE in the next year against the test score for NCEE takers of the year 2009, for science track and art track separately. The retake probability measures the proportion

---

[10] One may be concerned that our approach does not fully capture the retake behavior of students. For example, if a student chooses to move to another province to retake the NCEE, then she could not be detected in our sample. However, such possibility is unlikely to invalidate our results for two reasons. First, the hukou restrictions for the NCEE takers prevent students from arbitrarily choosing the province to take the NCEE. Second, even if the student can move to another province to take the NCEE, she will likely make such choice before her first take of the NCEE rather than doing so for retakes. As the exam content is often not the same across provinces, moving for retakes is very risky.

[11] Year 2010 is excluded from our analysis because we do not have the data for the next year, and are unable to identify whether the NCEE takers in year 2010 retake the exam in the next year or not.

of NCEE takers at each score that choose to retake in the next year. The patterns of the results are very similar for other years in our sample period. It is evident that there is a dramatic decline in retake probability at the tier-2 university cutoff, particularly for students in the science track. The retake probability is much lower for students around the tier-1 cutoff, and the decline in retake probability at the tier-1 university cutoff is much less pronounced.[12] This is because for students in Ningxia, admission into a tier-2 university is generally regarded as an educational success compared with tier-3 universities or technical colleges (Zhang et al., 2019). By contrast, just falling below the tier-1 cutoff, which indicates that the student is still eligible for admission into tier-2 universities, is much less viewed as a failure in the NCEE. Therefore, we focus on the tier-2 cutoff for the rest of the paper.

To examine how falling below the tier-2 university cutoff affects the retaking behavior, we estimate the following specification:

$$Retake_{i,y,tr} = \beta I\left(Score_{i,y,tr} < Cutoff_{y,tr}\right) + \gamma_1 f\left(Score_{i,y,tr} - Cutoff_{y,tr}\right)$$

$$+\gamma_2 I(Score_{i,y,tr} < Cutoff_{y,tr}) \times f(Score_{i,y,tr} - Cutoff_{y,tr}) + \theta X_i + \mu_{y,tr} + \varepsilon_{i,y,tr}, \quad (1)$$

where $Retake_{i,y,tr}$ is a binary indicator for whether individual i in year y and track tr (science or art) retakes the NCEE next year. $Score_{i,y,tr}$ is the test score of individual i, and $Cutoff_{y,tr}$ is the cutoff score for tier-2 university admission that varies across year-track. The indicator function $I\left(Score_{i,y,tr} < Cutoff_{y,tr}\right)$ equals 1 if the test score is below the cutoff. We include a function of the running variable, $Score_{i,y,tr} - Cutoff_{y,tr}$, the distance between the test score and the cutoff, and its interaction with the indicator of below the cutoff. We consider linear and quadratic functions in this parametric specification, as well as the local polynomial non-parametric estimation and inference procedure (Calonico et al., 2014). In the parametric specifications, we control for a set of individual characteristics X_i, including gender, ethnicity, age, household registration (*hukou*) status, and whether the individual is a first-time taker of the NCEE. Year-by-track fixed effects $\mu_{y,tr}$ are also controlled. For the baseline, we use a 15-point bandwidth and uniform kernel weights. Our results are robust to alternative bandwidths and kernel weights.

The standard errors are two-way clustered at the individual identifier level and the high school-year level. The former accommodates the fact that the same individual may appear multiple times in our estimation sample.[13] For example, if a student retakes once after her first take in the NCEE, and her scores are within

[12] The tier-1 cutoff is generally higher than the tier-2 cutoff by 30-60 points, depending on the year and the exam track, and the cutoff is extremely selective and only 10% of students score above the cutoff.

[13] The individual identifier is generated based on the name identifier, exact date of birth, gender, ethnicity, and exam track. It uniquely identifies an individual within the sample.

the 15-point bandwidth around the cutoffs in both years, then she will enter the estimation sample twice.[14] The latter allows arbitrary error correlation between schoolmates in the same school cohort. We follow the recommendation of Kolesár and Rothe (2018) and do not cluster the standard errors by the discrete running variable. However, the results are very similar when the standard errors are clustered at the running variable level (Lee and Card, 2008).

3.2 Effects of Falling Below the Tier-2 Cutoff on Retake Probability

Table 1 shows the summary statistics of the individual characteristics and the indicator of retaking the NCEE in the next year. Column (1) shows the summary statistics for the full sample, and column (2) shows the summary statistics for observations within the 15-point bandwidth, which is our RD estimation sample. Columns (3) and (4) show the summary statistics for the observations below and above the tier-2 cutoff, both still within the 15-point bandwidth. One can find that students below the tier-2 cutoff are more likely to retake the NCEE next year than those above the tier-2 cutoff. Overall, these summary statistics show that retaking the NCEE is not an uncommon choice for students—28% of the NCEE takers (20% for the RD sample) choose to retake next year. The retake probability is also very stable over time in our sample period.

Before presenting our main results, we present evidence to support the validity of our regression discontinuity design. The density distribution of the running variable around the tier-2 cutoff is shown in Figure 2. We apply the manipulation testing procedure proposed by Cattaneo, Jansson and Ma (2018) and obtain a p-value of 0.82, suggesting that there is no evidence of discontinuous density in test scores around the tier-2 cutoff. This confirms our research design because the cutoffs are determined after the NCEE, and students do not have the ability to sort around the cutoffs.

We also plot the individual characteristics of students against the distance to the cutoff in Figure A1. There is no substantial discontinuous jump for these pre-determined characteristics at the cutoff. The estimation results for the balancing tests are shown in Table A1. Indeed, there is no consistent evidence showing that a pre-determined characteristic has a substantial discontinuity at the cutoff for both linear and quadratic control specifications.[15] As explained above, students do not have the ability to sort around the cutoffs because of the institutional setting, and there is no reason that students of certain characteristics are more

[14] Approximately 92.7% of observations within the 15-point window are individuals that only appear once. Approximately 3.6% of the individuals within the 15-point window appear twice. Less than 0.1% of the individuals appear more than two times in the 15-point window.

[15] There is one coefficient significant at 10% level (first-time taker) when using the linear control specification, and one coefficient significant at 5% level (age) when using the quadratic control specification. However, none of the individual characteristics show significant coefficients under both specifications.

likely to appear on one side of the cutoff. Note that the graders have no information on students and the grading process is highly regulated, and thus discrimination based on individual characteristics is not possible.

Figure 3 plots the probability of retaking the NCEE in the next year against the distance to the tier-2 cutoff score.[16] There is a notable discontinuity in retake probability around the cutoff point. The retake probability is close to 10% and relatively stable above the cutoff point, but ranges from 20% to 40% below the cutoff point. The estimated discontinuity effect without any covariates is 0.081 when using the local polynomial non-parametric estimation and inference procedure in Calonico et al. (2014), with a robust 95% confidence interval [0.051, 0.095]. Table 2 presents the results using the parametric specification (Equation (1)), for both linear and quadratic controls. The results are consistent and show that falling below the tier-2 cutoff increases the probability of retaking the NCEE by eight percentage points, which is almost an 100% increase compared to being above the cutoff. In addition, whether including the individual characteristics in the regression or not barely changes the estimates of our main results, which further suggests that the discontinuity in retake probability at the cutoff is unlikely to be confounded.

Our results are robust to alternative specification choices and inference methods. Figure A2 plots the estimated discontinuity in retake probability at the tier-2 cutoff for alternative bandwidth choices and weighting methods. In addition to the 15-point bandwidth in the baseline, we also consider 10-point, 20-point, and the data-driven optimal bandwidth (Calonico et al., 2014), as well as using the triangular kernel weights instead of the uniform kernel weights in the baseline.[17] Our results remain robust. In addition, Table A2 shows that our results are not sensitive to using alternative inference methods, including clustering the standard errors by the discrete running variable and allowing error correlation between all NCEE takers in the same high school.

3.3 Effects of Falling Below the Tier-2 Cutoff and Retake on Exam Outcomes

To estimate the causal effects of retaking the NCEE on exam outcomes for the cutoff-induced retakers, we first estimate the reduced-form effects of falling below the tier-2 cutoff on exam outcomes:

$$Y^I{}_{i,y,tr} = \beta^I I\left(Score_{i,y,tr} < Cutoff_{y,tr}\right) + \gamma_1 f\left(Score_{i,y,tr} - Cutoff_{y,tr}\right)$$

[16] Stata package *rdplot* is used for the regression discontinuity plots. See Calonico et al. (2015) and Calonico et al. (2017) for details.
[17] The CCT optimal bandwidth (Calonico et al., 2014) is 8.1 points when using the uniform kernel weights.

$$+\gamma_2 I(Score_{i,y,tr} < Cutoff_{y,tr}) \times f(Score_{i,y,tr} - Cutoff_{y,tr}) + \theta X_i + \mu_{y,tr} + \varepsilon_{i,y,tr}, \quad (2)$$

$$Y^F{}_{i,y,tr} = \beta^F I(Score_{i,y,tr} < Cutoff_{y,tr}) + \gamma_1 f(Score_{i,y,tr} - Cutoff_{y,tr})$$

$$+\gamma_2 I(Score_{i,y,tr} < Cutoff_{y,tr}) \times f(Score_{i,y,tr} - Cutoff_{y,tr}) + \theta X_i + \mu_{y,tr} + \varepsilon_{i,y,tr}, \quad (3)$$

where $Y^I{}_{i,y,tr}$ is the outcome Y in the first year of this two-year period, which is referred to as the "initial outcome". $Y^F{}_{i,y,tr}$ is the final outcome Y over this two-year period, which is equal to the outcome in the first year for those who do not retake the NCEE in the next year, and is equal to the outcomes in the next year for those who retake the NCEE in the next year. It is the final payoff of the retake decision and is referred to as the "final outcome".[18] The summary statistics of the exam outcomes are shown in Table A3. The standardized score and ranking are generally higher in the final outcome than in the initial outcome.

We distinguish the initial and final outcomes for ease of interpretation. As the initial outcomes such as test scores are realized before the cutoff is determined, they should not be affected by the cutoff ($\beta^I = 0$). By contrast, β^F identifies the effect of falling below the tier-2 cutoff on the final payoff of the retake decision. Note that we can also use $Y^F{}_{i,y,tr} - Y^I{}_{i,y,tr}$ as the dependent variable of the same specification, and the coefficient would be equal to $\beta^F - \beta^I$, which can be interpreted as the reduced-form effects of falling below the tier-2 cutoff on the improvement in exam outcomes through retakes. By definition, $Y^I{}_{i,y,tr}$ and $Y^F{}_{i,y,tr}$ only differ for those who choose to retake the NCEE in the next year, and the differences in the effects can only come from retakes. Note that because $\beta^I = 0$, this coefficient reduces to β^F, and the coefficients when using $Y^F{}_{i,y,tr} - Y^I{}_{i,y,tr}$ or $Y^F{}_{i,y,tr}$ as the dependent variable identify the same parameter of interest, which is confirmed in Table 3. We use the specification with $Y^F{}_{i,y,tr} - Y^I{}_{i,y,tr}$ as the dependent variable as the baseline specification for measuring the return to retake because it has a clear interpretation as the causal effect on the improvement in exam performance, and can be directly compared with the improvement in exam performance for retakers that are not driven by falling below the cutoff (see Section 4 for more discussions).

In addition, we can use the discontinuity as an instrument and estimate the following two-stage least square (2SLS) specification:

$$Y^F{}_{i,y,tr} - Y^I{}_{i,y,tr} = \beta_{IV} Retake_{i,y,tr} + \gamma_1 f(Score_{i,y,tr} - Cutoff_{y,tr})$$

[18] We restrict the analysis to the retaking decisions and outcomes for next year and do not analyze the decisions to retake for multiple years. Unlike other admission-related exams that can be taken multiple times in a year such as SAT, the NCEE can only be taken once per year, and the decision to retake is better modelled as a sequential decision in each year. In addition, taking the NCEE more than two times is rare—there are only around 4% of the individuals who appear in our sample more than two times.

$$+\gamma_2 I(Score_{i,y,tr} < Cutoff_{y,tr}) \times f(Score_{i,y,tr} - Cutoff_{y,tr}) + \theta X_i + \mu_{y,tr} + \varepsilon_{i,y,tr}, \quad (4)$$

where $Retake_{i,y,tr}$ is instrumented by the indicator $I(Score_{i,y,tr} < Cutoff_{y,tr})$ as in Equation (1). The coefficient β_{IV} estimates the returns to NCEE retake driven by missing the tier-2 university cutoff in terms of exam outcomes.

Figure 4 plots the exam outcomes against the distance to the tier-2 cutoff. The left panel of the figure plots the initial outcomes of standardized score and ranking, and the right panel of the figure plots the final outcomes. There is no discernible discontinuity in the initial score and ranking, and the points above and below the cutoff are almost on the same line. This is reassuring because the cutoff is determined after the initial score and ranking outcomes are realized, and should not have any effects on these outcomes. By contrast, there are pronounced discontinuities in the final score and ranking outcomes: students just below the cutoff have higher final payoffs in terms of standardized score and ranking than students just above the cutoff, who have better initial outcomes. The only plausible explanation for these differences is through higher retake probabilities for students scoring just below the cutoff, and retaking improves the exam outcomes substantially.

Table 3 presents the results using the parametric specifications (Equations (2)-(4)). Panel A shows the effects on initial outcomes, and Panel B shows the effects on final outcomes. There is little evidence on effects on initial exam outcomes.[19] By contrast, falling below the tier-2 cutoff increases the final NCEE score by 0.04 standard deviations and increases the final ranking by 0.9 percentage points. Panel C shows the effects on the differences between the final and initial outcomes, which can be interpreted as the reduced-form estimates, i.e., the effects of falling below the cutoff on the improvement in exam performance, and the estimates are almost identical to Panel B. Panel D shows the 2SLS estimates of the effects of retaking the NCEE on the improvement of exam outcomes, in which we use the indicator of falling below the cutoff as an instrumental variable for retaking. The first-stage KP *F*-statistics are well above the Stock-Yogo critical value of 16.38 (Kleibergen and Paap, 2006), suggesting a strong first stage. The 2SLS results show that retaking the NCEE increases the standardized score by 0.47 standard deviations and increases the ranking by 11 percentage points. Together, they show that retaking the NCEE leads to a substantial improvement in the exam outcomes of students, and the returns to retake are high—students can beat an additional 11% of competitors if they retake the NCEE in the next year. Figure A3 plots the

[19] There is one statistically significant coefficient in column (3) for the initial ranking, when using the linear function specification. This is because the transformation from the raw test score to ranking is not a perfect linear transformation, and the estimated discontinuity happens to be statistically significant at the cutoff. Nevertheless, the point estimate for the discontinuity is small and economically insignificant, and becomes no longer statistically significant when using the quadratic function specification that accounts for the transformation from score to ranking more flexibly.

estimated returns to retake in terms of exam outcomes under different bandwidth and specification choices, and the results are similar.

