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

Le Kang

Nanjing University

Ziteng Lei

Renmin University of China

Yang Song

Colgate University

Peng Zhang

Chinese University of Hong Kong, Shenzhen, and Shenzhen Finance Institute

We document gender differences in reactions to failure in the National College Entrance Exam in China. Using administrative data from Ningxia province and a regression discontinuity design, we find that students who narrowly miss the tier-2 university cutoff exhibit an 8 percentage point increase in their likelihood of retaking the exam, and that retaking improves exam performance substantially. Notably, the response to this failure is much larger for men than for women. Survey evidence suggests that gender differences in psychological costs of retaking, parental education expectations, and some noncognitive traits can explain an important part of gender differences in retake willingness.

All authors contributed equally and are ordered alphabetically. We thank Ying Bai, Serena Canaan, Clément de Chaisemartin, Naijia Guo, Peter Kuhn, Shelly Lundberg, Pierre Mouganie, Gonzalo Vazquez-Bare, Huihua Xie, Shiyue Wang, and seminar and conference

Electronically published April 12, 2024

Journal of Political Economy Microeconomics, volume 2, number 2, May 2024.

https://doi.org/10.1086/729535

I. Introduction

Gender disparities in educational and labor market outcomes have attracted increasing attention. Previous studies have documented that gender differences in noncognitive 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 the 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 careers 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.

In this paper, we study how men and women respond differently to failures in the National College Entrance Exam (NCEE), an annual exam that solely determines the admission of almost all students into higher education institutions in China.[1] 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 college and for a less-selected group of individuals. 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 is 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

participants at the University of California, Santa Barbara, Jinan University, the Chinese University of Hong Kong, Peking University, the University of Vermont, the Sixth Economics of Education Workshop of the Institute of Labor Economics (IZA), the 2021 Asia Impact Evaluation Conference, the Thirteenth International Symposium on Human Capital and Labor Markets, the Society for Institutional and Organizational Economics Annual Conference, the Fifth Renmin University of China–Global Labor Organization Joint Research Conference on Chinese Labor Markets, the China Economics of Education Annual Conference, and the Forum on Empirical Education Research for helpful comments and suggestions. Zhang acknowledges support from the National Natural Science Foundation of China (grant no. 72203193) and the Natural Science Foundation of Guangdong Province, China (grant no. 2023B1515020065). Song acknowledges support from the Colgate University Research Council; Mind, Brain, and Behavior Initiative; and Public Affairs and Policy Research Initiative. This paper was edited by John List (editor) and Alec Brandon (associate editor).

[1] In 2013, e.g., nearly 95% of high school graduates participated in the NCEE nationwide.

get into college 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: 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 applications for universities in each tier.[2] We show evidence that these cutoffs are exogenously determined, and students cannot self-select around the cutoffs. Around 10 million students take the NCEE to compete for admission to highly selective universities each year, with only around 25% of students achieving scores that make them eligible to apply to the 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.[3] 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 narrowly miss the tier-2 cutoff, a signal of entering good universities and educational success, have an 8 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 improvements amount to a 2.7%–5.7% higher wage offer for the first job after college under a simple back-of-the-envelope analysis. 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.

[2] 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 NCEE scores above the tier cutoff can apply for universities in that specific tier. See sec. II for more discussion.

[3] 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 sec. III for more discussion.

We then focus on gender differences in reactions to the failure of missing 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.

In addition, we find that the gender differences in retake are large and of similar magnitude for individuals from urban and rural households, of different ethnicities, 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 remain consistent for diverse groups of individuals.

We then discuss several mechanisms that may help explain why women are less likely to retake the NCEE than men when scoring just below the cutoff, including gender differences in returns to retake, retake costs, noncognitive traits (such as causal attribution and locus of control), preferences and expectations, and family support. Among these potential mechanisms, females facing lower test score returns from retake is unlikely to be the explanation, as we find that the causal returns to retake in terms of exam outcomes for women are similar to or even higher than those for men, and such gender differences in returns cannot be explained by students rationally self-selecting into retake based on returns.

To further distinguish between the potential mechanisms, we collect survey data on high school students in Gansu province, a province that is geographically and economically close to Ningxia province.[4] We collect information on retake willingness and potential mechanisms and find that gender differences in psychological costs of retaking the NCEE and parental education expectations are important mechanisms behind the gender differences in reactions to failure. Gender differences in some noncognitive traits, such as confidence, competitiveness, and causal attribution, also explain part of our results. By contrast, gender differences in returns to retake and other psychological traits (such as risk preferences and locus of control) do not explain the gender gap in retaking.

[4] In terms of overall economy, both Ningxia and Gansu are among the less-developed provinces in China. In 2020, the gross regional product was 897.97 billion CNY in Gansu and 395.63 billion CNY in Ningxia. The per capita gross regional product was 35,848 CNY in Gansu and 55,021 CNY in Ningxia. The educational resources are also not very different in these two provinces. In 2020, the student-to-teacher ratio in high school was 11.21 in Gansu and 13.84 in Ningxia.

Our paper mainly contributes to two strands of the literature. First, we contribute to the broad literature on gender differences in educational choices and competitions (Niederle and Vesterlund 2007; Buser, Niederle, and Oosterbeek 2014; Flory, Leibbrandt, and List 2015; Berlin and Dargnies 2016; Buser, Peter, and Wolter 2017; Reuben, Wiswall, and Zafar 2017; Astorne-Figari and Speer 2019; Cai et al. 2019; Zhang 2019),[5] 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; Fang, Zhang, and Zhang 2021; Wasserman 2023). 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 United States (Ellison and Swanson 2018; Buser and Yuan 2019), in low-stakes Rubik's Cube competitions (Fang, Zhang, and Zhang 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 2023).

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 in 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. For example, Wasserman (2023) focuses on politicians, a group of individuals who survive extremely competitive careers and have unusual ambitions, and most of the general population would be excluded from the analysis. Landaud and Maurin (2020) focus on candidates of elite science graduate programs in France, which is also a selective sample.[6] Our setting, however, can greatly alleviate the concern of sample selection because almost everyone needs to take the NCEE to get into college in China. Our results can thus enhance the external validity of prior findings substantially. Finally, we provide rich discussions on the potential mechanisms for the gender gap using the administrative data and the unique survey data we collected. The regression discontinuity (RD)

[5] For example, Cai et al. (2019) look at gender difference in a high-stakes environment, comparing performances in a mock exam and those in the NCEE, as well as gender difference in the effect of negative shocks in morning exam performance on afternoon exam performance. Zhang (2019) studies how an egalitarian marriage reform, implemented on Han Chinese but not on the Yi minority, generated changes in female inclination to compete, as measured in a lab experiment.

[6] For example, over 82% of the observations in Landaud and Maurin (2020) are males, indicating that many females that are not devoted to entering science graduate programs, and also many people that do not pursue a graduate degree, are not represented by the sample.

results suggest that female students have similar, if not higher, returns in terms of exam performance improvement; thus this factor is unlikely an important driver of the gender differences in the tendency to retake. The survey data we collected provide a more detailed and direct way of measuring the explanatory power of various potential explanations. We find that parental educational expectations, psychological costs of retaking, and confidence in doing well when retaking are among the most important factors. Overall, these results contribute to the literature by providing a better understanding of the potential mechanisms of gender differences in reactions to failure.