One may be concerned that these improvements in exam performance may not translate to meaningful improvements in terms of admission. In order to further illustrate the magnitude of these improvements, we use an indicator of whether the test score is above or equal to the tier-1 cutoff score as the outcome variable.[20] By construction, the initial outcome for this indicator is always equal to 0 within the 15-point bandwidth, as the tier-1 cutoff is generally higher than the tier-2 cutoff by 30-60 points in our sample. However, as shown in Figure A4, the probability that students are eligible to apply for tier-1 universities is around 5% above the cutoff and around 10% below the cutoff, with a sharp discontinuity at the cutoff when we use final exam scores, i.e., exam scores in the next year for retakers and in the initial year for non-retakers. These results show that despite both ineligible to apply for tier-1 universities in the initial year, students scoring below the tier-2 cutoff are more likely to become eligible to apply for tier-1 universities next year than students scoring above the tier-2 cutoff because of the improvement of exam scores through retake. The 2SLS estimates (Equation (4)) show that retaking the NCEE increases the probability of being eligible to apply for tier-1 universities by 51-62 percentage points for the cutoff-induced retakers, indicating that these improvements in exam performance are consequential for admission—the retakers become eligible to apply for universities of higher quality that they would not be eligible to apply for otherwise.[21]

To conclude, retaking the NCEE leads to sizeable improvements in exam outcomes and a large return in terms of educational success for students. Our estimates (0.47 standard deviations increase) are comparable to and even larger than the estimates of the causal effects of retaking the SAT on the admission-relevant superscore (around 0.34 standard deviations increase) in Goodman et al. (2020). However, we are unable to estimate the optimal retaking strategy for students because the opportunity cost of retaking the NCEE—postponing the entrance into higher education by (at least) a year—may also be large and heterogeneous for different students. In addition, our estimates are for the local average treatment effects of the students who retake the NCEE because of falling just below the tier-2 cutoff—a group of students performing better than the general population but still have a large room for improvement—and should be carefully interpreted when extrapolating the effects to the general population of all students.

[20] Note that we cannot use an indicator of whether the test score is above or equal to the tier-2 cutoff score as the outcome variable, because there is a discontinuity from 0 to 1 in the initial outcome at the cutoff by construction, which violates the continuity assumption required by regression discontinuity design (Cattaneo, Idrobo and Titiunik, 2018). Therefore, we use whether the test score is above or equal to the tier-1 cutoff score to evaluate the consequence of the improvement in exam performance.

[21] The coefficients (not reported, available upon request) are all statistically significant at the 1% level.

4 Gender Differences in the Retaking Decisions

In the previous section, we have documented that students who confront the failure of scoring just below the tier-2 cutoff are more likely to retake the NCEE in the next year, and such cutoff-induced retakes generate large returns in terms of exam performance. In this section, we investigate the gender differences in the retaking decisions when confronting the failure of missing the cutoff, and the mechanisms and explanations for these gender differences.

4.1 Empirical Strategy

To investigate the gender differences in the propensity to retake induced by missing the cutoff, as well as the effects of retakes, we first split the sample by gender and estimate the baseline specifications separately. To formally test the statistical significance of the gender differences, we use the full RD sample and estimate the following specification with full gender interactions:

$$Retake_{i,y,tr} = \alpha Male_i + \beta I\left(Score_{i,y,tr} < Cutoff_{y,tr}\right) + \delta Male_i \times I\left(Score_{i,y,tr} < Cutoff_{y,tr}\right) + \gamma_1 f\left(Score_{i,y,tr} - Cutoff_{y,tr}\right) + \rho_1 Male_i \times f\left(Score_{i,y,tr} - Cutoff_{y,tr}\right) + \gamma_2 I\left(Score_{i,y,tr} < Cutoff_{y,tr}\right) \times f\left(Score_{i,y,tr} - Cutoff_{y,tr}\right) + \rho_2 Male_i \times I\left(Score_{i,y,tr} < Cutoff_{y,tr}\right) \times f(Score_{i,y,tr} - Cutoff_{y,tr}) + \theta_1 X_i + \theta_2 Male_i \times X_i + \mu_{y,tr,male} + \varepsilon_{i,y,tr}, \quad (5)$$

where $Male_i$ is a binary indicator for being male. With full gender interactions, the slopes are allowed to be different to the left and right of the cutoff, and be different for each gender. The individual characteristics are also interacted with the male indicator to allow for differential effects, and the fixed effects are now at year-by-track-by-gender level. The coefficient δ captures the gender differences in the propensity to retake induced by missing the cutoff, and is equal to the difference in the coefficients for male and female subsamples.

For gender differences in returns to retake, we follow the same strategy as the baseline and estimate the following specification:

$$Y^F{}_{i,y,tr} - Y^I{}_{i,y,tr} = \alpha Male_i + \beta_{IV} Retake_{i,y,tr} + \delta_{IV} Male_i \times Retake_{i,y,tr} + \gamma_1 f\left(Score_{i,y,tr} - Cutoff_{y,tr}\right) + \rho_1 Male_i \times f\left(Score_{i,y,tr} - Cutoff_{y,tr}\right) + \gamma_2 I\left(Score_{i,y,tr} < Cutoff_{y,tr}\right) \times f\left(Score_{i,y,tr} - Cutoff_{y,tr}\right) + \rho_2 Male_i \times I\left(Score_{i,y,tr} < Cutoff_{y,tr}\right) \times f(Score_{i,y,tr} - Cutoff_{y,tr}) + \theta_1 X_i + \theta_2 Male_i \times X_i + \mu_{y,tr,male} + \varepsilon_{i,y,tr,} \quad (6)$$

where $Retake_{i,y,tr}$ and $Male_i \times Retake_{i,y,tr}$ are instrumented by $I(Score_{i,y,tr} < Cutoff_{y,tr})$ and $Male_i \times I(Score_{i,y,tr} < Cutoff_{y,tr})$ as in Equation (5). The coefficient δ_{IV} estimates the gender differences in the effects of the NCEE retake driven by missing the tier-2 university cutoff on improvements in the exam outcomes, and is equal to the difference in estimated returns for male and female subsamples.

4.2 Main Results

We start by showing the predictors of retaking the NCEE in the next year from a linear probability model that regresses the retake indicator on a set of covariates. The results are presented in Table A4. Columns (1)-(2) show the results for our full sample. The results show that males are two to three percentage points more likely to retake than females, and the gender difference is persistent when more covariates are added. Note that these other covariates are also strong predictors of the retake probability—students of Han ethnicity are much more likely to retake, followed by students of Hui ethnicity, compared with students of other minority ethnicities. In addition, students with urban hukou, repeated takers, older students and students with higher test scores are less likely to retake. Columns (3)-(4) show the results for the sample within the 15-point bandwidth around the tier-2 cutoff. The pattern is similar, and the gender difference is more pronounced—males are six to eight percentage points more likely to retake than females when they score around the tier-2 cutoff.

Given that retaking the NCEE is an endogenous choice that correlates with many unobservable personal traits, the gender differences in retakes could arise from gender differences in many aspects, such as confidence and goal-setting. We focus on the retakes induced by missing the tier-2 cutoff and examine whether males and females differ in the likelihood of retaking when confronting this exogenous failure. Figure 5 plots the probability of retaking the NCEE in the next year against the distance to the tier-2 cutoff separately for males and females. It is clear that males have a higher retake probability than females on both sides of the cutoff, and the gender differences are much more pronounced to the left of the cutoff. More importantly, the discontinuity in retake probability at the cutoff is much more pronounced for males than for females.

Note that our analysis on gender differences relies on the validity of the regression discontinuity design for each gender. We plot the density distribution of the running variable around the tier-2 cutoff for each gender in Figure A5, and there is no evidence of discontinuous density in test scores around the tier-2 cutoff for males or females.[22] We also plot the individual characteristics of students against the distance to the cutoff for males and females separately in Figure A6, and the estimation results for the balancing tests are shown in Table A5. Again, there is no substantial discontinuous jump for these pre-determined characteristics at the cutoff for both males and females under both linear and quadratic controls, which reassures the validity of our research design.[23]

Table 4 presents the results of the parametric specifications in Equation (5). Columns (1)-(2) present the results for males and females using the linear control separately. Males are 11 percentage points more likely to retake when falling just below the tier-2 cutoff, while females are only 5.5 percentage points more likely to retake when falling just below the cutoff. The gender difference in the retaking probability induced by the cutoff is around 5.6 percentage points, and is statistically significant at 1% level when using the full gender interaction model as in column (3). Columns (4)-(6) similarly show the results using the quadratic control. The results are very similar, and the estimated gender difference is even larger (7.3 percentage points). The gender differences in the effects are quite substantial: the discontinuity effect for males is more than twice of such effect for females.

In Table A6, we present results using the full RD sample but relaxing the model assumption of full gender interactions to assess the robustness of the results. In column (1), we do not include any individual characteristics and use year-by-track fixed effects, but still allow the slopes to be different to the left and right of the cutoff, and be different for each gender. In column (2), we include individual characteristics, but still do not include gender interactions with these covariates. Then in column (3), the effects of individual characteristics and fixed effects are allowed to vary by gender, and the specification is the same as Equation (5) and in column (3) of Table 4. The advantage of columns (1)-(2) of Table A6 is that the coefficient of the male dummy will not be absorbed as in the full gender interaction specification, and we can clearly observe the gender difference in retake probability when scoring just above the cutoff. The results in columns (1)-(3) show that, under the linear control specification, the gender difference in the retaking probability induced by the cutoff is robust to the inclusion of covariates and their gender interactions. In addition, columns (1)-(2) show that men are five to six percentage points more likely to retake than women when scoring just above the tier-2 cutoff, and such gender difference becomes around twice larger when scoring just below the cutoff. Columns (4)-(6) show the results using the quadratic control

[22] The p-value of the manipulation testing procedure proposed by Cattaneo, Jansson and Ma (2018) is 0.61 for males and 0.84 for females.

[23] There is one coefficient significant at 10% level (Han) when using the linear control specification, and one coefficient significant at 5% level (urban) when using the quadratic control specification for males. There is one coefficient significant at 10% level (Han) and one coefficient significant at 5% level (Hui) when using the linear control specification, and no significant coefficient when using the quadratic control specification for females. Again, none of the individual characteristics show significant coefficients under both specifications.

specification, and the results are similar. Figure A7 plots the estimated coefficients of gender differences (the interaction terms in columns (3) and (6) of Table 4) under different bandwidth and specification choices and shows the robustness of the results. Indeed, the estimated gender differences are large and statistically significant across various specifications.

Therefore, we can conclude that males are more likely to be motivated by missing the tier-2 cutoff and retake the NCEE next year than females. Our findings are consistent with previous studies that females are more likely to stop participating in competitions after failures (Buser and Yuan, 2019; Wasserman, 2020), and our results are for a context with much higher stakes and for a much larger population. By contrast, our results differ from Goodman et al. (2020) who find that females are more likely to retake the SAT than males, although the sense of competition against others is less clear and retake is less costly in SAT than that in the NCEE.

4.3 Understanding Gender Differences in Reactions to Failure

Why are females less inspired to retake after the failure of missing the cutoff? As retaking the NCEE is a risky choice that has high opportunity costs and uncertain returns, such gender differences may be explained by gender differences in several aspects of the decision-making process. First, the returns to retake may be different across gender. For example, if males in general have better performance and higher returns when retaking the NCEE, then it is rational for them to participate in the retakes more frequently. Second, the opportunity costs may be different across gender, as postponing the time of entering higher education and the labor market by a year could have differential impacts on men and women, especially with fertility concerns. Third, the gender differences in retake decisions may also be explained by gender differences in non-cognitive traits and preferences. For example, females may have different causal attribution than males. Men tend to attribute success to internal factors such as talent, and failure to external factors such as luck, whereas women tend to do the opposite (Dweck et al., 1978; Ryckman and Peckham, 1987; Beyer, 1998). Females who fail the cutoff may be more likely to attribute the failure to own ability and be less confident about the prospect of the retakes, and thus are less motivated to retake than males. Fourth, the gender differences could come from differences in risk preferences (Boring and Brown, 2016; Saygin, 2016; Reuben et al., 2017), that females may be less motivated to retake than males because of stronger risk aversion. Finally, although the decision to retake is made by students, parents may also have a significant influence on the decision-making process. If the financial or emotional support from parents is weaker for females because of gender differences in social norms, then females may be less likely to retake as well.

Although distinguishing between these competing explanations is hard, we are able to directly test the hypothesis of differential returns by examining whether the returns to retake in terms of exam outcomes are higher for males. Table 5 presents the results for exam outcomes using the linear control specification. The results show that the return to retake is on average 0.42 standard deviations in test scores for males, but is 0.58 standard deviations in test scores for females. The difference is statistically significant at the 10% level. When measuring the return in terms of the relative ranking, females also show a larger return than males, although the difference is insignificant. Figure A8 plots the estimated gender differences in returns to retake in terms of exam outcomes under different bandwidth and specification choices. The results are very robust—the estimates are either negative or statistically insignificant. These results show that females in general have similar or even higher returns to retake than males in terms of exam outcomes. Therefore, higher returns for males in terms of exam outcomes are unlikely to explain the results.

One may be concerned that the gender differences in returns to retake are in fact driven by gender differences in selection into retake. For example, suppose males have higher returns than females in general, and both males and females select into retake if their returns are sufficiently high, then there would be more males choosing to retake than females, and their average returns become no longer higher than females as they are more selected into retake. Indeed, the implicit assumption of comparing the returns to retake for males and females in our sample is that students do not systematically select into retake based on their expected returns to retake. We believe this assumption is likely to hold because it is hard for students to predict the returns to retake and select into retake, as the performance in NCEE—one of the toughest exams in the world—is hard to predict.[24] In addition, we provide some additional direct evidence against this rational selection hypothesis. Specifically, we compare the returns to retake for the cutoff-induced retakers ("compliers") and for the retakers that will choose to retake regardless of being above or below the cutoff ("always-retakers") at the cutoff.

Let $D_i(1)$ and $D_i(0)$ denote the potential treatment value (whether chooses to retake or not) when the individual i is assigned to be below and above the cutoff, respectively. Under the monotonicity condition (Imbens and Angrist, 1994; Cattaneo, Idrobo and Titiunik, 2018) $D_i(1) \geq D_i(0)$, the individuals can be classified into three types: "always-retakers" who always choose to retake regardless of being above or below the cutoff ($D_i(1) = D_i(0) = 1$), "compliers" who only choose to retake if being below the cutoff ($D_i(1) = 1$ and $D_i(0) = 0$), and "never-retakers" who never choose to retake ($D_i(1) = D_i(0) = 0$). Then the coefficient β in Equation (1) estimates the proportion of "compliers", and the coefficient β_{IV} in

[24] In addition, the assumption can also be guaranteed to hold when assuming that the expected returns to retake are homogeneous for students scoring just below the cutoff within each gender. Although the expected returns to retake may be heterogeneous in general, it may be plausible to assume that to be homogeneous for students who score just below the cutoff—a group of students showing similar ability and receiving the same feedback signal.

Equation (4) estimates the local average treatment effect (LATE) of retake on the improvement of the outcome for the "compliers" at the cutoff under continuity assumptions (Cattaneo, Idrobo and Titiunik, 2018).

One important feature of our setting is that we can also estimate the LATE of retake on the improvement of the outcome for the "always-retakers" as well. Specifically, we can estimate the mean of the improvement of the outcome for individuals who choose to retake and scoring exactly at the cutoff to recover the LATE for the "always-retakers" at the cutoff.[25] The intuition is that the individuals who choose to retake when scoring above the cutoff must be "always-retakers" by construction.[26] If the rational selection hypothesis is true, then the LATE for the "always-retakers" at the cutoff should be higher than the LATE for the "compliers" at the cutoff, because the former group has strong motivation to retake regardless of the cutoff, and should be the group with the strongest incentive to retake. In Appendix A, this intuition is formalized using a simple rational selection model.

We thus estimate and compare the LATE for the "compliers" and the "always-retakers" at the cutoff. The LATE on the improvement of the test score is 0.47 standard deviations for the "compliers" (as shown in Table 3, Column (1)), and is 0.40 standard deviations for the "always-retakers" (N=122, s.e.=0.03). Therefore, there is no evidence that the LATE for the "always-retakers" is higher than for the "compliers". When conducting the analysis by gender, men have a LATE of 0.42 standard deviations for the "compliers" (as shown in Table 5, Column (1)) and a LATE of 0.40 for the "always-retakers" (N=79, s.e.=0.04), while women have a LATE of 0.58 standard deviations for the "compliers" (as shown in Table 5, Column (2)) and a LATE of 0.39 for the "always-retakers" (N=43, s.e.=0.05). Again, the results are inconsistent with the rational selection model as "compliers" show higher LATE than "always-retakers", suggesting that it is unlikely that students can rationally predict their returns and select into retake. The results are similar when using the relative ranking to measure exam outcomes. Together, these results indicate that our results are not explained by rational selection.

However, the returns we examined here only refer to improvements in exam scores, not other pecuniary or non-pecuniary returns that may be associated with higher scores in the long run. For example, even if the returns to retake are similar for males and females in terms of exam performance, admission into a selective university may translate into higher pecuniary or non-pecuniary returns for males because of

---

[25] See Appendix A for more discussions on these results.

[26] Note that in fact we can similarly use the same method to recover the treatment effect for retakers at any score, but the group of retakers is in general endogenous. Therefore, we focus on the RD estimates of the LATE for the "compliers" at the cutoff when analyzing returns to retake.

labor-market conditions (Cai et al., 2019). Our research design does not allow us to rule out this potential explanation.

4.4 Heterogeneous Effects and Mechanisms

In this section, we explore the heterogeneity in gender differences in cutoff effects to further clarify the mechanisms. We focus on the linear control specification throughout the heterogeneous effect analysis, and the results are in general similar when using the quadratic control specification. Table 6 presents the estimation results of the gender differences in the cutoff effects on retake probability by individual characteristics. The gender differences are pronounced and similar for Han and minority ethnicity students, and for urban and rural students. These results suggest that the gender differences in the retake decision when confronting the failure may not heavily depend on family background and financial resources, as well as differential cultural and social norms across ethnicities. In addition, the gender differences are similar for students in the science track, where females are less represented, and students in the art track.[27] Note that none of the gender differences (the interaction term) are significantly different across these subsamples. However, the gender differences are smaller for repeated takers, who typically have more experiences and information on their own ability, than for first-time takers. The gender differences for repeated takers and first-time takers are significantly different (p-value=0.04). This implies that gender differences in over-confidence and causal attributions may explain the results, as the gender differences become smaller for students with more experience and better judgment on their own ability, who are more likely to overcome the behavioral biases in over-confidence and causal attributions. However, we cannot rule out the possibility that these results are explained by gender differences in risk preferences. For example, the gender differences in risk aversion may be smaller for repeated takers who are a selective group of people that choose to tolerate the risk of retaking in the first place than for first-time takers, which may also lead to smaller gender differences for repeated takers.

We also present the estimated gender differences in the cutoff effects on retake probability by age in Figure 6.[28] The results show that the gender differences are pronounced for younger cohorts, but much smaller for older cohorts, especially for those above 21 years old. This implies that the gender differences in opportunity cost may not play an important role in explaining our results. As women in older cohorts may

[27] In our sample, the proportion of males is around 60% in the science track, and is around 35% in the art track.

[28] A typical student enters primary school at age 6-7, and thus attends their first NCEE at age 18-19. Most of the observations in the regression discontinuity sample are of age 18 or 19. Observations of age 17 or below are likely to be individuals who enter primary school early or skip grades. Observations of age 20 or above are likely to be individuals who enter primary school late, repeat grades, or are retaking the NCEE.

face larger marital market social norm pressure, we may expect that the gender difference in the opportunity cost of spending another year is likely larger for older cohorts, which would predict more pronounced gender differences in retake probability for older cohorts. However, our results show a decline rather than an increase in the gender differences as the cohort becomes older. Note that the age of an individual is highly correlated with the probability that the individual is a repeated taker, and we are unable to distinguish the age differences from the previously-described repeated taker differences.

Table 7 presents the estimation results of the gender differences in the cutoff effects on retake probability by high school and county characteristics. In columns (1)-(2), we divide the sample based on the quality of the high school.[29] High-quality high schools have better educational resources and peer groups, and students in high-quality high schools may have higher returns and lower non-pecuniary costs if they choose to retake. However, the results show that the gender differences are large and of similar magnitude in high-quality and low-quality high schools. In columns (3)-(4) and (5)-(6), we divide the sample based on the sex ratio of the high school cohort (columns (3)-(4)) and the sex ratio of the county (columns (5)-(6)).[30] Students in places with different levels of sex imbalance may face different gender-related social norms and family support. For example, in places with high boys-to-girls ratios, girls may receive little family support and thus are less likely to retake. However, we find large and similar gender differences when the students are exposed to places with different levels of sex ratio, suggesting that social norms and family support are unlikely to fully explain our results. Finally, in columns (7)-(8), we divide the sample based on the GDP per capita of the counties.[31] Students in poor counties may have lower economic returns and higher opportunity costs of retake, and such differences in benefits and costs of retake between rich and poor counties may differ across gender. However, the results show that the gender differences are again large and of similar magnitude for students in rich and poor counties, indicating that benefits and costs may play a limited role in explaining the gender differences.[32]

[29] The quality of the high school is measured by the median of the standardized NCEE score of the students in the high school, separately measured for each year-track. A student in our RD sample is defined to be in high-quality school if the quality of her high school is above or equal to the median of the quality of high school in the RD sample in the given year-track. There are a small proportion of NCEE takers (less than 1%) that do not have valid information on high school, and they are excluded from this analysis.

[30] The sex ratio of the high school is measured by the proportion of male students in the high school, separately measured for each year. A student in our RD sample is defined to be in school with high sex ratio if the sex ratio of her high school is above or equal to the median of the sex ratio of high school in the RD sample in the given year. The NCEE takers without valid information on high school are again excluded from this analysis. The sex ratio of the county is measured by the proportion of males in the total population. A student in our RD sample is defined to be in county with high sex ratio if the sex ratio of her county is above or equal to the median of the sex ratio of county in the RD sample in the given year.

[31] A student in our RD sample is defined to be in county with high GDP if the GDP per capita of her county is above or equal to the median of the GDP per capita of county in the RD sample in the given year.

[32] Note that none of the gender differences (the interaction term) are significantly different across these subsamples.

All these results on heterogeneous effects show that the gender differences in reactions to failure are not driven by certain groups of individuals, but are pronounced for all types of individuals. These results provide substantial support for the external validity of our findings. In addition, these results suggest that gender differences in returns, opportunity costs and family support are less likely to explain our results, while gender differences in causal attributions, confidence, and risk preferences may well explain our results, as these gender differences in non-cognitive traits may exist for all types of students. Nevertheless, we are unable to fully distinguish between these potential explanations.

4.5 Implications

We conclude this section by doing a simple back-of-the-envelope calculation for the economic meaning of the gender gap in retake tendency. One may be concerned that the returns to retake at the cutoff may be different from the general population. Therefore, we conduct the calculation for the sample within 15-point bandwidth around the tier-2 cutoff only, as extrapolating the estimated returns may be more plausible within this sample. Females have a 0.06-unit higher standardized score in terms of the final outcome over the two-year period than males (0.89 vs. 0.83) in this sample. If the gender gap in retake probabilities vanishes (7.82 percentage points, as shown in Table A4, Column (4)), and assuming the returns to retake can be extrapolated to this sample (0.4730, as shown in Table 3, Column (1)), then females would have an additional 0.037-unit advantage in terms of the final standardized score in this counterfactual case. This case would expand the current gender gap in exam performance in this sample by 60%. Despite having better exam performance on average, females have a 3.3-percentage point lower probability of finally being eligible to apply for higher-quality tier-1 universities over the two-year period than males (6.3% vs. 9.6%) in this sample because of a lower tendency to retake. If the gender gap in retake probabilities vanishes, then females would be substantially more represented in high-quality universities. These effects may have important implications for the gender disparities in the labor market.

5 Conclusion

We document the gender differences in reactions to failure in high-stakes competition in an important field setting—the NCEE in China. Using unique administrative data on the universe of NCEE takers in Ningxia and exploiting a regression-discontinuity design, we show that students who score just below the tier-2 cutoff have an eight percentage point higher probability (an almost 100% increase compared to being above the cutoff) of retaking the NCEE in the next year. We then exploit the discontinuity in the probability of retaking the NCEE around the cutoff to address endogenous retaking and estimate the causal returns to retaking the NCEE. The results show that retaking the NCEE increases the test scores for admission by 0.47 standard deviation, and increases the relative ranking among competitors by 11 percentage points. Our results show that retaking the NCEE generates large returns in terms of exam performance and educational success.

We then document large gender differences in the propensity to retake in the next year. We find consistent evidence that women are less likely to retake the NCEE than men with similar exam performance. The cutoff-induced retakes from the regression discontinuity design, which reflect the desire to participate in the competition again inspired by the exogenous failure of scoring below the cutoff, are also much more pronounced for men than for women. Our results suggest that these gender differences are not explained by gender differences in returns to retake in terms of exam outcomes, or gender differences in benefits, opportunity costs and family support, but may be explained by gender differences in non-cognitive traits, such as causal attribution, confidence, and risk preferences. Our estimates suggest that if females are equally likely to retake as males, females would have better final exam performance and be substantially more represented in the high-quality universities, which may in turn, have important implications for the gender equality in the labor market.

Unfortunately, due to data limitations, we cannot examine the effects of retaking NCEE on long-term outcomes, such as labor market and marital outcomes. In addition, we are unable to fully disentangle the potential explanations for the gender differences in reactions to failure in the NCEE. Further research is needed to test these hypotheses, which could be important for policy designs to address the gender gap.

References

Alan, Sule, Teodora Boneva, and Seda Ertac. "Ever failed, try again, succeed better: Results from a randomized educational intervention on grit." *The Quarterly Journal of Economics* 134, no. 3 (2019): 1121-1162.

Astorne-Figari, Carmen, and Jamin D. Speer. "Are changes of major major changes? The roles of grades, gender, and preferences in college major switching." Economics of Education Review 70 (2019): 75-93.

Berlin, Noémi, and Marie-Pierre Dargnies. "Gender differences in reactions to feedback and willingness to compete." Journal of Economic Behavior & Organization 130 (2016): 320-336.

Beyer, Sylvia. "Gender differences in causal attributions by college students of performance on course examinations." Current Psychology 17, no. 4 (1998): 346-358.

Borghans, Lex, Angela Lee Duckworth, James J. Heckman, and Bas Ter Weel. "The economics and psychology of personality traits." *Journal of human Resources* 43, no. 4 (2008): 972-1059.

Boring, Anne, and Jennifer Brown. Gender, Competition and Choices in Higher Education. Tech. rep. Working Paper, 2016.

Buser, Thomas, Muriel Niederle, and Hessel Oosterbeek. "Gender, competitiveness, and career choices." The Quarterly Journal of Economics 129, no. 3 (2014): 1409-1447.

Buser, Thomas, Noemi Peter, and Stefan C. Wolter. "Gender, competitiveness, and study choices in high school: Evidence from Switzerland." American economic review 107, no. 5 (2017): 125-30.

Buser, Thomas, and Huaiping Yuan. "Do women give up competing more easily? Evidence from the lab and the Dutch math olympiad." American Economic Journal: Applied Economics 11, no. 3 (2019): 225-52.

Cai, Xiqian, Yi Lu, Jessica Pan, and Songfa Zhong. "Gender Gap under Pressure: Evidence from China's National College Entrance Examination." Review of Economics and Statistics 101, no. 2 (2019): 249-263.

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

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. "Optimal data-driven regression discontinuity plots." Journal of the American Statistical Association 110, no. 512 (2015): 1753-1769.

Calonico, Sebastian, Matias D. Cattaneo, Max H. Farrell, and Rocio Titiunik. "rdrobust: Software for regression-discontinuity designs." The Stata Journal 17, no. 2 (2017): 372-404.

Calonico, Sebastian, Matias D. Cattaneo, Max H. Farrell, and Rocio Titiunik. "Regression discontinuity designs using covariates." Review of Economics and Statistics 101, no. 3 (2019): 442-451.

Cattaneo, Matias D., Nicolás Idrobo, and Rocıo Titiunik. "A Practical Introduction to Regression Discontinuity Designs: Volume II." (2018).

Cattaneo, Matias D., Michael Jansson, and Xinwei Ma. "Manipulation testing based on density discontinuity." The Stata Journal 18, no. 1 (2018): 234-261.

Cattaneo, Matias D., Michael Jansson, and Xinwei Ma. "Simple local polynomial density estimators." Journal of the American Statistical Association 115, no. 531 (2020): 1449-1455.

Delaney, Judith M., and Paul J. Devereux. "Gender and Educational Achievement: Stylized Facts and Causal Evidence." (2021).

Dweck, Carol S., William Davidson, Sharon Nelson, and Bradley Enna. "Sex differences in learned helplessness: II. The contingencies of evaluative feedback in the classroom and III. An experimental analysis." Developmental psychology 14, no. 3 (1978): 268.

Ellison, Glenn, and Ashley Swanson. Dynamics of the gender gap in high math achievement. No. w24910. National Bureau of Economic Research, 2018.

Fang, Chao, Ernest Zhang, and Junfu Zhang. "Do women give up competing more easily? Evidence from speedcubers." Economics Letters (2021): 109943.

Flory, Jeffrey A., Andreas Leibbrandt, and John A. List. "Do competitive workplaces deter female workers? A large-scale natural field experiment on job entry decisions." The Review of Economic Studies 82, no. 1 (2015): 122-155.

Golsteyn, Bart HH, Hans Grönqvist, and Lena Lindahl. "Adolescent time preferences predict lifetime outcomes." *The Economic Journal* 124, no. 580 (2014): F739-F761.

Goodman, Joshua, Oded Gurantz, and Jonathan Smith. "Take two! SAT retaking and college enrollment gaps." American Economic Journal: Economic Policy 12, no. 2 (2020): 115-58.

Ha, Wei, Le Kang, and Yang Song. "College matching mechanisms and matching stability: Evidence from a natural experiment in China." Journal of Economic Behavior & Organization 175 (2020): 206-226.

Heckman, James J., Jora Stixrud, and Sergio Urzua. "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior." *Journal of Labor economics* 24, no. 3 (2006): 411-482.

Imbens, Guido W., and Joshua D. Angrist. "Identification and estimation of local average treatment effects." Econometrica 62, no. 2 (1994): 467-475.

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

Kleibergen, Frank, and Richard Paap. "Generalized reduced rank tests using the singular value decomposition." *Journal of econometrics* 133, no. 1 (2006): 97-126.

Kolesár, Michal, and Christoph Rothe. "Inference in regression discontinuity designs with a discrete running variable." American Economic Review 108, no. 8 (2018): 2277-2304.

Krishna, Kala, Sergey Lychagin, and Veronica Frisancho. "Retaking in High Stakes Exams: Is Less More?." International Economic Review 59, no. 2 (2018): 449-477.

Landaud, Fanny, and Éric Maurin. "Aim High and Persevere! Competitive Pressure and Access Gaps in Top Science Graduate Programs." Working Paper. (2020).

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

Moffitt, Terrie E., Louise Arseneault, Daniel Belsky, Nigel Dickson, Robert J. Hancox, HonaLee Harrington, Renate Houts et al. "A gradient of childhood self-control predicts health, wealth, and public safety." *Proceedings of the national Academy of Sciences* 108, no. 7 (2011): 2693-2698.

Niederle, Muriel, and Lise Vesterlund. "Do women shy away from competition? Do men compete too much?." The Quarterly Journal of Economics 122, no. 3 (2007): 1067-1101.