Second, we contribute to the literature on the causal effects of exam retakes, particularly in high-stakes settings that are admission relevant (Krishna, Lychagin, and Frisancho 2018; Zhang et al. 2019; Goodman, Gurantz, and Smith 2020). Many countries use standardized tests such as the SAT and ACT to rank students for college admission, yet limited research has delved into students' decisions to retake these exams. One exception is Goodman, Gurantz, and Smith (2020), who estimate the effects of retaking the SAT using discontinuous jumps in retake probability at multiples of 100, and find that retaking substantially improves SAT scores and 4-year college enrollment rates. Interestingly, Goodman, Gurantz, and Smith (2020) find that females are more likely to retake the SAT than males, which differs from our results. While the SAT and the NCEE are both deciding factors for college admission, crucial distinctions exist between our setting and that of Goodman, Gurantz, and Smith (2020). The most important difference is that the NCEE is an annual exam, while the SAT can potentially be taken multiple times a year. Therefore, the sense of competition may be more pronounced in our setting, and the cost of retaking the NCEE is considerably higher. In other words, the key factors contributing to gender differences in retake behavior that we find, such as the psychological costs associated with retaking the exam and competitiveness, may play a much less important role in the context of SAT retaking, which entails lower costs and presumably lower competitiveness. These distinctions may help elucidate why our results diverge from those of Goodman, Gurantz, and Smith (2020). Nevertheless, our findings show that despite the high opportunity cost of waiting for another year, retaking the NCEE yields substantial returns in terms of exam performance that are consequential for admission outcomes. More interestingly, while females are less likely to retake the NCEE than males, the returns for females are similar to or sometimes even higher than for males. Our comprehensive results, based on both administrative and survey data, also advance our understanding of the potential mechanisms underpinning gender differences in retake behavior.

At the end of this section, we follow the SANS conditions, that is, selection, attrition, naturalness, and scaling, as outlined by List (2020), to provide a brief discussion on the external validity of our empirical findings.[7] First, regarding selection, that is, representativeness of the studied group compared to the underlying population, our sample is the universe of NCEE takers in Ningxia province, and there is no selection problem within our target population in Ningxia. In addition, despite that Ningxia is a relatively smaller and less-developed province in China, the content of high school education and the NCEE is not so different from that in other provinces, and important demographics of the NCEE takers, such as gender ratio and urban-rural ratio, are similar to those of NCEE takers nationwide.[8] Second, concerning attrition, that is, attrition and noncompliance of subjects under treatment, our setting faces no attrition problem, as almost every high school student has to take the NCEE to get into college. Third, regarding naturalness, that is, naturalness of the choice task, setting, and time frame, our empirical context is the NCEE, for which students have done practice exams many times. This context is thus well within the realm of students' experience and should not be considered unnatural. Finally, in terms of scaling, that is, program effects if the program is scaled, our results show that the gender differences in reactions to failure manifest across diverse groups of individuals, suggesting that they are likely to remain consistent at a larger scale.

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

II. Institutional Background and Data

A. Institutional Background

The NCEE, which is also commonly known as *gaokao*, is an annual examination held on June 7 and 8 that determines the admission of almost all students into higher education institutions in China.[9] The NCEE 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.[10] More than 2,000 universities in China

[7] An example of using the SANS conditions to discuss the generalizability of empirical results is Holz et al. (2020).

[8] For example, the proportion of male NCEE takers is 47.9% in Ningxia, similar to Beijing (47.3%) and Shanghai (47.4%), slightly smaller than national average (49.7%) in 2013. In addition, the proportion of urban NCEE takers is 44.3% in Ningxia, slightly larger than national average (40.2%).

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

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

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 the art (social science) track after the tenth 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 the science track or the 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 4-year universities that grant bachelor's degrees, whereas tertiary technical colleges mostly offer programs lasting 2–3 years. Admission into tier-1 and tier-2 universities is generally considered an educational success, while admission into tier-3 universities or tertiary technical colleges is often considered 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.[11] The college assignment is organized sequentially by tier: tier-1 universities first finish their assignment, then tier-2 universities recruit, followed by tier-3 universities and tertiary

[11] 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, Kang, and Song (2020) for more discussion on the timing of the college application submission in China.

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 superselective 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 one can take the NCEE, but taking the NCEE more than two times is rare.

B. Data

Our administrative data include the test scores and demographic information for the universe of NCEE takers in Ningxia province (or the 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.[12] Ningxia is a small province in China, with a population of around 7 million. Recently, there have been 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.[13] We also hand-collect the year-track cutoff points for the tier-1 and tier-2 universities in Ningxia province from publicly available records.[14]

To identify whether NCEE takers retake the exam in the following year and their exam performance, we match observations in the 2 consecutive years based on the name identifier (which uniquely identifies a full

[12] The test score discussed in this paper is the total score for admission purposes, 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, this will not confound the decision to retake.

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

[14] Admission to tier-3 universities is much less competitive since 40%–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.

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 retaken the NCEE in the next year.[15] 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.[16]

III. Cutoff-Induced NCEE Retakes and the Effects on Exam Outcomes

A. *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 IV. 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.[17] 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

[15] 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, because the hukou restrictions for the NCEE takers prevent students from arbitrarily choosing the province in which to take the NCEE.

[16] The 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 2010 retake the exam in the next year or not.

[17] Note that the test score is standardized only when used as an outcome variable, and its standardization is clearly noted as “standardized score” in the figures and tables.

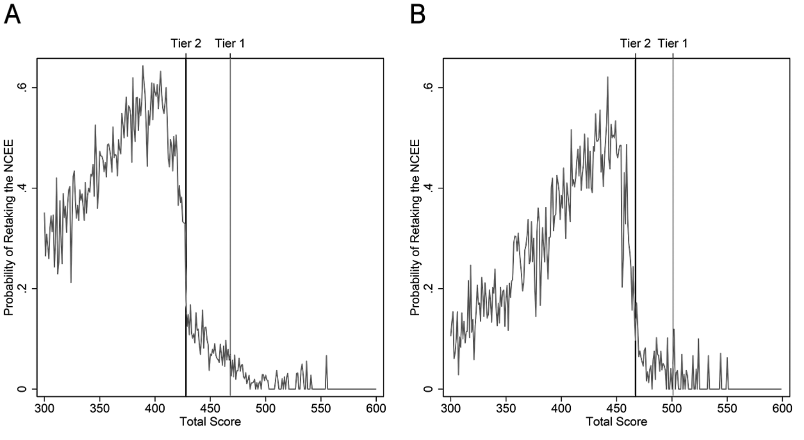

FIG. 1.—Plot of the probability of retaking the NCEE in the next year against the test score for NCEE takers of the year 2009, for the science track (*A*) and the art track (*B*), separately. The retake probability measures the proportion of NCEE takers at each score that choose to retake in the next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

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 III.B.

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 the science track and the art track separately. The retake probability measures the proportion 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, and they are shown in figures A1 and A2 (figs. A1–A11 are available online). 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.[18] 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:

[18] 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 more selective and only 10% of students score above the cutoff.

$$\begin{aligned}\text{Retake}_{i,y,\text{tr}} = {} & \beta I(\text{Score}_{i,y,\text{tr}} < \text{Cutoff}_{y,\text{tr}}) + \gamma_1 f(\text{Score}_{i,y,\text{tr}} - \text{Cutoff}_{y,\text{tr}}) \\ & + \gamma_2 I(\text{Score}_{i,y,\text{tr}} < \text{Cutoff}_{y,\text{tr}}) \times f\left(\text{Score}_{i,y,\text{tr}} - \text{Cutoff}_{y,\text{tr}}\right) \\ & + \theta X_i + \mu_{y,\text{tr}} + \varepsilon_{i,y,\text{tr}},\end{aligned} \tag{1}$$

where $\text{Retake}_{i,y,\text{tr}}$ is a binary indicator for whether individual i in year y and track tr (science or art) retakes the NCEE next year. $\text{Score}_{i,y,\text{tr}}$ is the test score of individual i, and $\text{Cutoff}_{y,\text{tr}}$ is the cutoff score for tier-2 university admission that varies across year-track. The indicator function $I(\text{Score}_{i,y,\text{tr}} < \text{Cutoff}_{y,\text{tr}})$ equals 1 if the test score is below the cutoff. We include a function of the running variable, $\text{Score}_{i,y,\text{tr}} - \text{Cutoff}_{y,\text{tr}}$, 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 nonparametric estimation and inference procedure (Calonico, Cattaneo, and Titiunik 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 $\mu_{y,\text{tr}}$ 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.[19] For example, if a student retakes once after her first take in the NCEE, and her scores are within the 15-point bandwidth around the cutoffs in both years, then she will enter the estimation sample twice.[20] 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).

B. 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

[19] 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.

[20] 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.