Reuben, Ernesto, Matthew Wiswall, and Basit Zafar. "Preferences and biases in educational choices and labour market expectations: Shrinking the black box of gender." The Economic Journal 127, no. 604 (2017): 2153-2186.

Ryckman, David B., and Percy D. Peckham. "Gender differences in attributions for success and failure." The Journal of Early Adolescence 7, no. 1 (1987): 47-63.

Saygin, Perihan Ozge. "Gender differences in preferences for taking risk in college applications." Economics of Education Review 52 (2016): 120-133.

Sutter, Matthias, Martin G. Kocher, Daniela Glätzle-Rützler, and Stefan T. Trautmann. "Impatience and uncertainty: Experimental decisions predict adolescents' field behavior." *American Economic Review* 103, no. 1 (2013): 510-31.

Wasserman, Melanie. "Gender Differences in Politician Persistence." (2020).

Zhang, Yan, Yanqing Ding, Honghui Li, Wei Li, Yujie Ma and Xiaoyang Ye. "Retaking College Entrance Exam: Evidence from a Regression Discontinuity Design." Global Education 48, no. 7 (2019): 111-128. (in Chinese)

Appendix A Theoretical Results on Testing Rational Selection Hypothesis

Let $Z_i = I(S_i < c)$ denote the treatment assignment for individual i with running variable S_i, which is equal to 1 if the running variable is below the cutoff c. Let the treatment take-up status $D_i = Z_i \times D_i(1) + (1 - Z_i) \times D_i(0)$ denote whether individual i chooses to retake. Specifically, $D_i(1)$ and $D_i(0)$ denote the potential treatment value when the individual is assigned to be below and above the cutoff, respectively. The monotonicity condition (Imbens and Angrist, 1994; Cattaneo, Idrobo and Titiunik, 2018) requires that $D_i(1) \geq D_i(0)$, which rules out the possibility that the individual only chooses to retake when scoring above the cutoff ($D_i(1) = 0$ and $D_i(0) = 1$). Then the individuals can be classified into three types: "always-retakers" who always choose to retake regardless of being above or below the cutoff ($D_i(1) = D_i(0) = 1$), "compliers" who only choose to retake if being below the cutoff ($D_i(1) = 1$ and $D_i(0) = 0$), and "never-retakers" who never choose to retake ($D_i(1) = D_i(0) = 0$). Note that the first-stage RD estimand $lim_{s\uparrow c}\, E(D_i|S_i = c) - lim_{s\downarrow c}\, E(D_i|S_i = c)$, which is estimated by the coefficient β in Equation (1), identifies the proportion of "compliers" at the cutoff under continuity assumptions (Cattaneo, Idrobo and Titiunik, 2018):

$$lim_{s\uparrow c}\, E(D_i|S_i = c) - lim_{s\downarrow c}\, E(D_i|S_i = c) = E(D_i(1) - D_i(0)|S_i = c).$$

Let the outcome $Y_i = D_i \times Y_i(1) + (1 - D_i) \times Y_i(0)$ denote the initial and final outcomes such as standardized test score. Note that the 1 and 0 in the function $Y_i(.)$ refer to the treatment take-up status $D_i = 1$ and $D_i = 0$, not the treatment assignment status Z_i. In addition, for initial outcomes that are realized before the cutoff is determined, such as standardized test score, ${Y^I}_i(1) = {Y^I}_i(0)$ holds by construction. Then, the 2SLS RD estimand identifies the local average treatment effect (LATE) of retake on the outcome for the "compliers" at the cutoff under continuity assumptions (Cattaneo, Idrobo and Titiunik, 2018):

$$\frac{lim_{s\uparrow c}\, E(Y_i|S_i = c) - lim_{s\downarrow c}\, E(Y_i|S_i = c)}{lim_{s\uparrow c}\, E(D_i|S_i = c) - lim_{s\downarrow c}\, E(D_i|S_i = c)} = E(Y_i(1) - Y_i(0)|S_i = c, D_i(1) = 1, D_i(0) = 0).$$

Note that ${Y^F}_i(0) - {Y^I}_i(0) = 0$ holds by construction, which means that the final outcomes and the initial outcomes would be the same if not retaking the exam. Therefore, the coefficient β_{IV} in Equation (4) estimates the LATE of retake on the improvement of the outcome for the "compliers" at the cutoff, $E({Y^F}_i(1) - {Y^I}_i(1)|S_i = c, D_i(1) = 1, D_i(0) = 0)$.

One important feature of our setting is that we can also estimate the LATE of retake on the improvement of the outcome for the "always-retakers" as well. Specifically, we have

$$E({Y^F}_i - {Y^I}_i|S_i = c, D_i = 1) = E({Y^F}_i(1) - {Y^I}_i(1)|S_i = c, D_i(1) = D_i(0) = 1).$$

Therefore, we can estimate the mean of the improvement of the outcome for individuals who choose to retake and scoring exactly at the cutoff to recover the LATE for the "always-retakers" at the cutoff. The intuition is that the individuals who choose to retake when scoring above the cutoff must be "always-retakers" by construction. If the rational selection hypothesis is true, then the LATE for the "always-retakers" at the cutoff should be higher than the LATE for the "compliers" at the cutoff, because the former group has strong motivation to retake regardless of the cutoff, and should be the group with the strongest incentive to retake.

To formalize this intuition, consider the following simple model where individual i chooses to retake if

$$R_i(S_i) - \theta_i > 0,$$

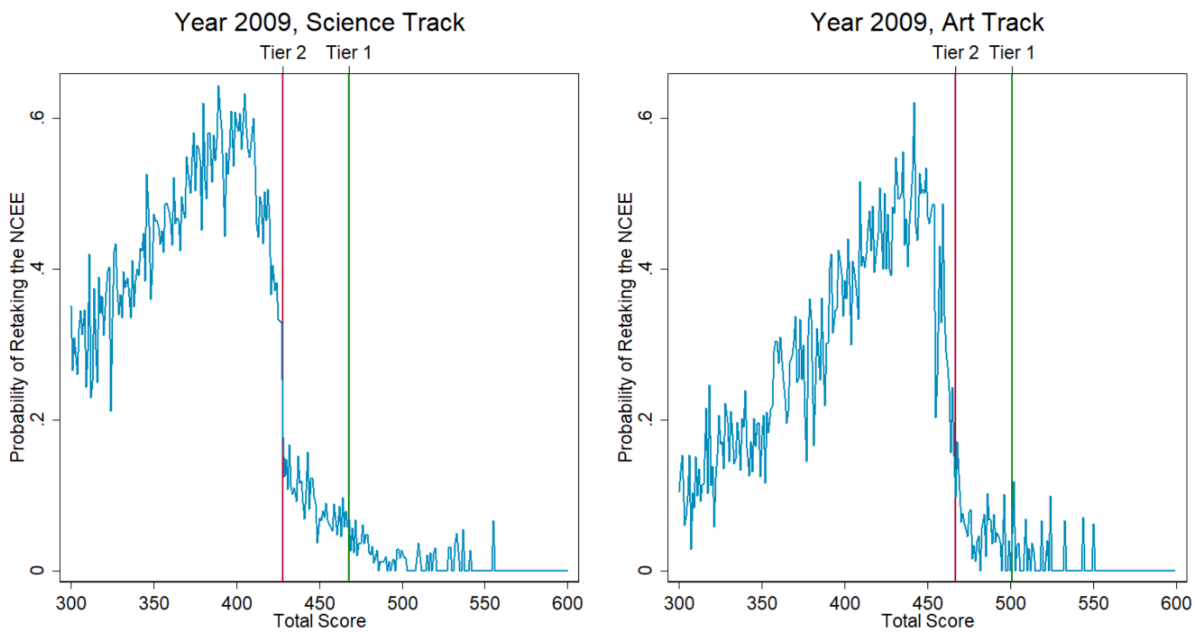
that the pecuniary return to retake R_i for individual i with test score is higher than the pecuniary cost of retake θ_i. Let $R_i(S_i) = r_i + I(S_i < c) \times \mu_i$, that the return to retake is equal to a random variable r_i drawn from the distribution $F(.)$ if the test score is announced to be above or equal to the cutoff, and is equal to r_i plus an additional pecuniary benefit term $\mu_i > 0$ if the test score is announced to be below the cutoff. The term μ_i reflects the potential benefit of the possibility to move up in the admission tiers in the next take as the test score is very close to but below the cutoff in this year. In fact, the term μ_i is the source of the discontinuity in retake probability at the cutoff. For simplicity, let θ_i and μ_i to be constant.

Under this simple model with rational selection, individuals are divided into different types based on their return. For individuals with test score $S_i = c$, they are "always-retakers" if $r_i > \theta$, as they choose to retake regardless of whether the cutoff is below their test scores or not. Individuals with $\theta - \mu < r_i \leq \theta$ are "compliers", as their returns are not sufficiently high for them to retake if their scores are already above the cutoff, but would be sufficiently high for them to retake if their scores are below the cutoff. Individuals with $r_i \leq \theta - \mu$ are "never-retakers" that have low returns even if their scores are below the cutoff. Therefore, at any score S_i, the expected return for "always-retakers" should always be higher than the expected return for "compliers":

$$E(r_i | r_i > \theta) > \theta > E(r_i | \theta - \mu < r_i \leq \theta).$$

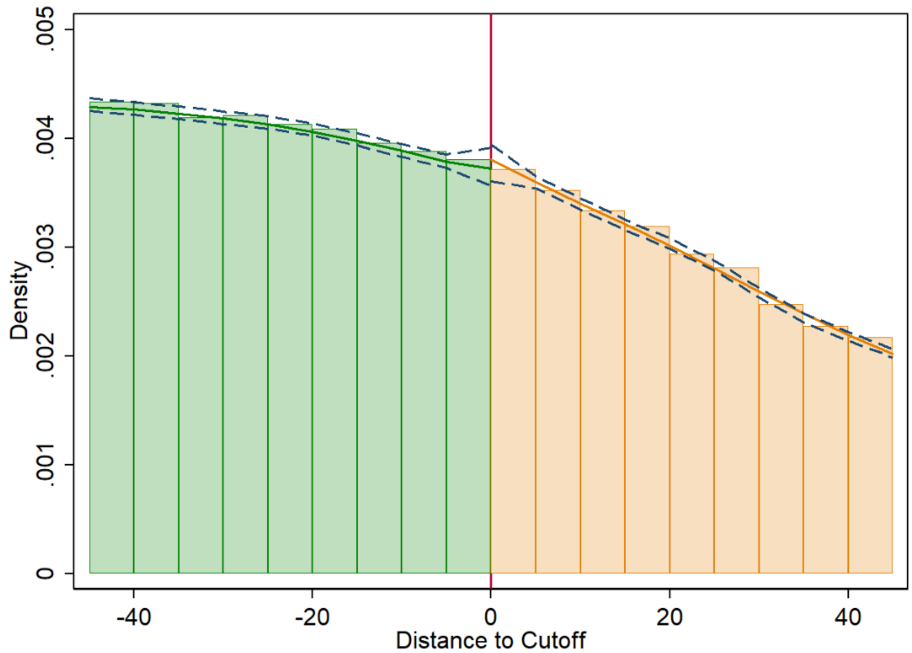
Figures and Tables

Figure 1: The Relationship between the Probability of NCEE Retaking and Test Scores



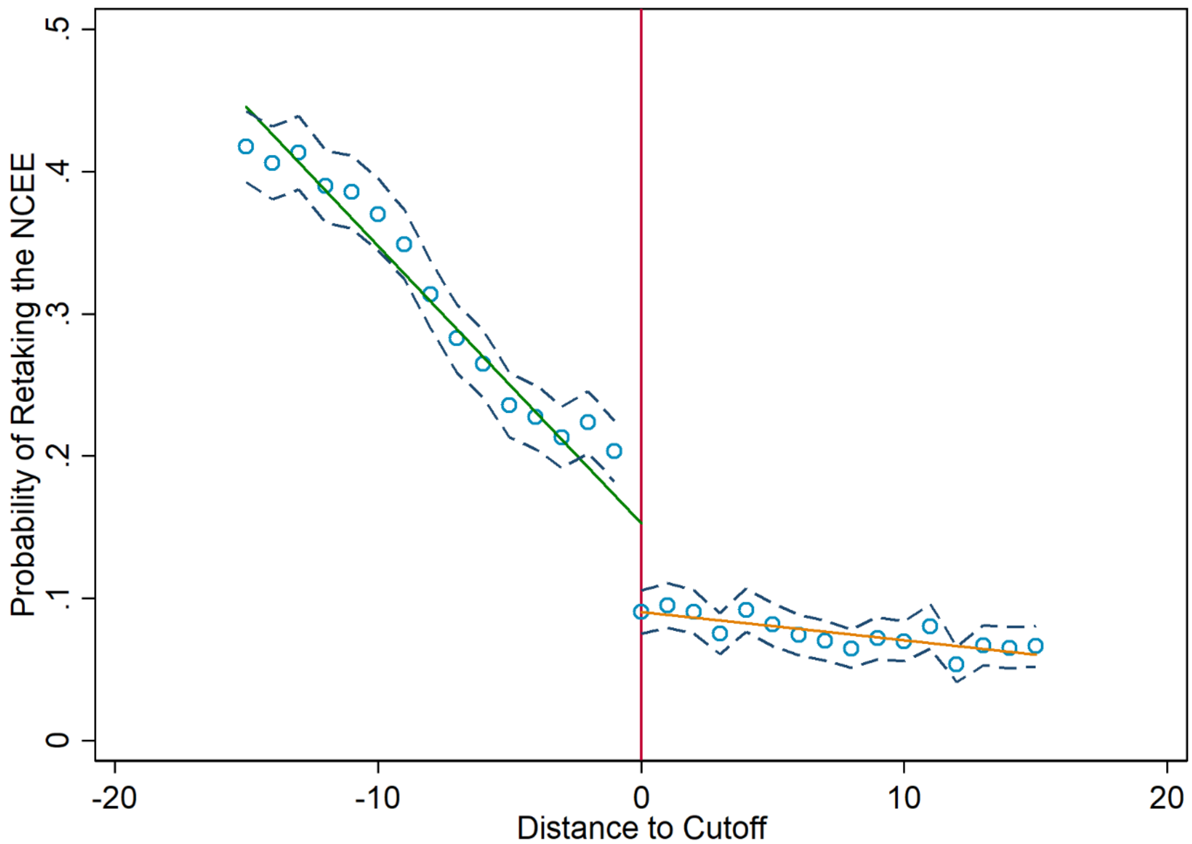
Notes: This figure plots the probability of retaking the NCEE in the next year against the test score for NCEE takers of year 2009, for science track (left) and art track (right) separately. The retake probability measures the proportion of NCEE takers at each score that choose to retake in the next year. The lines in each panel of the figure represent the cutoff scores for tier-2 (left) and tier-1 (right) university admission for each track.