TABLE 1
SUMMARY STATISTICS

	Full (1)	[−15, 15] (2)	[−15, 0) (3)	[0, 15] (4)
Observations	362,592	41,477	21,123	20,354
Male	.52	.51	.51	.51
	(.50)	(.50)	(.50)	(.50)
Ethnicity: Han	.78	.79	.79	.79
	(.41)	(.41)	(.41)	(.40)
Ethnicity: Hui	.20	.19	.19	.19
	(.40)	(.39)	(.39)	(.39)
Urban	.45	.46	.44	.47
	(.50)	(.50)	(.50)	(.50)
First-time taker	.73	.57	.56	.58
	(.44)	(.50)	(.50)	(.49)
Age	19.15	19.16	19.19	19.12
	(1.23)	(1.24)	(1.24)	(1.23)
Retake	.28	.20	.31	.08
	(.45)	(.40)	(.46)	(.27)

NOTE.—This table shows the summary statistics (means, with standard deviations in parentheses) 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 col. 2 that is below the cutoff. Column 4 is for the sample in col. 2 that is above or equal to the cutoff.

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 .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 cannot sort around the cutoffs.

We also plot the individual characteristics of students against the distance to the cutoff in figure A3. There is no substantial discontinuous jump for these predetermined characteristics at the cutoff. The estimation results for the balancing tests are shown in table A1 (tables A1–A9 are available online). Indeed, there is no consistent evidence showing

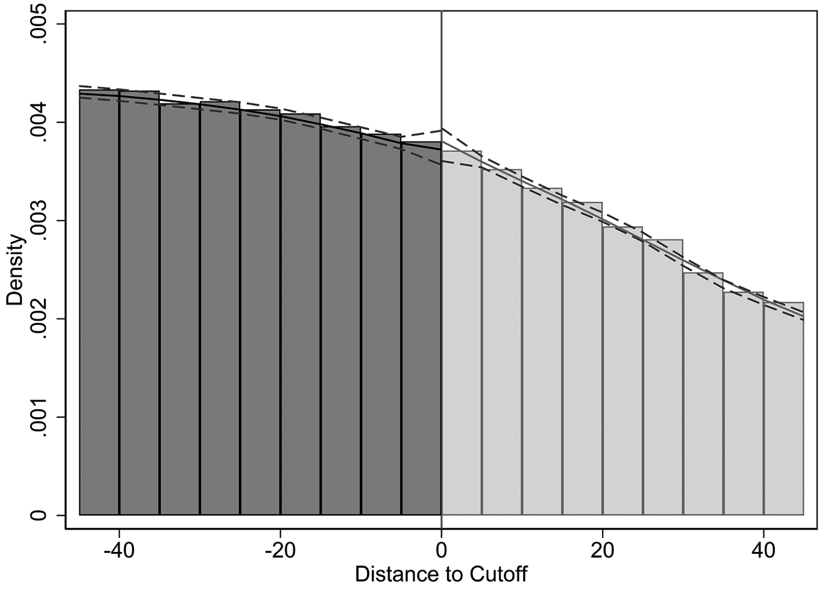

FIG. 2.—Plot of the density of the running variable (the distance to the tier-2 cutoff score) following the manipulation testing procedure in Cattaneo, Jansson, and Ma (2018). The bars represent the density distribution of the running variable over 5-point bins. The solid curves represent the estimated density to the left and to the right of the cutoff using the local polynomial density estimators proposed in Cattaneo, Jansson, and Ma (2020). The dashed curves represent the lower and upper bounds of the 95% confidence interval for the estimated density.

that a predetermined characteristic has a substantial discontinuity at the cutoff for both linear and quadratic control specifications.[21] As explained above, students cannot sort around the cutoffs because of the institutional setting, and there is no reason that students of certain characteristics are more 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.[22] 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

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

[22] The Stata package *rdplot* is used for the regression discontinuity plots. See Calonico, Cattaneo, and Titiunik (2015) and Calonico et al. (2017) for details.

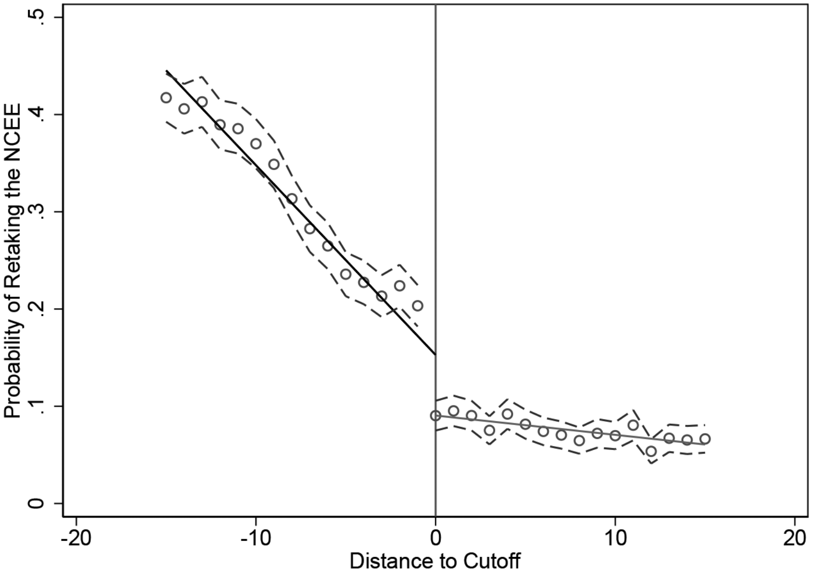

Fig. 3.—Plot of the probability of retaking the NCEE in the next year against the distance to the tier-2 cutoff score. The sample consists of observations within the 15-point bandwidth around the cutoff. Each circle corresponds to 1 point in the test score. The curves represent the fitted linear functions to the left and to the right of the cutoff. The dashed curves represent the lower and upper bounds of the 95% confidence interval for the sample mean of the outcome variable within the corresponding bin.

from 20% to 40% below the cutoff point. The estimated discontinuity effect without any covariates is 0.081 when using the local polynomial nonparametric estimation and inference procedure in Calonico, Cattaneo, and Titiunik (2014), with a robust 95% confidence interval [0.051, 0.095]. Table 2 presents the results using the parametric specification (eq. [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 8 percentage points, which is almost a 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 A4 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 the 10-point, 20-point, and data-driven optimal bandwidth (Calonico, Cattaneo, and Titiunik 2014), as well as using the triangular kernel weights instead of the uniform kernel weights in the

TABLE 2
Effects of Below Tier-2 University Cutoff on Retake Probability

	Dependent Variable: Retake			
	(1)	(2)	(3)	(4)
Below cutoff	.0805***	.0831***	.0737***	.0753***
	(.0081)	(.0076)	(.0122)	(.0115)
Observations	41,477	41,477	41,477	41,477
R^2	.117	.220	.117	.220
Bandwidth	15	15	15	15
Interaction controls	Linear	Linear	Quadratic	Quadratic
Individual characteristics	No	Yes	No	Yes
Year-track fixed effects	Yes	Yes	Yes	Yes

Note.—Standard errors in parentheses. 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.

*** $p < .01$.

baseline.[23] 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.

One interesting pattern in figure 3 is that the retake probability declines as the test score approaches the cutoff from below. One may rather expect an increase in retake probability because the regret of missing the cutoff may be larger when getting closer to the cutoff from below. However, a higher test score also implies better college admission options in the current year—as the tier-3 or worse colleges may still differ in quality and other characteristics, such as college and major reputation, and location. Therefore, students who have higher test scores would have better outside options other than retaking the exam and are supposed to have a lower likelihood of retaking the exam, which appears to be the dominant effect empirically. Note that the negative slope pattern is not unique to our study—Landaud and Maurin (2020) also find that the retake probability declines as the ranking increases on both sides of the cutoff in a similar regression discontinuity design investigating the entrance exam of highly selected elite science graduate programs in France. Therefore, the slopes may be negative below the cutoff, especially

[23] The CCT optimal bandwidth (Calonico, Cattaneo, and Titiunik 2014) is 8.1 points when using the uniform kernel weights.

in settings of educational selection systems based on exams where the running variable is positively associated with the outside options. In addition, all these effects only matter for the slopes around the cutoff but not the discontinuity at the cutoff, and should not confound our RD design as we are comparing students just below and just above the cutoff only.