Figure 2: Test of Running Variable Density Smoothness around Tier-2 Cutoff



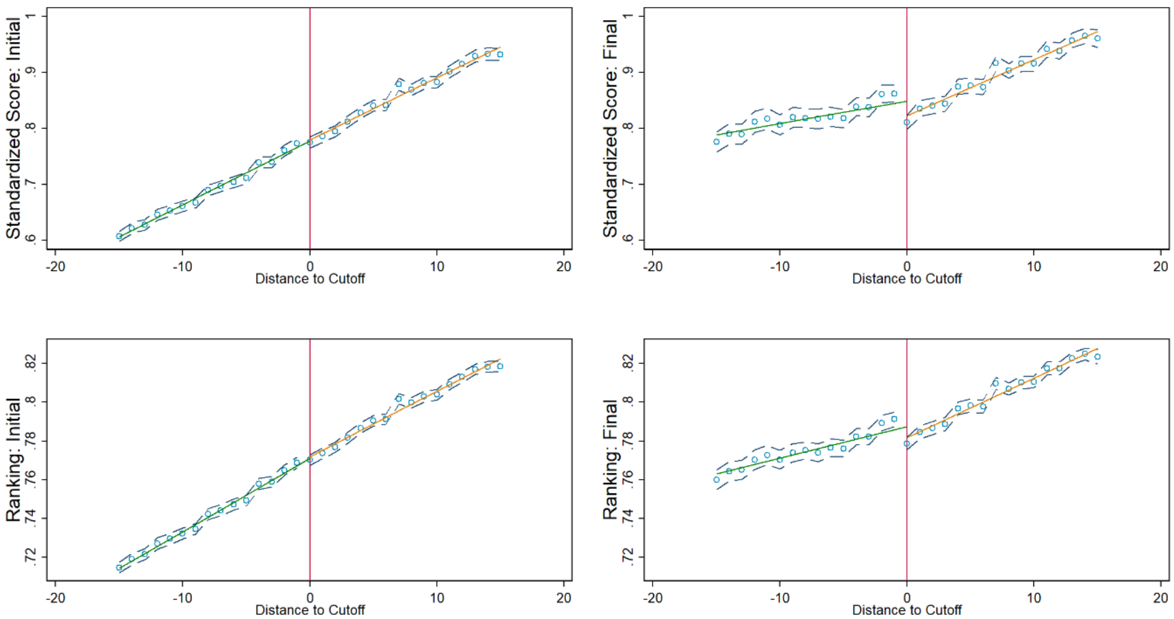
Notes: This figure plots the density of the running variable (the distance to the tier-2 cutoff score) following the manipulation testing procedure in Cattaneo, Jansson and Ma (2018). The bars in this figure represent the density distribution of the running variable over 5-point bins. The straight lines represent the estimated density to the left and to the right of the cutoff using the local polynomial density estimators proposed in Cattaneo et al. (2020). The dashed lines represent the lower and upper bounds of the 95% confidence interval for the estimated density.

Figure 3: The Probability of NCEE Retaking vs. Distance to Tier-2 University Cutoff Score



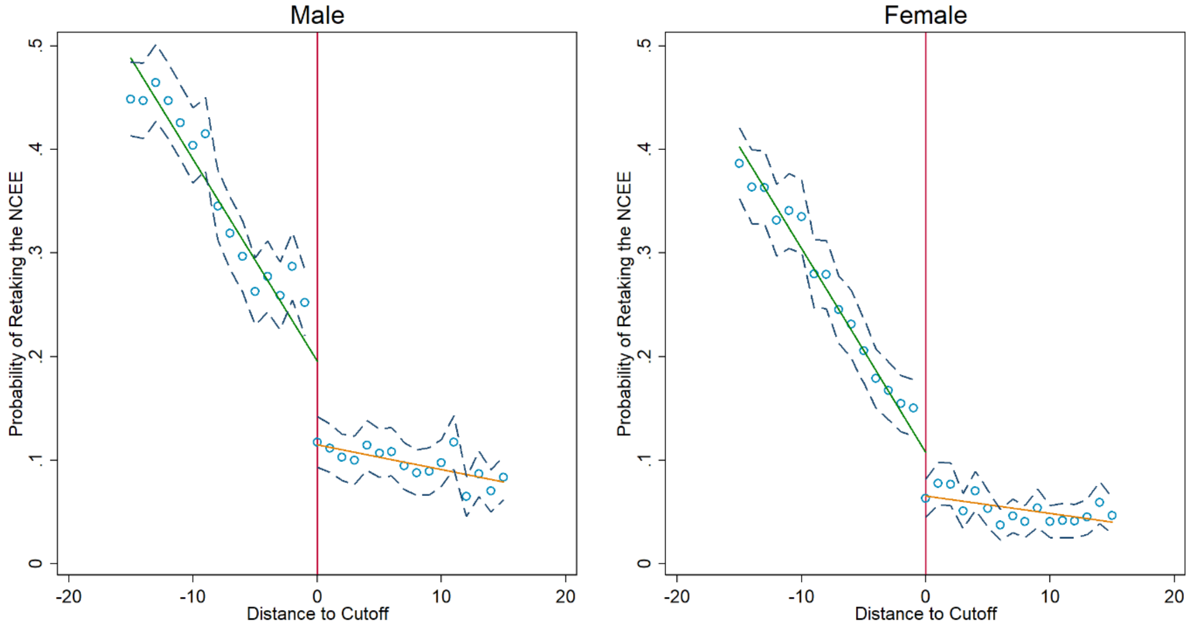
Notes: This figure plots the probability of retaking the NCEE in the next year against the distance to the tier-2 cutoff score. The sample consists of observations within the 15-point bandwidth around the cutoff. Each circle corresponds to one point in the test score. The straight lines represent the fitted linear functions to the left and to the right of the cutoff. The dashed lines represent the lower and upper bounds of the 95% confidence interval for the sample mean of the outcome variable within the corresponding bin.

Figure 4: Exam Outcomes vs. Distance to Tier-2 University Cutoff Score



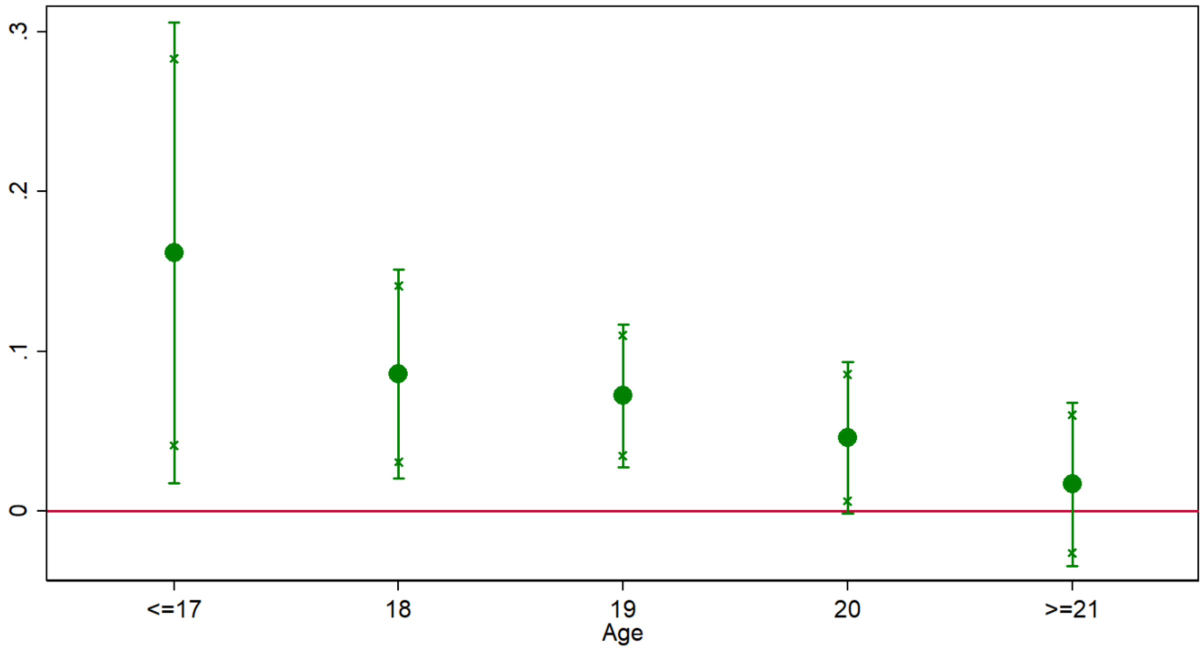
Notes: This figure plots the exam outcomes against the distance to the tier-2 cutoff score. The upper panel uses the standardized score as the outcome variable. The lower panel uses the ranking within the year-track as the outcome variable. The left panel shows the initial outcomes, i.e., the dependent variables in the current year. The right panel shows the final outcomes, i.e. the final payoffs of the dependent variables, which are equal to the dependent variables in the current year if the individual does not retake the NCEE in the next year, and are equal to the dependent variables in the next year if the individual retakes the NCEE in the next year. The sample consists of observations within the 15-point bandwidth around the cutoff. Each circle corresponds to one point in the test score. The straight lines represent the fitted linear functions to the left and to the right of the cutoff. The dashed lines represent the lower and upper bounds of the 95% confidence interval for the sample mean of the outcome variable within the corresponding bin.

Figure 5: The Probability of NCEE Retaking vs. Distance to Tier-2 University Cutoff Score, by Gender



Notes: This figure plots the probability of retaking the NCEE in the next year against the distance to the tier-2 cutoff score separately for males and females. The left panel is for males, and the right panel is for females. The sample consists of observations within the 15-point bandwidth around the cutoff. Each circle corresponds to one point in the test score. The straight lines represent the fitted linear functions to the left and to the right of the cutoff. The dashed lines represent the lower and upper bounds of the 95% confidence interval for the sample mean of the outcome variable within the corresponding bin.

Figure 6: Gender Differences in the Effects of Below Cutoff on Retake Probability, By Age



Notes: This figure plots the estimated gender differences in the effects of below the tier-2 cutoff on retake probability for different age groups. The sample consists of observations within the 15-point bandwidth around the cutoff. The linear function of the running variable and its interaction with the indicator of below the cutoff, and their gender interactions are controlled in all regressions. Individual characteristics (ethnicity, hukou status, whether the individual is a first-time taker, age) and their gender interactions, as well as year-by-track-by-gender fixed effects, are also controlled in all regressions. Standard errors are two-way clustered at the individual identifier level and the high school-year level. "x" markers represent bounds of 90% confidence interval. "-" markers represent bounds of 95% confidence interval.

Table 1: Summary Statistics

	(1)	(2)	(3)	(4)
Sample	Full	[-15,15]	[-15,0)	[0,15]
Observations	362,592	41,477	21,123	20,354
	Mean (S.D.)			
Male	0.52	0.51	0.51	0.51
	(0.50)	(0.50)	(0.50)	(0.50)
Ethnicity: Han	0.78	0.79	0.79	0.79
	(0.41)	(0.41)	(0.41)	(0.40)
Ethnicity: Hui	0.20	0.19	0.19	0.19
	(0.40)	(0.39)	(0.39)	(0.39)
Urban	0.45	0.46	0.44	0.47
	(0.50)	(0.50)	(0.50)	(0.50)
First-Time Taker	0.73	0.57	0.56	0.58
	(0.44)	(0.50)	(0.50)	(0.49)
Age	19.15	19.16	19.19	19.12
	(1.23)	(1.24)	(1.24)	(1.23)
Retake	0.28	0.20	0.31	0.08
	(0.45)	(0.40)	(0.46)	(0.27)

Notes: This table shows the summary statistics of individual characteristics and the indicator of retaking the NCEE in the next year. Column (1) is for the full sample. Column (2) is for the sample within the 15-point bandwidth around the tier-2 cutoff. Column (3) is for the sample in column (2) that is below the cutoff. Column (4) is for the sample in column (2) that is above or equal to the cutoff.

Table 2: Effects of Below Tier-2 University Cutoff on Retake Probability

	(1)	(2)	(3)	(4)
Variables		Retake		
Below Cutoff	0.0805***	0.0831***	0.0737***	0.0753***
	(0.0081)	(0.0076)	(0.0122)	(0.0115)
Observations	41,477	41,477	41,477	41,477
R-squared	0.117	0.220	0.117	0.220
Bandwidth	15	15	15	15
Interaction Controls	Linear	Linear	Quadratic	Quadratic
Individual Characteristics	No	Yes	No	Yes
Year-Track FE	Yes	Yes	Yes	Yes

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The sample consists of observations within the 15-point bandwidth around the cutoff. The dependent variable is an indicator of retaking the NCEE in the next year. Columns (1) and (2) control for a linear function of the running variable and its interaction with the indicator of below the cutoff. Columns (3) and (4) control for a quadratic function of the running variable and its interaction with the indicator of below the cutoff. Columns (1) and (3) do not control for individual characteristics. Columns (2) and (4) control for a set of individual characteristics, including gender, ethnicity, hukou status, whether the individual is a first-time taker, and age. Year-by-track fixed effects are controlled in all columns. Standard errors are two-way clustered at the individual identifier level and the high school-year level.

Table 3: Effects of Below Tier-2 University Cutoff and Retake on Exam Outcomes

	(1)	(2)	(3)	(4)
Variables	Standardized Score		Ranking	
Panel A: Initial Outcomes				
Below Cutoff	-0.0001	-0.0001	-0.0004***	-0.0000
	(0.0001)	(0.0002)	(0.0001)	(0.0001)
Panel B: Final Outcomes				
Below Cutoff	0.0392***	0.0374***	0.0090***	0.0097***
	(0.0046)	(0.0073)	(0.0011)	(0.0018)
Panel C: Differences in Outcomes				
Below Cutoff	0.0393***	0.0375***	0.0094***	0.0097***
	(0.0046)	(0.0073)	(0.0011)	(0.0018)
Panel D: Differences in Outcomes, 2SLS				
Retake	0.4730***	0.4978***	0.1127***	0.1293***
	(0.0389)	(0.0679)	(0.0094)	(0.0169)
1st-stage KP F-stat	120.1	42.8	120.1	42.8
Observations	41,477	41,477	41,477	41,477
Bandwidth	15	15	15	15
Interaction Controls	Linear	Quadratic	Linear	Quadratic
Individual Characteristics	Yes	Yes	Yes	Yes
Year-Track FE	Yes	Yes	Yes	Yes

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The sample consists of observations within the 15-point bandwidth around the cutoff. The dependent variable in columns (1) and (2) is the standardized score. The dependent variable in columns (3) and (4) is the ranking within the year-track. Panel A shows the results using the initial outcomes, i.e. the dependent variables in the current year. Panel B shows the results using the final outcomes, i.e. the final payoffs of the dependent variables, which are equal to the dependent variables in the current year if the individual does not retake the NCEE in the next year, and are equal to the dependent variables in the next year if the individual retakes the NCEE in the next year. Panel C shows the results using the differences between the final outcomes and the initial outcomes as the dependent variables. Panel D uses the same dependent variables as Panel C, but uses a 2SLS specification and uses the indicator of below the cutoff as an instrument for the indicator of retaking the NCEE in the next year. Columns (1) and (3) control for a linear function of the running variable and its interaction with the indicator of below the cutoff. Columns (2) and (4) control for a quadratic function of the running variable and its interaction with the indicator of below the cutoff. Gender, ethnicity, hukou status, whether the individual is a first-time taker, age, and year-by-track fixed effects are controlled in all columns. Standard errors are two-way clustered at the individual identifier level and the high school-year level.

Table 4: Gender Differences in the Effects of Below Tier-2 University Cutoff on Retake Probability

	(1)	(2)	(3)	(4)	(5)	(6)
Sample	Male	Female	Full	Male	Female	Full
Variables			Retake			
Male*Below Cutoff			0.0558***			0.0732***
			(0.0133)			(0.0203)
Below Cutoff	0.1107***	0.0549***	0.0549***	0.1109***	0.0377***	0.0377***
	(0.0108)	(0.0093)	(0.0093)	(0.0167)	(0.0137)	(0.0137)
Observations	21,162	20,315	41,477	21,162	20,315	41,477
Bandwidth	15	15	15	15	15	15
Interaction Controls	Linear	Linear	Linear	Quadratic	Quadratic	Quadratic
Individual Characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Year-Track FE	Yes	Yes	By Gender	Yes	Yes	By Gender

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The sample consists of observations within the 15-point bandwidth around the cutoff. The dependent variable is an indicator of retaking the NCEE in the next year. Columns (1) and (4) are using only the male sample. Columns (2) and (5) are using only the female sample. Columns (1)-(2) control for a linear function of the running variable and its interaction with the indicator of below the cutoff. Columns (4)-(5) control for a quadratic function of the running variable and its interaction with the indicator of below the cutoff. Individual characteristics (ethnicity, hukou status, whether the individual is a first-time taker, age) and year-by-track fixed effects are controlled in columns (1)-(2) and (4)-(5). Columns (3) and (6) are using the full sample with full gender interactions. Columns (3) controls for a linear function of the running variable and its interaction with the indicator of below the cutoff, and their gender interactions. Columns (6) controls for a quadratic function of the running variable and its interaction with the indicator of below the cutoff, and their gender interactions. Columns (3) and (6) control for individual characteristics (ethnicity, hukou status, whether the individual is a first-time taker, age) and their gender interactions, as well as year-by-track-by-gender fixed effects. Standard errors are two-way clustered at the individual identifier level and the high school-year level.

Table 5: Gender Differences in the Effects of Retake on Exam Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
Sample	Male	Female	Full	Male	Female	Full
Variables	Standardized Score			Ranking		
Male*Retake			-0.1566*			-0.0246
			(0.0904)			(0.0223)
Retake	0.4220***	0.5785***	0.5785***	0.1049***	0.1295***	0.1295***
	(0.0460)	(0.0765)	(0.0765)	(0.0119)	(0.0179)	(0.0179)
1st-stage KP F-stat	105.0	35.2	17.6	105.0	35.2	17.6
Observations	21,162	20,315	41,477	21,162	20,315	41,477
Bandwidth	15	15	15	15	15	15
Interaction Controls	Linear	Linear	Linear	Linear	Linear	Linear
Individual Characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Year-Track FE	Yes	Yes	By Gender	Yes	Yes	By Gender

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The sample consists of observations within the 15-point bandwidth around the cutoff. The dependent variables are the differences between the final outcomes and the initial outcomes of the exam outcomes (standardized score in columns (1)-(3), the ranking within the year-track in columns (4)-(6)). The indicator of below the cutoff (and its interaction with male dummy) are used as instruments for the indicator of retaking the NCEE in the next year (and its interaction with male dummy). Columns (1) and (4) are using only the male sample. Columns (2) and (5) are using only the female sample. Columns (1)-(2) and (4)-(5) control for a linear function of the running variable and its interaction with the indicator of below the cutoff. Individual characteristics (ethnicity, hukou status, whether the individual is a first-time taker, age) and year-by-track fixed effects are controlled in columns (1)-(2) and (4)-(5). Columns (3) and (6) are using the full sample with full gender interactions. The linear function of the running variable and its interaction with the indicator of below the cutoff, and their gender interactions are controlled in columns (3) and (6). Columns (3) and (6) control for individual characteristics (ethnicity, hukou status, whether the individual is a first-time taker, age) and their gender interactions, as well as year-by-track-by-gender fixed effects. Standard errors are two-way clustered at the individual identifier level and the high school-year level.

Table 6: Gender Differences by Individual Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sample	Han	Minority	Urban	Rural	Art	Science	First-Time Taker	Repeated Taker
Variables					Retake			
Male*Below Cutoff	0.0506***	0.0739***	0.0504**	0.0627***	0.0498*	0.0587***	0.0822***	0.0276**
	(0.0150)	(0.0247)	(0.0205)	(0.0188)	(0.0260)	(0.0158)	(0.0216)	(0.0136)
Below Cutoff	0.0683***	0.0017	0.0646***	0.0464***	0.0567***	0.0531***	0.0902***	0.0077
	(0.0109)	(0.0151)	(0.0121)	(0.0135)	(0.0140)	(0.0117)	(0.0137)	(0.0093)
Observations	32,896	8,581	18,885	22,592	12,556	28,921	23,643	17,834
Bandwidth	15	15	15	15	15	15	15	15
Interaction Controls	Linear	Linear	Linear	Linear	Linear	Linear	Linear	Linear
Individual Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Track-Gender FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

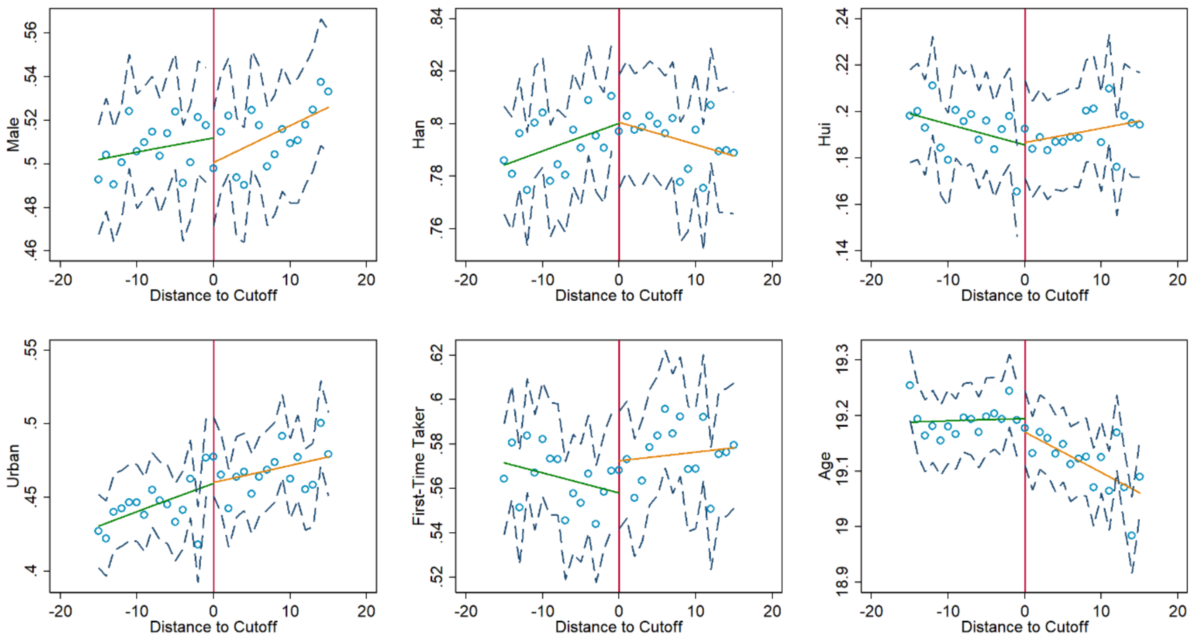
Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The sample consists of observations within the 15-point bandwidth around the cutoff. The dependent variable is an indicator of retaking the NCEE in the next year. Column (1) uses the sample of observations with Han ethnicity. Column (2) uses the sample of observations with minority ethnicity (Hui and other minority ethnicities). Column (3) uses the sample of observations with urban hukou status. Column (4) uses the sample of observations with rural hukou status. Column (5) uses the sample of observations in art track. Column (6) uses the sample of observations in science track. Column (7) uses the sample of observations that are first-time takers. Column (8) uses the sample of observations that are repeated takers. All columns control for a linear function of the running variable and its interaction with the indicator of below the cutoff, and their gender interactions. Individual characteristics (ethnicity, hukou status, whether the individual is a first-time taker, age) and their gender interactions, as well as year-by-track-by-gender fixed effects are controlled in all columns. Standard errors are two-way clustered at the individual identifier level and the high school-year level.

Table 7: Gender Differences by High School and County Characteristics

Sample	(1) High Quality High School	(2) Low Quality High School	(3) High Sex Ratio High School	(4) Low Sex Ratio High School	(5) High Sex Ratio County	(6) Low Sex Ratio County	(7) High GDP County	(8) Low GDP County
Variables	Retake							
Male*Below Cutoff	0.0642***	0.0461***	0.0553***	0.0556***	0.0466**	0.0638***	0.0478***	0.0604***
	(0.0205)	(0.0177)	(0.0195)	(0.0179)	(0.0181)	(0.0195)	(0.0172)	(0.0210)
Below Cutoff	0.0527***	0.0598***	0.0579***	0.0528***	0.0657***	0.0427***	0.0586***	0.0550***
	(0.0128)	(0.0134)	(0.0144)	(0.0116)	(0.0129)	(0.0131)	(0.0112)	(0.0155)
Observations	21,098	20,059	21,035	20,122	22,286	19,191	21,785	19,692
Bandwidth	15	15	15	15	15	15	15	15
Interaction Controls	Linear	Linear	Linear	Linear	Linear	Linear	Linear	Linear
Individual Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Track-Gender FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

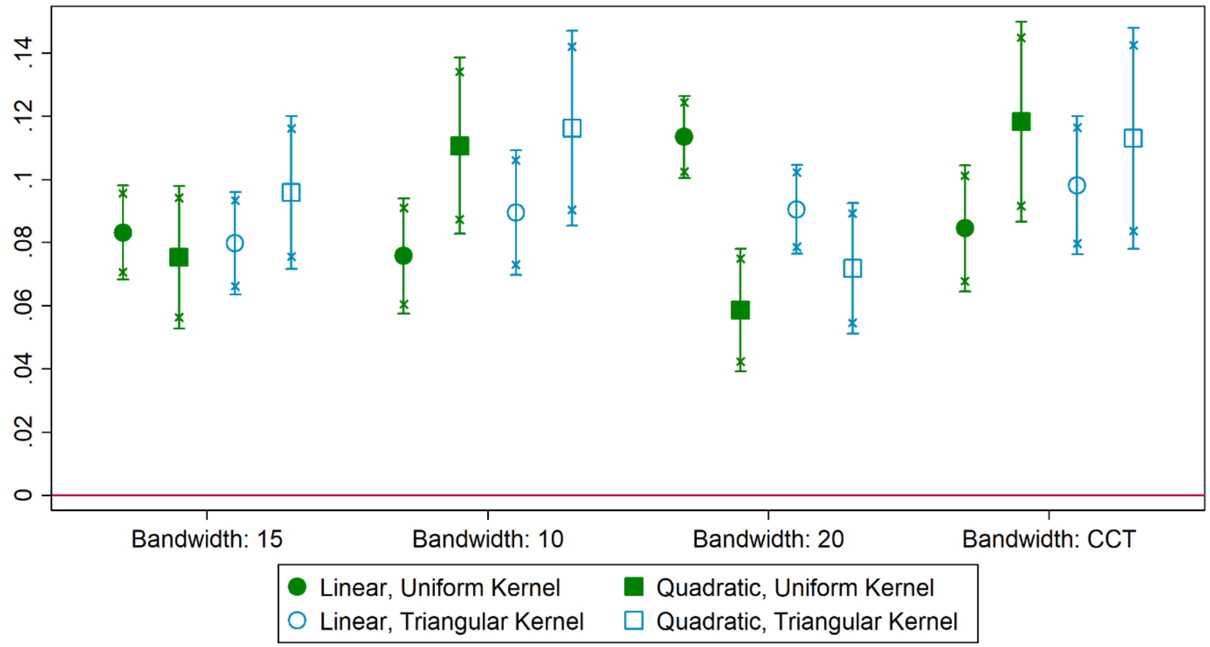
Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The sample consists of observations within the 15-point bandwidth around the cutoff. The dependent variable is an indicator of retaking the NCEE in the next year. Column (1) uses the sample of observations in high-quality high schools. Column (2) uses the sample of observations in low-quality high schools. Column (3) uses the sample of observations in high schools with high male-female sex ratio. Column (4) uses the sample of observations in high schools with low male-female sex ratio. Column (5) uses the sample of observations in counties with high male-female sex ratio. Column (6) uses the sample of observations in counties with low male-female sex ratio. Column (7) uses the sample of observations in counties with high GDP per capita. Column (8) uses the sample of observations in counties with low GDP per capita. All columns control for a linear function of the running variable and its interaction with the indicator of below the cutoff, and their gender interactions. Individual characteristics (ethnicity, hukou status, whether the individual is a first-time taker, age) and their gender interactions, as well as year-by-track-by-gender fixed effects are controlled in all columns. Standard errors are two-way clustered at the individual identifier level and the high school-year level.

Figure A1: Covariates Balancedness



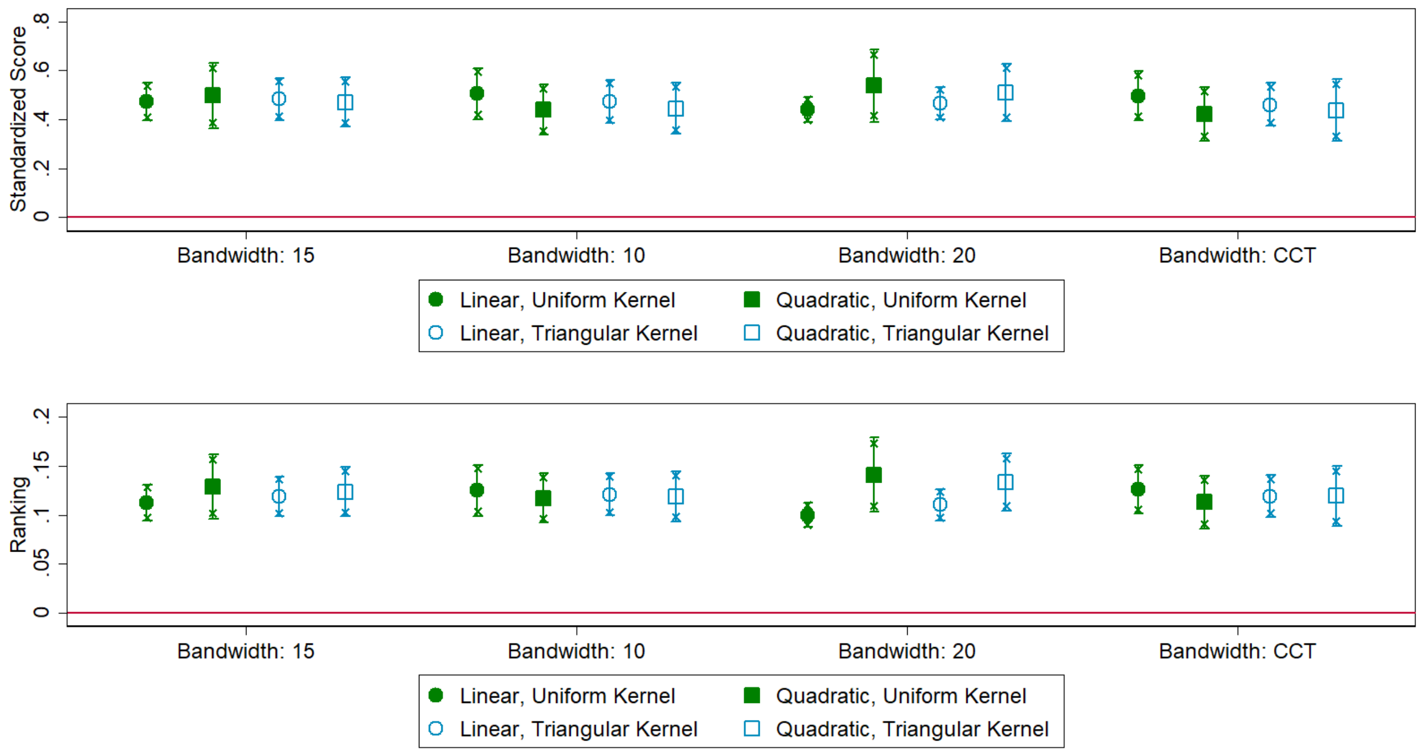
Notes: This figure plots the individual characteristics (gender, Han ethnicity, Hui ethnicity, hukou status, being a first-time taker, and age) against the distance to the tier-2 cutoff score. The sample consists of observations within the 15-point bandwidth around the cutoff. Each circle corresponds to one point in the test score. The straight lines represent the fitted linear functions to the left and to the right of the cutoff. The dashed lines represent the lower and upper bounds of the 95% confidence interval for the sample mean of the outcome variable within the corresponding bin.