C. *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:

$$\begin{aligned} Y^I_{i,y,\mathrm{tr}} = {} & \beta^I I(\mathrm{Score}_{i,y,\mathrm{tr}} < \mathrm{Cutoff}_{y,\mathrm{tr}}) + \gamma_1 f(\mathrm{Score}_{i,y,\mathrm{tr}} - \mathrm{Cutoff}_{y,\mathrm{tr}}) \\ & + \gamma_2 I(\mathrm{Score}_{i,y,\mathrm{tr}} < \mathrm{Cutoff}_{y,\mathrm{tr}}) \times f(\mathrm{Score}_{i,y,\mathrm{tr}} - \mathrm{Cutoff}_{y,\mathrm{tr}}) \\ & + \theta X_i + \mu_{y,\mathrm{tr}} + \varepsilon_{i,y,\mathrm{tr}}, \end{aligned} \tag{2}$$

$$\begin{aligned} Y^F_{i,y,\mathrm{tr}} = {} & \beta^F I(\mathrm{Score}_{i,y,\mathrm{tr}} < \mathrm{Cutoff}_{y,\mathrm{tr}}) + \gamma_1 f(\mathrm{Score}_{i,y,\mathrm{tr}} - \mathrm{Cutoff}_{y,\mathrm{tr}}) \\ & + \gamma_2 I(\mathrm{Score}_{i,y,\mathrm{tr}} < \mathrm{Cutoff}_{y,\mathrm{tr}}) \times f(\mathrm{Score}_{i,y,\mathrm{tr}} - \mathrm{Cutoff}_{y,\mathrm{tr}}) \\ & + \theta X_i + \mu_{y,\mathrm{tr}} + \varepsilon_{i,y,\mathrm{tr}}, \end{aligned} \tag{3}$$

where $Y^I_{i,y,\mathrm{tr}}$ is the outcome Y in the first year of this 2-year period, which is referred to as the "initial outcome," and $Y^F_{i,y,\mathrm{tr}}$ is the final outcome Y over this 2-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."[24] 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 ($\beta^I = 0$).

[24] 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 the SAT, the NCEE can only be taken once per year, and the decision to retake is better modeled as a sequential decision in each year. In addition, taking the NCEE more than two times is rare—only around 4% of the individuals appear in our sample more than two times.

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 $Y^F_{i,y,\mathrm{tr}} - Y^I_{i,y,\mathrm{tr}}$ 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.[25] We use the specification with $Y^F_{i,y,\mathrm{tr}} - Y^I_{i,y,\mathrm{tr}}$ 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 who are not driven by falling below the cutoff (see sec. IV for further discussion).

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

$$\begin{aligned} Y^F_{i,y,\mathrm{tr}} - Y^I_{i,y,\mathrm{tr}} = {} & \beta_{\mathrm{IV}}\mathrm{Retake}_{i,y,\mathrm{tr}} + \gamma_1 f(\mathrm{Score}_{i,y,\mathrm{tr}} - \mathrm{Cutoff}_{y,\mathrm{tr}}) \\ & + \gamma_2 I(\mathrm{Score}_{i,y,\mathrm{tr}} < \mathrm{Cutoff}_{y,\mathrm{tr}}) \times f(\mathrm{Score}_{i,y,\mathrm{tr}} - \mathrm{Cutoff}_{y,\mathrm{tr}}) \qquad (4) \\ & + \theta X_i + \mu_{y,\mathrm{tr}} + \varepsilon_{i,y,\mathrm{tr}}, \end{aligned}$$

where $\mathrm{Retake}_{i,y,\mathrm{tr}}$ is instrumented by the indicator $I(\mathrm{Score}_{i,y,\mathrm{tr}} < \mathrm{Cutoff}_{y,\mathrm{tr}})$ 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 panels of the figure plot the initial outcomes of standardized score and ranking, and the right panels plot 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 (eqq. [2]–[4]). First we show the effects on initial outcomes; next we show the effects

[25] By definition, $Y^I_{i,y,\mathrm{tr}}$ and $Y^F_{i,y,\mathrm{tr}}$ 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 $Y^F_{i,y,\mathrm{tr}} - Y^I_{i,y,\mathrm{tr}}$ or $Y^F_{i,y,\mathrm{tr}}$ as the dependent variable identify the same parameter of interest, which is confirmed in table 3.

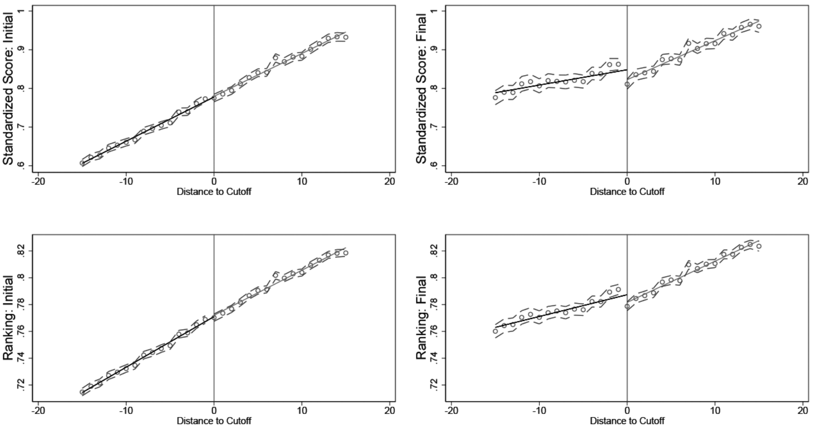

FIG. 4.—Plot of the exam outcomes against the distance to the tier-2 cutoff score. The top panels use the standardized score as the outcome variable. The bottom panels use the ranking within the year-track as the outcome variable. The left panels show the initial outcomes, that is, the dependent variables in the current year. The right panels show the final outcomes, that is, 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 1 point in the test score. The solid curves represent the fitted linear functions to the left and to the right of the cutoff. The dashed curves represent the lower and upper bounds of the 95% confidence interval for the sample mean of the outcome variable within the corresponding bin.

on final outcomes. There is little evidence on the effects on initial exam outcomes.[26] 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. Then we show the effects on the differences between the final and initial outcomes, which can be interpreted as the reduced-form estimates, that is, the effects of falling below the cutoff on the improvement in exam performance, and the estimates are almost identical to the effects on final outcomes. Finally, we show the 2SLS estimates of the effects of retaking the NCEE on the improvement of exam outcomes, where we use the indicator of falling below the cutoff as an instrumental variable for retaking. The first-stage Kleibergen-Paap (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

[26] There is one statistically significant coefficient in col. 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.

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

	Standardized Score		Ranking	
	(1)	(2)	(3)	(4)
Initial outcomes:				
Below cutoff	−.0001	−.0001	−.0004***	−.0000
	(.0001)	(.0002)	(.0001)	(.0001)
Final outcomes:				
Below cutoff	.0392***	.0374***	.0090***	.0097***
	(.0046)	(.0073)	(.0011)	(.0018)
Differences in outcomes:				
Below cutoff	.0393***	.0375***	.0094***	.0097***
	(.0046)	(.0073)	(.0011)	(.0018)
Differences in outcomes, 2SLS:				
Retake	.4730***	.4978***	.1127***	.1293***
	(.0389)	(.0679)	(.0094)	(.0169)
First-stage KP F-statistic	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 fixed effects	Yes	Yes	Yes	Yes

Note.—Standard errors in parentheses. The sample consists of observations within the 15-point bandwidth around the cutoff. The dependent variable in cols. 1 and 2 is the standardized score. The dependent variable in cols. 3 and 4 is the ranking within the year-track. We first show the results using the initial outcomes, i.e., the dependent variables in the current year. Next we show 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. Then we show the results using the differences between the final outcomes and the initial outcomes as the dependent variables. Finally, we use these same dependent variables, but we use a 2SLS specification and use 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.