Figure A2: Robustness: Effects of Below Tier-2 University Cutoff on Retake Probability



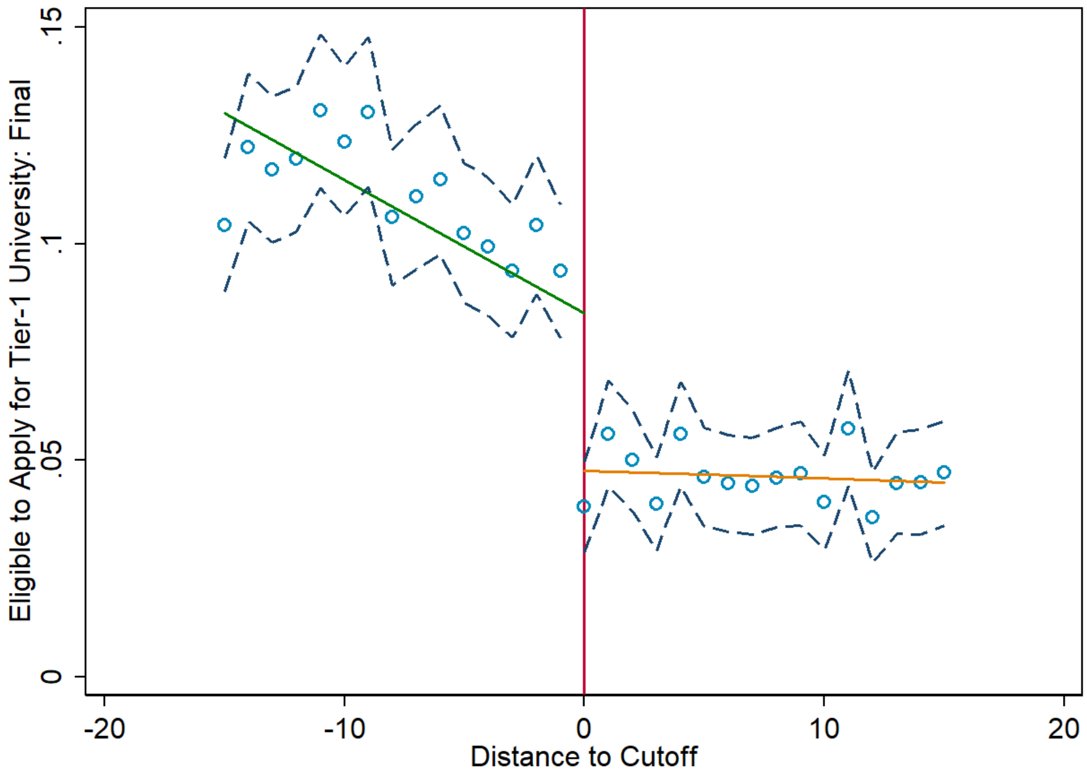
Notes: This figure plots the estimated effects of below the tier-2 cutoff on retake probability using different bandwidths (15-point, 10-point, 20-point, and the CCT optimal bandwidth (8.1-point) proposed by Calonico et al. (2014)) and specifications (linear and quadratic controls, uniform and triangular kernel weights). The sample consists of observations within the 15-point bandwidth around the cutoff. The parametric function of the running variable and its interaction with the indicator of below the cutoff are controlled in all regressions. Gender, ethnicity, hukou status, whether the individual is a first-time taker, age and year-by-track fixed effects are controlled in all regressions. Standard errors are two-way clustered at the individual identifier level and the high school-year level. “x” markers represent bounds of 90% confidence interval. “-” markers represent bounds of 95% confidence interval.

Figure A3: Robustness: Effects of Retake on Exam Outcomes



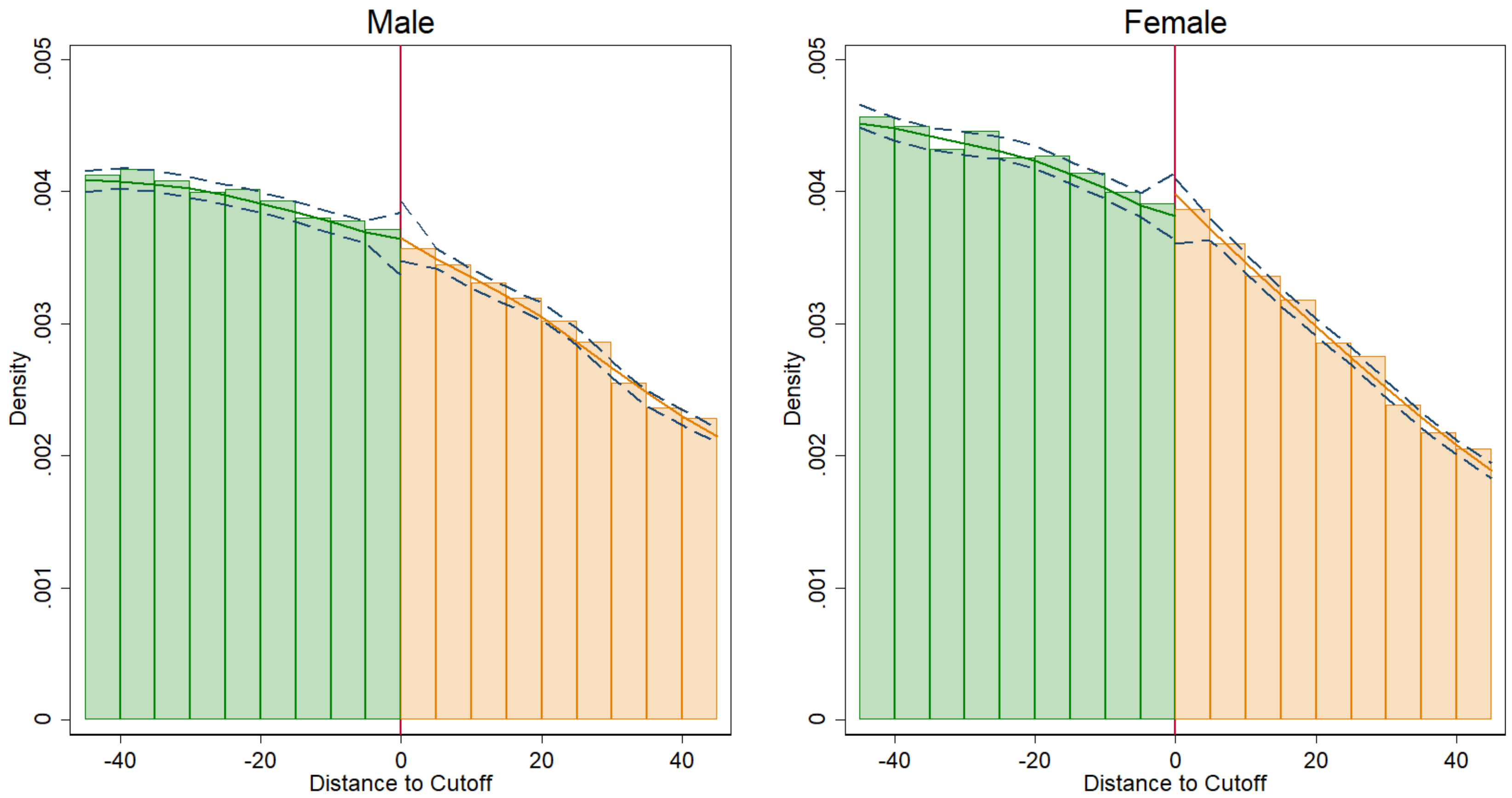
Notes: This figure plots the estimated effects of retaking the NCEE on exam outcomes using different bandwidths (15-point, 10-point, 20-point, and the CCT optimal bandwidth (8.1-point) proposed by Calonico et al. (2014)) and specifications (linear and quadratic controls, uniform and triangular kernel weights). The sample consists of observations within the 15-point bandwidth around the cutoff. The indicator of below the cutoff is used as an instrument for the indicator of retaking the NCEE in the 2SLS specification, with the differences between the final outcomes and the initial outcomes (standardized score, ranking within the year-track) as the dependent variables. The parametric function of the running variable and its interaction with the indicator of below the cutoff are controlled in all regressions. Individual characteristics (gender, ethnicity, hukou status, whether the individual is a first-time taker, age) and year-by-track fixed effects are controlled in all regressions. Standard errors are two-way clustered at the individual identifier level and the high school-year level. "x" markers represent bounds of 90% confidence interval. "-" markers represent bounds of 95% confidence interval.

Figure A4: Eligibility to Apply for Tier-1 University in Terms of Final Outcome vs. Distance to Tier-2 University Cutoff Score



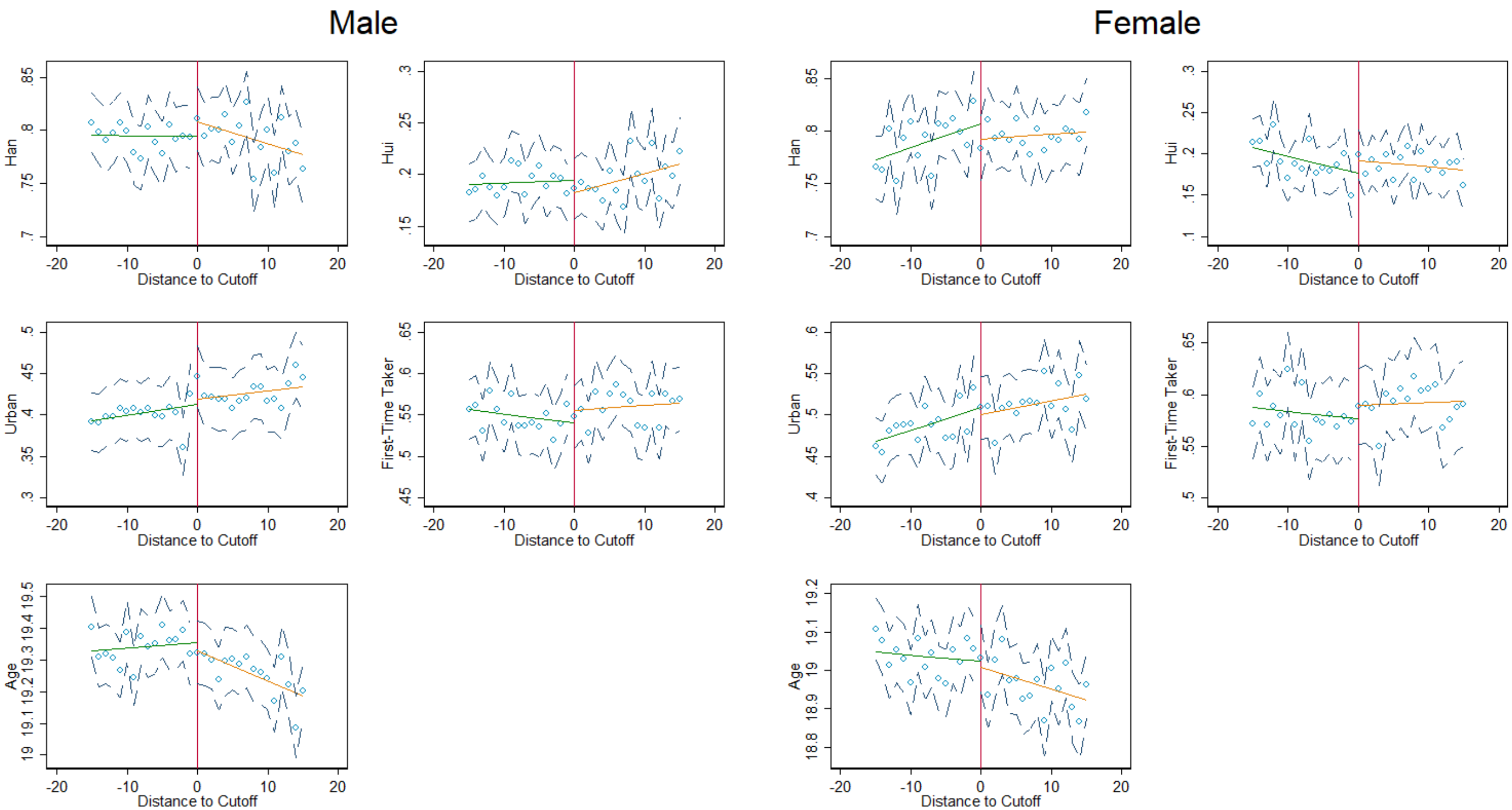
Notes: This figure plots the probability of being eligible to apply for tier-1 universities in terms of the final outcome, which is equal to the dependent variable in the current year if the individual does not retake the NCEE in the next year, and is equal to the dependent variable in the next year if the individual retakes the NCEE in the next year, against the distance to the tier-2 cutoff score. The individual is defined as eligible to apply for tier-1 universities if the score of the individual is above or equal to the tier-1 cutoff in the corresponding year-track. The sample consists of observations within the 15-point bandwidth around the cutoff. Each circle corresponds to one point in the test score. The straight lines represent the fitted linear functions to the left and to the right of the cutoff. The dashed lines represent the lower and upper bounds of the 95% confidence interval for the sample mean of the outcome variable within the corresponding bin.

Figure A5: Test of Running Variable Density Smoothness around Tier-2 Cutoff by Gender



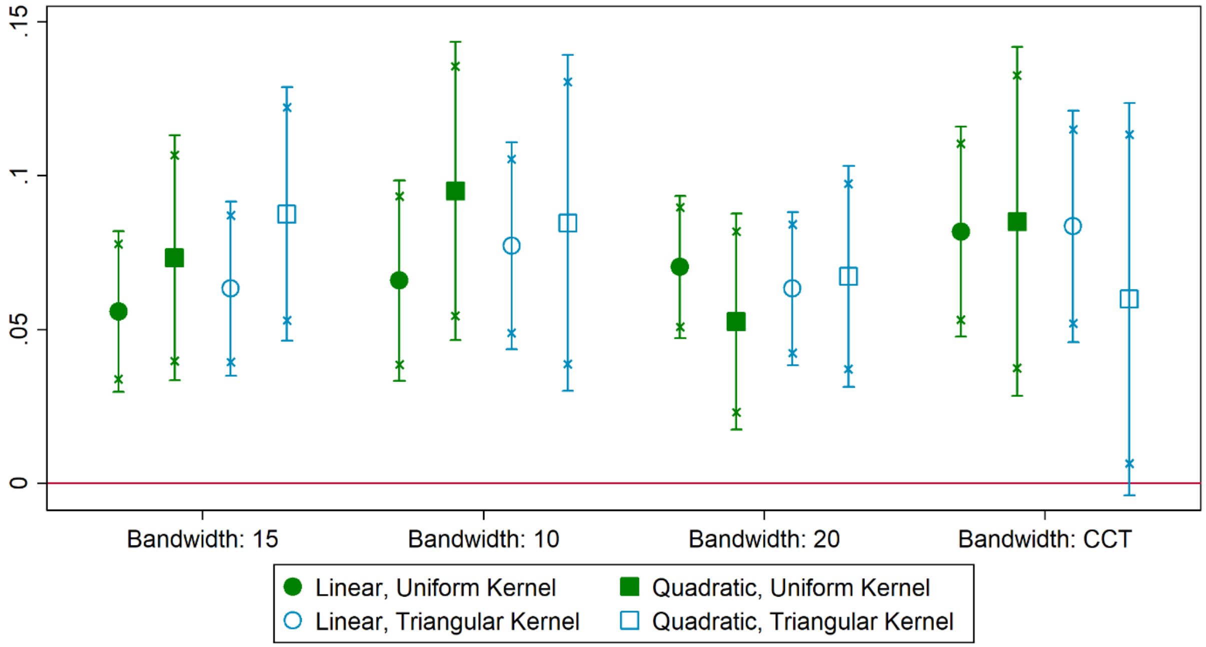
Notes: This figure plots the density of the running variable (the distance to the tier-2 cutoff score) following the manipulation testing procedure in Cattaneo, Jansson and Ma (2018) for males and females separately. The bars in this figure represent the density distribution of the running variable over 5-point bins. The straight lines represent the estimated density to the left and to the right of the cutoff using the local polynomial density estimators proposed in Cattaneo et al. (2020). The dashed lines represent the lower and upper bounds of the 95% confidence interval for the estimated density.