*** $p < .01$.

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 A5 plots the 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. 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.[27] 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 A6, 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, that is, exam scores in the next year for retakers and in the initial year for non-retakers. These results show that despite both being 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 (eq. [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.[28] Moreover, these improvements are also consequential for their labor market opportunities—Jia and Li (2021) show that being eligible to apply for tier-1 universities translates to a 5.2%–9.2% higher wage offer for the first job after college. Therefore, under a simple back-of-the-envelope analysis, our estimates suggest that retaking the NCEE increases the first-job wage by around 2.7%–5.7% for students around the tier-2 cutoff.

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 deviation increase) are comparable to and even larger than the estimates of the causal effects of retaking the SAT on the admission-relevant superscore (an increase of around 0.34 standard deviations) in Goodman, Gurantz, and Smith (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

[27] 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.

[28] The coefficients (not reported) are all statistically significant at the 1% level.

than the general population but who still have much room for improvement—and should be carefully interpreted when extrapolating the effects to the general population of all students.

IV. 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 discuss the mechanisms and explanations for these gender differences.

A. 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:

$$\begin{aligned}\text{Retake}_{i,y,\text{tr}} = {} & \alpha\text{Male}_i + \beta I(\text{Score}_{i,y,\text{tr}} < \text{Cutoff}_{y,\text{tr}}) + \delta\text{Male}_i \\ & \times I(\text{Score}_{i,y,\text{tr}} < \text{Cutoff}_{y,\text{tr}}) + \gamma_1 f(\text{Score}_{i,y,\text{tr}} - \text{Cutoff}_{y,\text{tr}}) \\ & + \rho_1\text{Male}_i \times f(\text{Score}_{i,y,\text{tr}} - \text{Cutoff}_{y,\text{tr}}) \\ & \times \gamma_2 I(\text{Score}_{i,y,\text{tr}} < \text{Cutoff}_{y,\text{tr}}) \\ & \times f(\text{Score}_{i,y,\text{tr}} - \text{Cutoff}_{y,\text{tr}}) + \rho_2\text{Male}_i \\ & \times I(\text{Score}_{i,y,\text{tr}} < \text{Cutoff}_{y,\text{tr}}) \\ & \times f(\text{Score}_{i,y,\text{tr}} - \text{Cutoff}_{y,\text{tr}}) + \theta_1 X_i + \theta_2\text{Male}_i \\ & \times X_i + \mu_{y,\text{tr},\text{male}} + \varepsilon_{i,y,\text{tr}},\end{aligned} \tag{5}$$

where Male_i is a binary indicator of being male. With full gender interactions, the slopes are allowed to be different to the left and right of the cutoff, and to 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:

$$\begin{aligned} Y^{F}_{i,y,\mathrm{tr}} - Y^{I}_{i,y,\mathrm{tr}} = {} & \alpha \mathrm{Male}_i + \beta_{\mathrm{IV}} \mathrm{Retake}_{i,y,\mathrm{tr}} + \delta_{\mathrm{IV}} \mathrm{Male}_i \\ & \times \mathrm{Retake}_{i,y,\mathrm{tr}} + \gamma_1 f(\mathrm{Score}_{i,y,\mathrm{tr}} - \mathrm{Cutoff}_{y,\mathrm{tr}}) \\ & + \rho_1 \mathrm{Male}_i \times f(\mathrm{Score}_{i,y,\mathrm{tr}} - \mathrm{Cutoff}_{y,\mathrm{tr}}) \\ & \times \gamma_2 I(\mathrm{Score}_{i,y,\mathrm{tr}} < \mathrm{Cutoff}_{y,\mathrm{tr}}) \times f(\mathrm{Score}_{i,y,\mathrm{tr}} - \mathrm{Cutoff}_{y,\mathrm{tr}}) \\ & + \rho_2 \mathrm{Male}_i \times I(\mathrm{Score}_{i,y,\mathrm{tr}} < \mathrm{Cutoff}_{y,\mathrm{tr}}) \\ & \times f(\mathrm{Score}_{i,y,\mathrm{tr}} - \mathrm{Cutoff}_{y,\mathrm{tr}}) + \theta_1 X_i + \theta_2 \mathrm{Male}_i \times X_i \\ & + \mu_{y,\mathrm{tr},\mathrm{male}} + \varepsilon_{i,y,\mathrm{tr}}, \end{aligned} \tag{6}$$

where $\mathrm{Retake}_{i,y,\mathrm{tr}}$ and $\mathrm{Male}_i \times \mathrm{Retake}_{i,y,\mathrm{tr}}$ are instrumented by $I(\mathrm{Score}_{i,y,\mathrm{tr}} < \mathrm{Cutoff}_{y,\mathrm{tr}})$ and $\mathrm{Male}_i \times I(\mathrm{Score}_{i,y,\mathrm{tr}} < \mathrm{Cutoff}_{y,\mathrm{tr}})$ 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.

B. *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 and 2 show the results for our full sample. The results show that males are 2–3 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 and 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 6–8 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

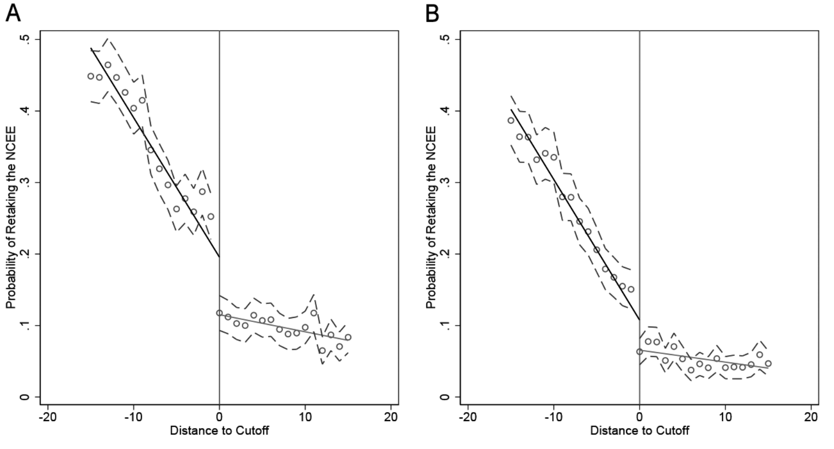

FIG. 5.—Plot of the probability of retaking the NCEE in the next year against the distance to the tier-2 cutoff score, separately for males (*A*) and females (*B*). The sample consists of observations within the 15-point bandwidth around the cutoff. Each circle corresponds to 1 point in the test score. The solid curves represent the fitted linear functions to the left and to the right of the cutoff. The dashed curves represent the lower and upper bounds of the 95% confidence interval for the sample mean of the outcome variable within the corresponding bin.

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 the 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 A7, and there is no evidence of discontinuous density in test scores around the tier-2 cutoff for males or females.[29] We also plot the individual characteristics of students against the distance to the cutoff for males and females separately in figure A8, and the estimation results for the balancing tests are shown in table A5. Again, there is no substantial discontinuous jump for these predetermined characteristics at the cutoff for both males and females under both linear and quadratic controls, which reassures the validity of our research design.[30]

[29] The *p*-value of the manipulation testing procedure proposed by Cattaneo, Jansson, and Ma (2018) is .61 for males and .84 for females.

[30] There is one coefficient significant at the 10% level (Han) when using the linear control specification, and one coefficient significant at the 5% level (urban) when using the

Table 4 presents the results of the parametric specifications in equation (5). Columns 1 and 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 the 1% level when using the full gender interaction model as in column 3. In column 4, we relax the model assumption of full gender interactions to assess the robustness of the results. Specifically, we do not interact individual characteristics and year-by-track fixed effects with the gender indicator, but still allow the slopes to be different to the left and right of the cutoff, and to be different for each gender. The advantage of this specification 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 show that the gender difference in the retaking probability induced by the cutoff is robust to the exclusion of gender interactions of covariates. In addition, column 4 shows that men are 5–6 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 5–8 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, as shown in col. 7). The gender differences in the effects are quite substantial: the discontinuity effect for males is more than twice that for females. Figure A9 plots the estimated coefficients of gender differences (the interaction terms in cols. 3 and 7 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.

One may be concerned that our sample consists of both first-time NCEE takers and repeated NCEE takers, who have experienced the NCEE before and may have very different decision-making processes for retaking the NCEE. Therefore, we present the results for the sample with first-time NCEE takers only in table A6. As shown in table A6, using the linear control specification, males are 17 percentage points more likely to retake when falling just below the tier-2 cutoff, while females are only 9 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

quadratic control specification for males. There is one coefficient significant at the 10% level (Han) and one coefficient significant at the 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.

TABLE 4

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

	Dependent Variable: Retake							
	Male (1)	Female (2)	Full Sample (3)	Full Sample (4)	Male (5)	Female (6)	Full Sample (7)	Full Sample (8)
Male × below cutoff			.0558***	.0548***			.0732***	.0720***
			(.0133)	(.0133)			(.0203)	(.0203)
Below cutoff	.1107***	.0549***	.0549***	.0553***	.1109***	.0377***	.0377***	.0384***
	(.0108)	(.0093)	(.0093)	(.0093)	(.0167)	(.0137)	(.0137)	(.0138)
Male				.0584***				.0465***
				(.0072)				(.0097)
Observations	21,162	20,315	41,477	41,477	21,162	20,315	41,477	41,477
Bandwidth	15	15	15	15	15	15	15	15
Mean of dependent variable in [0, 15]	.10	.05	.08	.08	.10	.05	.08	.08
Interaction controls	Linear	Linear	Linear	Linear	Quadratic	Quadratic	Quadratic	Quadratic
Individual characteristics	Yes	Yes	With gender interaction	Without gender interaction	Yes	Yes	With gender interaction	Without gender interaction
Year-track fixed effects	Yes	Yes	By gender	Yes	Yes	Yes	By gender	Yes

Note.—Standard errors in parentheses. 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 5 are using only the male sample. Columns 2 and 6 are using only the female sample. Columns 1 and 2 control for a linear function of the running variable and its interaction with the indicator of below the cutoff. Columns 5 and 6 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 cols. 1–2 and 5–6. Columns 3–4 and 7–8 are using the full sample. Columns 3 and 4 control for a linear function of the running variable and its interaction with the indicator of below the cutoff, and their gender interactions. Columns 7 and 8 control 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 7 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. Columns 4 and 8 control for individual characteristics (ethnicity, hukou status, whether the individual is a first-time taker, age) and year-by-track fixed effects. Standard errors are two-way clustered at the individual identifier level and the high school year level.

*** $p < .01$.

8.2 percentage points, and is statistically significant at the 1% level when using the full gender interaction model as in column 3. The results are similar when excluding gender interactions of covariates, or using the quadratic control specification. Overall, the pattern of the results is very similar to table 4, that the gender difference in retake probability becomes around twice as large when moving from scoring just above the cutoff to scoring just below the cutoff. The point estimates of the effects are larger in table A6, as repeated NCEE takers who in general have much lower retake probability are excluded from the analysis. In conclusion, the results in table A6 are reassuring to our main findings.

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; Landaud and Maurin 2020; Wasserman 2023), and our results are for a context with much higher stakes and for a much larger and less selected population. Specifically, Wasserman (2023) focuses on politicians, a group of individuals who survive extremely competitive careers and have unusual ambitions, and most of the general population would be excluded from the analysis. In addition, Landaud and Maurin (2020) focus on candidates of elite science graduate programs in France, a setting most similar to ours but still with a very selective sample. For example, over 82% of the observations in Landaud and Maurin (2020) are males, indicating that many females who are not devoted to entering science graduate programs, and also many people who do not pursue a graduate degree, are excluded from the sample in the first place. By contrast, the gender composition is much more balanced in our setting (51%–52% males, as shown in table 1), and almost everyone needs to take the NCEE to get into college in China. Therefore, our setting can alleviate the concern of sample selection that individuals who do not like competition may choose not to participate in the competition in the first place.

By contrast, our results differ from Goodman, Gurantz, and Smith (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 for the SAT than for the NCEE, as the SAT can be taken many times in a year while the NCEE is an annual exam.

C. Heterogeneous Effects

In this section, we explore the heterogeneity in gender differences in cutoff effects to further improve our understanding on the gender differences in reactions to failure and illustrate the potential generalizability of our results. 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 5 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. In addition, the gender differences are similar for students in the science track, where females are less represented, and students in the art track.[31] Finally, the gender differences are presented for both first-time takers and repeated takers. These results 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.

We also present the estimated gender differences in the cutoff effects on retake probability by age in figure A10.[32] One possible explanation for the gender gap in retaking is a higher marriage market cost of delaying 1 year of college entrance for females. If so, we should expect the gender gap in retaking to be larger among older students. However, the results show that the gender differences are smaller for older cohorts, especially for those above 21 years old. This result does not provide support for the marriage market cost being a main driver for the gender gap we see in retaking. 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 repeated taker differences.

Table 6 presents the estimation results of the gender differences in the cutoff effects on retake probability by high school and county characteristics. In columns 1 and 2, we divide the sample based on the quality of the high school.[33] In columns 3–4 and 5–6, we divide the sample based on the sex ratio of the high school cohort (cols. 3–4) and the sex ratio of the county (cols. 5–6).[34] Finally, in columns 7 and 8, we divide the sample

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

[32] 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.

[33] 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 a 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 is a small proportion of NCEE takers (less than 1%) who do not have valid information on high school, and they are excluded from this analysis.

[34] 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 a school with a high sex ratio if the sex ratio of her high school is above or equal to the median of the high school sex ratio 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 a high sex ratio if the sex ratio of her county is above or equal to the median of the county sex ratio in the RD sample in the given year.

TABLE 5

Gender Differences in the Effects of Below Tier-2 University Cutoff on Retake Probability, by Individual Characteristics

	Dependent Variable: Retake							
	Han (1)	Minority (2)	Urban (3)	Rural (4)	Art (5)	Science (6)	First-Time Taker (7)	Repeated Taker (8)
Male × below cutoff	.0506***	.0739***	.0504**	.0627***	.0498*	.0587***	.0822***	.0276**
	(.0150)	(.0247)	(.0205)	(.0188)	(.0260)	(.0158)	(.0216)	(.0136)
Below cutoff	.0683***	.0017	.0646***	.0464***	.0567***	.0531***	.0902***	.0077
	(.0109)	(.0151)	(.0121)	(.0135)	(.0140)	(.0117)	(.0137)	(.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
MHT-adjusted p-value	.001	.019	.047	.011	.053	.001	.001	.087
Interaction controls	Linear	Linear	Linear	Linear	Linear	Linear	Linear	Linear
Individual characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-track-gender fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note.—Standard errors in parentheses. 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 the art track. Column 6 uses the sample of observations in the 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. The MHT-adjusted p-value shows the p-value for the interaction term of the male indicator and the indicator of below the cutoff after implementing the multiple hypothesis testing adjustment procedure (List, Shaikh, and Xu 2019; Barsbai et al. 2020).

* $p < .1$.

** $p < .05$.

*** $p < .01$.

TABLE 6

Gender Differences in the Effects of Below Tier-2 University Cutoff on Retake Probability, by High School and County Characteristics

	Dependent Variable: Retake							
	High Schools				Counties			
	High Quality (1)	Low Quality (2)	High Sex Ratio (3)	Low Sex Ratio (4)	High Sex Ratio (5)	Low Sex Ratio (6)	High GDP (7)	Low GDP (8)
Male × below cutoff	.0642***	.0461***	.0553***	.0556***	.0466**	.0638***	.0478***	.0604***
	(.0205)	(.0177)	(.0195)	(.0179)	(.0181)	(.0195)	(.0172)	(.0210)
Below cutoff	.0527***	.0598***	.0579***	.0528***	.0657***	.0427***	.0586***	.0550***
	(.0128)	(.0134)	(.0144)	(.0116)	(.0129)	(.0131)	(.0112)	(.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
MHT-adjusted p-value	.021	.040	.030	.011	.041	.001	.041	.026
Interaction controls	Linear	Linear	Linear	Linear	Linear	Linear	Linear	Linear
Individual characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-track-gender fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note.—Standard errors in parentheses. 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. The MHT-adjusted p-value shows the p-value for the interaction term of the male indicator and the indicator of below the cutoff after implementing the multiple hypothesis testing adjustment procedure (List, Shaikh, and Xu 2019; Barsbai et al. 2020).