Figure A6: Covariates Balancedness by Gender



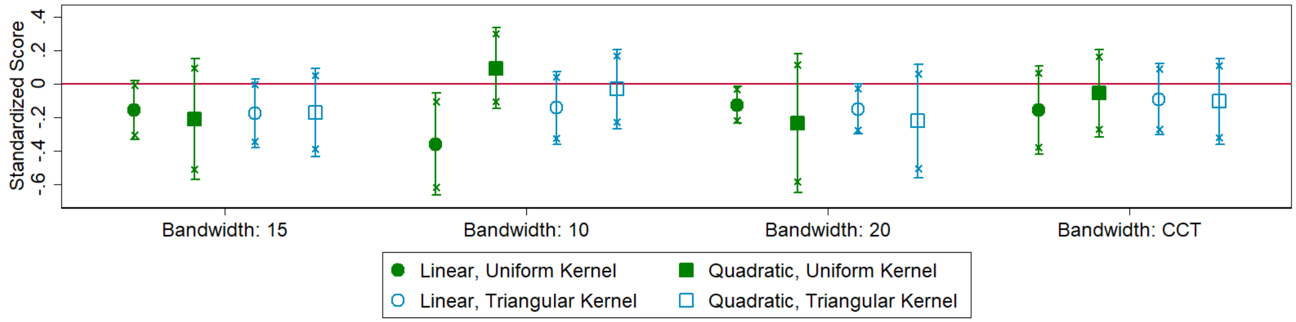
Notes: This figure plots the individual characteristics (Han ethnicity, Hui ethnicity, hukou status, being a first-time taker, and age) against the distance to the tier-2 cutoff score for males and females separately. The sample consists of observations within the 15-point bandwidth around the cutoff. Each circle corresponds to one point in the test score. The straight lines represent the fitted linear functions to the left and to the right of the cutoff. The dashed lines represent the lower and upper bounds of the 95% confidence interval for the sample mean of the outcome variable within the corresponding bin.

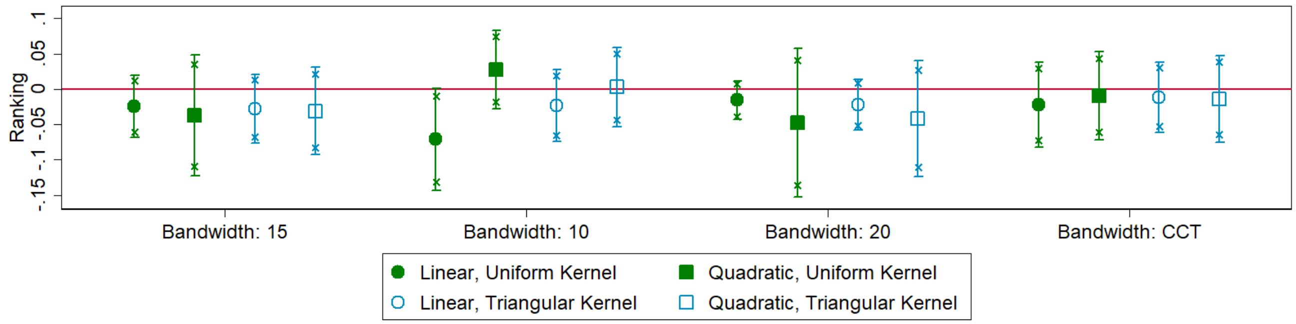
Figure A7: Robustness: Gender Differences in the Effects of Below Cutoff on Retake Probability



Notes: This figure plots the estimated gender differences in the effects of below the tier-2 cutoff on retake probability using different bandwidths (15-point, 10-point, 20-point, and the CCT optimal bandwidth (8.1-point) proposed by Calonico et al. (2014)) and specifications (linear and quadratic controls, uniform and triangular kernel weights). The sample consists of observations within the 15-point bandwidth around the cutoff. The parametric function of the running variable and its interaction with the indicator of below the cutoff, and their gender interactions are controlled in all regressions. Individual characteristics (ethnicity, hukou status, whether the individual is a first-time taker, age) and their gender interactions, as well as year-by-track-by-gender fixed effects, are also controlled in all regressions. Standard errors are two-way clustered at the individual identifier level and the high school-year level. “x” markers represent bounds of 90% confidence interval. “-” markers represent bounds of 95% confidence interval.

Figure A8: Robustness: Gender Differences in the Effects of Retake on Exam Outcomes





Notes: This figure plots the estimated gender differences in the effects of retaking the NCEE on exam outcomes using different bandwidths (15-point, 10-point, 20-point, and the CCT optimal bandwidth (8.1-point) proposed by Calonico et al. (2014)) and specifications (linear and quadratic controls, uniform and triangular kernel weights). The sample consists of observations within the 15-point bandwidth around the cutoff. The indicator of below the cutoff and its interaction with male dummy are used as instruments for the indicator of retaking the NCEE and its interaction with male dummy in the 2SLS specification, with the differences between the final outcomes and the initial outcomes (standardized score, ranking within the year-track) as the dependent variables. The parametric function of the running variable and its interaction with the indicator of below the cutoff, and their gender interactions are controlled in all regressions. Individual characteristics (ethnicity, hukou status, whether the individual is a first-time taker, age) and their gender interactions, as well as year-by-track-by-gender fixed effects, are also controlled in all regressions. Standard errors are two-way clustered at the individual identifier level and the high school-year level. “x” markers represent bounds of 90% confidence interval. “-” markers represent bounds of 95% confidence interval.

Table A1: Balancing Tests

Variables	(1) Male	(2) Han	(3) Hui	(4) Urban	(5) First-Time Taker	(6) Age
Panel A: Linear Control						
Below Cutoff	0.0146	0.0002	-0.0030	-0.0052	-0.0175*	0.0295
	(0.0092)	(0.0075)	(0.0073)	(0.0099)	(0.0103)	(0.0270)
Panel B: Quadratic Control						
Below Cutoff	0.0095	0.0056	-0.0086	-0.0159	-0.0100	0.0759**
	(0.0142)	(0.0116)	(0.0116)	(0.0154)	(0.0141)	(0.0373)
Observations	41,477	41,477	41,477	41,477	41,477	41,477
Bandwidth	15	15	15	15	15	15
Year-Track FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The sample consists of observations within the 15-point bandwidth around the cutoff. Panel A controls for a linear function of the running variable and its interaction with the indicator of below the cutoff. Panel B controls for a quadratic function of the running variable and its interaction with the indicator of below the cutoff. Year-by-track fixed effects are controlled in all columns. Standard errors are two-way clustered at the individual identifier level and the high school-year level.

Table A2: Effects of Below Tier-2 University Cutoff on Retake Probability, Alternative Clustering

	(1)	(2)	(3)	(4)
Variables		Retake		
Panel A: Baseline, Two-way Clustering by Individual Identifier and High School-Year				
Below Cutoff	0.0805***	0.0831***	0.0737***	0.0753***
	(0.0081)	(0.0076)	(0.0122)	(0.0115)
Panel B: Two-way Clustering by Individual Identifier and High School				
Below Cutoff	0.0805***	0.0831***	0.0737***	0.0753***
	(0.0079)	(0.0076)	(0.0121)	(0.0119)
Panel C: Two-way Clustering by Running Variable and Individual Identifier				
Below Cutoff	0.0805***	0.0831***	0.0737***	0.0753***
	(0.0089)	(0.0081)	(0.0148)	(0.0124)
Panel D: Two-way Clustering by Running Variable and High School-Year				
Below Cutoff	0.0805***	0.0831***	0.0737***	0.0753***
	(0.0090)	(0.0081)	(0.0142)	(0.0119)
Observations	41,477	41,477	41,477	41,477
Bandwidth	15	15	15	15
Interaction Controls	Linear	Linear	Quadratic	Quadratic
Individual Characteristics	No	Yes	No	Yes
Year-Track FE	Yes	Yes	Yes	Yes

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The sample consists of observations within the 15-point bandwidth around the cutoff. The dependent variable is an indicator of retaking the NCEE in the next year. Columns (1) and (2) control for a linear function of the running variable and its interaction with the indicator of below the cutoff. Columns (3) and (4) control for a quadratic function of the running variable and its interaction with the indicator of below the cutoff. Columns (1) and (3) do not control for individual characteristics. Columns (2) and (4) control for a set of individual characteristics, including gender, ethnicity, hukou status, whether the individual is a first-time taker, and age. Year-by-track fixed effects are controlled in all columns. In Panel A, standard errors are two-way clustered at the individual identifier level and the high school-year level. In Panel B, standard errors are two-way clustered at the individual identifier level and the high school level. In Panel C, standard errors are two-way clustered at the running variable level and the individual identifier level. In Panel D, standard errors are two-way clustered at the running variable level and the high school-year level.

Table A3: Summary Statistics of Exam Outcomes

	(1)	(2)	(3)	(4)
Sample	Full	[-15,15]	[-15,0)	[0,15]
Observations	362,592	41,477	21,123	20,354
	Mean (S.D.)			
Standardized Score: Initial	0.00	0.77	0.69	0.86
	(1.00)	(0.21)	(0.19)	(0.19)
Ranking: Initial	0.50	0.77	0.74	0.80
	(0.29)	(0.06)	(0.06)	(0.05)
Standardized Score: Final	0.13	0.86	0.82	0.90
	(1.03)	(0.30)	(0.33)	(0.25)
Ranking: Final	0.54	0.79	0.77	0.80
	(0.30)	(0.08)	(0.09)	(0.06)

Notes: This table shows the summary statistics of the exam outcomes. Initial outcomes are the variables in the current year. Final outcomes are the final payoffs of the outcomes, and are equal to the variables in the current year if the individual does not retake the NCEE in the next year, and are equal to the variables in the next year if the individual retakes the NCEE in the next year. Column (1) is for the full sample. Column (2) is for the sample within the 15-point bandwidth around the tier-2 cutoff. Column (3) is for the sample in column (2) that is below the cutoff. Column (4) is for the sample in column (2) that is above or equal to the cutoff.

Table A4: Predicting the NCEE Retakes

	(1)	(2)	(3)	(4)
Sample	Full	Full	[-15,15]	[-15,15]
Variables		Retake		
Male	0.0207***	0.0273***	0.0620***	0.0782***
	(0.0019)	(0.0019)	(0.0039)	(0.0037)
Han		0.1131***		0.1756***
		(0.0060)		(0.0125)
Hui		0.0415***		0.0100
		(0.0080)		(0.0132)
Urban		-0.1264***		-0.0599***
		(0.0055)		(0.0056)
First-Time Taker		0.1403***		0.1843***
		(0.0050)		(0.0065)
Age		-0.0248***		-0.0306***
		(0.0014)		(0.0018)
Standardized Score		-0.0161***		-1.1963***
		(0.0041)		(0.0282)
Observations	362,592	362,572	41,477	41,477
R-squared	0.004	0.055	0.013	0.207
Year-Track FE	Yes	Yes	Yes	Yes

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is an indicator of retaking the NCEE in the next year. Columns (1) and (2) use the full sample. Columns (3) and (4) use the sample within the 15-point bandwidth around the cutoff. Year-by-track fixed effects are controlled in all columns. Standard errors are two-way clustered at the individual identifier level and the high school-year level.

Table A5: Balancing Tests by Gender

Variables	(1) Han	(2) Hui	(3) Urban	(4) First-Time Taker	(5) Age
Panel A: Male, Linear Control					
Below Cutoff	-0.0181*	0.0132	-0.0174	-0.0169	0.0324
	(0.0105)	(0.0105)	(0.0134)	(0.0137)	(0.0381)
Panel B: Male, Quadratic Control					
Below Cutoff	-0.0034	-0.0037	-0.0429**	-0.0069	0.0775
	(0.0168)	(0.0166)	(0.0206)	(0.0204)	(0.0579)
Observations	21,162	21,162	21,162	21,162	21,162
Panel C: Female, Linear Control					
Below Cutoff	0.0200*	-0.0204**	0.0093	-0.0164	0.0164
	(0.0109)	(0.0102)	(0.0143)	(0.0147)	(0.0339)
Panel D: Female, Quadratic Control					
Below Cutoff	0.0163	-0.0149	0.0129	-0.0135	0.0700
	(0.0161)	(0.0154)	(0.0219)	(0.0209)	(0.0478)
Observations	20,315	20,315	20,315	20,315	20,315
Bandwidth	15	15	15	15	15
Year-Track FE	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The sample consists of observations within the 15-point bandwidth around the cutoff. Panels A and B are for the male sample. Panels C and D are for the female sample. Panels A and C control for a linear function of the running variable and its interaction with the indicator of below the cutoff. Panels B and D control for a quadratic function of the running variable and its interaction with the indicator of below the cutoff. Year-by-track fixed effects are controlled in all columns. Standard errors are two-way clustered at the individual identifier level and the high school-year level.

Table A6: Gender Differences in the Effects of Below Tier-2 University Cutoff on Retake Probability, Robustness

	(1)	(2)	(3)	(4)	(5)	(6)
Variables			Retake			
Male*Below Cutoff	0.0492***	0.0548***	0.0558***	0.0724***	0.0720***	0.0732***
	(0.0144)	(0.0133)	(0.0133)	(0.0218)	(0.0203)	(0.0203)
Below Cutoff	0.0545***	0.0553***	0.0549***	0.0360**	0.0384***	0.0377***
	(0.0098)	(0.0093)	(0.0093)	(0.0145)	(0.0138)	(0.0137)
Male	0.0453***	0.0584***		0.0334***	0.0465***	
	(0.0071)	(0.0072)		(0.0093)	(0.0097)	
Observations	41,477	41,477	41,477	41,477	41,477	41,477
Bandwidth	15	15	15	15	15	15
Interaction Controls	Linear	Linear	Linear	Quadratic	Quadratic	Quadratic
Individual Characteristics	No	Without Gender Interaction	With Gender Interaction	No	Without Gender Interaction	With Gender Interaction
Year-Track FE	Yes	Yes	By Gender	Yes	Yes	By Gender

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The sample consists of observations within the 15-point bandwidth around the cutoff. The dependent variable is an indicator of retaking the NCEE in the next year. Columns (1)-(3) control for a linear function of the running variable and its interaction with the indicator of below the cutoff, as well as their gender interactions. Columns (4)-(6) control for a quadratic function of the running variable and its interaction with the indicator of below the cutoff, as well as their gender interactions. Columns (1) and (4) control for year-by-track fixed effects. Columns (2) and (5) control for individual characteristics (ethnicity, hukou status, whether the individual is a first-time taker, age) and year-by-track fixed effects. Columns (3) and (6) control for individual characteristics (ethnicity, hukou status, whether the individual is a first-time taker, age) and their gender interactions, as well as year-by-track-by-gender fixed effects. Standard errors are two-way clustered at the individual identifier level and the high school-year level.