** $p < .05$.

*** $p < .01$.

based on the GDP per capita of the counties.[35] The results show that the gender differences are large and of similar magnitude for students in high-quality and low-quality high schools, in places with different levels of sex ratios, as well as in rich and poor counties.

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 suggest our findings may have strong generalizability to other groups of people and more developed areas as well.

One may be concerned about the multiple hypothesis testing problem as we perform the test for multiple subgroups, which is one of the three major areas of multiple hypothesis testing (List, Shaikh, and Xu 2019). Therefore, we implement the multiple hypothesis testing adjustment procedure proposed by Barsbai et al. (2020), which is based on List, Shaikh, and Xu (2019) and modified to be used in a multivariate regression setting.[36] We present the adjusted p-values for our main coefficients of interest in table 5 and table 6. The results show that the statistical significance of the estimated gender differences is mostly unaffected by the multiple hypothesis testing adjustment, and most coefficients are still statistically significant at the 5% level, except for the coefficients for students in the art track and repeated takers, which are statistically significant at the 10% level. Therefore, the multiple hypothesis testing problem may not be a severe concern in our results.

D. *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. In this section, we briefly discuss several potential explanations for the gender differences in reactions to failure, and provide survey-based evidence on the relative importance of these potential explanations. Table 7 provides a brief summary of the potential explanations.

[35] 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.

[36] The Stata package *mhtreg* is used for multiple hypothesis testing adjustment. See List, Shaikh, and Xu (2019) and Barsbai et al. (2020) for details. When conducting multiple hypothesis testing adjustment, the cluster is specified at the high school year level and the number of bootstraps is 2,000. All of the 16 subgroups shown in table 5 and table 6 are included jointly in the multiple hypothesis testing adjustment procedure.

TABLE 7
Summary of Potential Explanations for Gender Differences in Retake Behavior

Categories	Potential Explanations for Gender Differences in Retake Behavior
Gender differences in returns to retake	Gender differences in returns in exam performance (differential improvements of exam performance)
	Gender differences in returns in labor market outcomes (differential labor market returns to access to better universities)
Gender differences in retake costs	Gender differences in opportunity costs of postponing the time of entering higher education and labor market by a year, for example differential marriage market impacts
	Gender differences in psychological costs (differential levels of disutility from the extra year of preparation)
Gender differences in noncognitive traits, preferences, and expectations	Gender differences in causal attribution
	Gender differences in risk preferences
	Gender differences in confidence
	Gender differences in competitiveness
	Gender differences in locus of control
	Gender differences in exam expectations
Gender differences in family support	Gender differences in financial or emotional support from parents
	Gender differences in educational expectations from parents

1. Gender Differences in Returns to Retake

One potential explanation for our results is that the returns to retake may be higher for males than for females. Specifically, higher returns to retake could take different forms, such as higher improvement of exam performance in the next try, or higher labor market return to access to better universities. 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. On the other hand, if the labor market return to access to better universities is higher for males, then it is also rational for them to participate in the retakes more frequently as retaking the exam substantially improves the opportunity of admission into better universities.

We can directly test the hypothesis of differential returns in exam performance improvement by examining whether the returns to retake in terms of exam outcomes are higher for males. Table 8 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 and 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

TABLE 8
GENDER DIFFERENCES IN THE EFFECTS OF RETAKE ON EXAM OUTCOMES

	Standardized Score			Ranking		
	Male (1)	Female (2)	Full Sample (3)	Male (4)	Female (5)	Full Sample (6)
Male × retake			−.1566*			−.0246
			(.0904)			(.0223)
Retake	.4220***	.5785***	.5785***	.1049***	.1295***	.1295***
	(.0460)	(.0765)	(.0765)	(.0119)	(.0179)	(.0179)
First-stage KP F-statistic	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 fixed effects	Yes	Yes	By gender	Yes	Yes	By gender

NOTE.—Standard errors in parentheses. 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 cols. 1–3; the ranking within the year-track in cols. 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 cols. 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 cols. 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.

* $p < .1$.

*** $p < .01$.

a larger return than males, although the difference is insignificant. Figure A11 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. Note that 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 discuss the plausibility of this implicit assumption and provide direct evidence in support of this assumption in appendix A (apps. A and B are available online), where we refer to the work of Imbens and Angrist (1994), among others. To conclude, higher returns for males in terms of exam outcomes are not why we see the gender gap in retake.

However, the returns we examined here only refer to improvements in exam scores, not other pecuniary or nonpecuniary 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 nonpecuniary returns for males because of labor market conditions. Our research design does not allow us to rule out this potential explanation, and gender differences in labor market returns may still be an important potential explanation for our results. Previous studies find suggestive evidence that the labor market return to elite college attendance may in fact be higher for females than for males in China (Li et al. 2012). However, it is the gender difference in perceived return, rather than actual return, that matters for the gender differences in retake behavior, so if the perceived labor market return to elite college attendance is higher for males, it can still explain our results.

2. Gender Differences in Retake Costs

Another possible explanation is that the potential costs associated with retaking the NCEE may be different across gender. The potential costs may include opportunity costs of retaking the NCEE, and also psychological costs associated with the retake decision. For example, 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. In addition, females may perceive retaking the exam to be a greater burden (i.e., their disutility from the extra year of preparation is greater), and thus have higher psychological costs than males. If these potential costs of retaking the exam are lower for males, it is rational for them to participate in the retakes more frequently as well.

3. Gender Differences in Noncognitive Traits, Preferences, and Expectations

Our results can also be explained by gender differences in noncognitive traits, preferences, and expectations. 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 their own ability and be less confident about the prospect of the retakes, and thus are less motivated to retake than males. In addition, the gender differences could come from differences in risk preferences (Boring and Brown 2016; Saygin 2016; Reuben, Wiswall, and Zafar 2017), in that females may be less motivated to retake than males because of stronger risk aversion. Gender differences in other psychological factors, such as confidence, competitiveness, and locus of control, may also explain our results. Moreover, males might simply have higher expectations for their performance in these exams and are more likely to persist for this reason.

4. Gender Differences in Family Support

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. In addition, parents may have differential expectations in terms of educational attainment for males and females, and such parental influence may also help explain our results.

5. Survey Evidence

To assess whether these potential explanations can indeed explain the gender differences in retake behavior, and to evaluate the relative importance of these potential explanations, we collect survey data on the willingness to retake and the potential mechanism variables for high school students in all grades in two high schools in Gansu province in April 2023.[37]

[37] We conduct the survey for high school students in all grades (grades 10, 11, and 12) and do not restrict the sample for analysis for students in grade 12 who are going to take the NCEE soon (in 2 months) for two reasons. First, because of recent reforms of textbooks and subjects in the NCEE, the form and content may change in 2024 in Gansu province. Therefore, for students currently in grade 12, the incentives for retaking the NCEE may be very different because they may face the new NCEE if they choose to retake in 2024. Indeed, the average willingness to retake is similar among students in grades 10 and 11 in

Unlike our analysis based on regression discontinuity design, we are unable to causally estimate the gender differences in reactions to exogenous-type of failure in survey data, so we use two questions to measure the willingness to retake under different scenarios of failure, including scoring just below the tier-1 cutoff and scoring just below the target type of university, on a scale of 1 (least likely) to 4 (most likely).[38] We use the average of these two measures as an index measure for the willingness to retake in reactions to failure.

We also use survey questions to measure some potential mechanisms. For perceived returns to retake (especially in the form of labor market return to elite college attendance), we ask students to gauge the pecuniary return to tier-1 universities versus tier-2 universities, and the pecuniary return to tier-2 universities versus tertiary technical colleges, both for the general population on average and for the person herself. For costs, we ask for the perceived marriage market cost and psychological cost of retaking the NCEE. For family support, we ask for the extent of parental support for retaking the NCEE, and also the parental expectation for the level and quality of education. For risk preference, we directly ask for the extent of risk loving. For confidence, we ask for the expected improvement in test scores in the next try in the NCEE and in low-stakes regular tests separately. For competitiveness, we ask for the willingness to compete in important competitions. We measure causal attribution by asking the extent to which the respondent attributes the failure in recent exams and NCEE to bad luck,

our sample, but much lower among students in grade 12 (2.59, 2.49, and 1.94, respectively, on a scale of 1 [least likely] to 4 [most likely]). Therefore, although grade-12 students are the closest to the NCEE, focusing on this specific cohort may be misleading because of the coming NCEE reforms. Note that our main results remain similar if we restrict the sample to students in grades 10 and 11 only. Second, since the survey data are only from two high schools, surveying students in all grades could increase our sample size.

The two high schools surveyed in Gansu are relatively good ones compared to the general population of the province. The representativeness of the sample is one limitation of our survey evidence.

[38] The survey questions are “How likely would you retake the NCEE if your score were just below the tier-1 cutoff?” and “How likely would you retake the NCEE if your score were just below your target type of university?” Before the second question, we ask for the target type of university (choosing from 985 universities, 211 universities, other tier-1 universities, tier-2 universities, and tertiary technical colleges, with the prestige from the highest to the lowest) of the respondent.

We explicitly ask for having scores just below the tier-1 cutoff rather than tier-2 cutoff for several reasons. First, due to the development in higher education in China, what is considered as failure has likely changed compared to our sample period more than 10 years ago. For instance, the tier-3 has been gradually combined into the tier-2 between 2015 and 2019 in different provinces. At the time of our survey, there are no more tier-3 universities. Second, the tier-2 cutoff plays a negligible role in our survey sample, partly due to the reasons mentioned above. Specifically, when asking for their target type of university, more than 80% of respondents report that their primary targets are tier-1 or better universities. Third, we still allow for explicitly asking about tier-2 universities in the second question about when having scores just below the target type of university, which include the cutoff for tier-2 universities for the students who report that tier-2 universities are their primary targets.

with two dummy variables representing causal attribution to external factors. We collect respondents' locus of control by eliciting the extent to which the respondent believes that the future is determined by personal effort and education. The survey also collects demographic and family background of respondents, including parental education, family income, ethnicity, and number of siblings, and includes a set of dummy variables for these covariates in our analysis. Detailed survey questions and variable definitions are reported in appendix B.

Table A7 shows the summary statistics of our survey data. Indeed, males on average have a stronger willingness to retake in reactions to failure in the NCEE than females, regardless of whether the index or the two survey questions are used as the measure for retake willingness, and the differences are all statistically significant. As for family and individual characteristics, there is no evidence of any gender differences in parental education or ethnicity, while there is evidence that females are in general less likely to be in richer households and have more siblings.

Most of the potential mechanism variables exhibit gender differences consistent with prior literature.[39] For instance, males are more risk loving, more confident, more competitive, and more likely to attribute failure to external factors such as luck, in comparison to females. These gender differences are all statistically significant. As shown in table A8, the confidence variables are positively associated with retake willingness, though the association is smaller and not statistically significant for the confidence for monthly testing.[40] Given that females tend to have lower confidence than males, these gender disparities in confidence may play a vital role in explaining our findings. In addition, competitiveness and causal attribution to external factors are both positively associated with retake willingness, though not statistically significant, and gender differences in these psychological factors may also partly explain our results. By contrast, risk loving is surprisingly not positively associated with retake willingness in our

[39] The results are mostly similar if we control for school-by-grade-by-track fixed effects rather than examining the raw gender differences.

[40] In table A8, we examine how much the coefficient of the male dummy in the OLS regression of retake willingness index on it (and other covariates) changes when we include the mechanism variables in the regression. In col. 1, we do not control for any covariates. In col. 2, we control for school-by-grade-by-track fixed effects to adjust for invariant differences in different schools, grades, and tracks flexibly, for instance, differential school quality, differential exposure to NCEE reforms across grades, and differential retake tendency across tracks. In col. 3, we additionally control for family and individual characteristics, including dummy variables for father's and mother's education, family income, ethnicity, and number of siblings. In col. 4, we use the same specification as in col. 3, but in the sample with complete information on all mechanism variables. Finally, in col. 5, we add all the mechanism variables into the regression, and the gender differences in retake willingness remain large and statistically significant. Table A9 presents the results using the two survey questions rather than the index as the dependent variables for analysis, and the results are similar.

sample. Therefore, although males are more risk loving than females, such gender differences in risk preference are unable to explain the gender differences in willingness to retake.

Regarding the returns to retaking the NCEE, we do find that males perceive a higher return to tier-1 universities for the general population, but not for their own return, and there is no evidence on gender differences in return perceptions for tier-2 universities. In addition, table A8 suggests that the return perceptions do not show strong positive association with retake willingness, which indicates that differential returns may only play a minor role in explaining our results. As for the costs associated with retaking the NCEE, we find that females report higher psychological costs for preparing for retake for a year than males. This may explain the gender differences in retake willingness that we observe, as psychological cost is strongly negatively associated with retake willingness. Interestingly, females report lower, rather than higher, perceived marriage market costs than males. This, along with the gender retaking gap being higher for younger students (as shown in sec. IV.C), suggests that concerns about marriage market opportunity costs may not significantly contribute to explaining our results.

As for family support, we find that parents have lower educational expectations for females compared to males. Such differential expectations may transform into differential retake willingness throughout the education journey. On the other hand, female students do not report statistically different levels of expected family support for retaking the NCEE compared to male students. Both family support and expectation variables are strongly positively associated with retake willingness, and gender differences in parental education expectations may offer another crucial explanation for gender differences in retake willingness. Finally, the belief of the importance of personal effort is similar for females and males, while there is evidence that females actually have stronger belief of the importance of education, which is positively correlated with retake willingness, indicating that gender differences in locus of control are unlikely to explain our results.

To further assess the relative importance of these potential mechanisms, we use the Oaxaca-Blinder decomposition to decompose the gender differences in willingness to retake, and the results are presented in table 9. Out of the total gender gap in willingness to retake (−0.1240), 44.3% of the gender gap can be "explained" by gender differences in observed characteristics, with the remaining 55.7% of the gender gap "unexplained," that is, driven by differences in the importance of these factors. Within the explained gap, cost variables (15.1%), confidence variables (14.1%), and family support and expectation variables (7.7%) are the mechanisms with the most substantial explanatory power, followed by competitiveness variables (3.3%) and causal attribution variables (3.0%). Family and individual characteristics, risk preference, returns to retake, and locus of control do not

TABLE 9
DECOMPOSITION OF GENDER RETAKE WILLINGNESS GAP

	Retake Willingness Index	
	Gap Explained	Percent of Gap Explained
Family and individual characteristics variables	.0035	−2.85
Risk preference variables	.0044	−3.52
Labor market return perception variables	.0020	−1.61
Retake cost variables	−.0187	15.09
Family support and expectation variables	−.0096	7.71
Confidence variables	−.0175	14.09
Willingness to compete variables	−.0041	3.30
Causal attribution variables	−.0037	3.01
Locus of control variables	.0086	−6.91
School-grade-track indicator variables	−.0198	15.99
Total explained gap	−.0549	44.31
Total unexplained gap	−.0690	55.69
Total gap	−.1240	100.00

NOTE.—This table shows the results of Oaxaca-Blinder decomposition of the gender differences in retake willingness index. Family and individual characteristics variables include dummy variables for parental education, family income, ethnicity, and number of siblings. Risk preference variables include the extent of risk loving. Labor market return perception variables include the perceived pecuniary return to tier-1 universities vs. tier-2 universities, and the perceived pecuniary return to tier-2 universities vs. tertiary technical colleges, for both the general population and the person herself. Retake cost variables include the perceived marriage market cost and psychological cost of retaking the NCEE. Family support and expectation variables include the extent of parental support for retaking the NCEE, and also the parental expectation for the level and quality of education. Confidence variables include the expected improvement in the NCEE test score and in low-stakes regular test score. Willingness to compete variables include the willingness to compete in important competitions. Causal attribution variables include the extent that the respondent attributes the failure in recent exams and NCEE to bad luck. Locus of control variables include the extent that the respondent believes that the future is determined by personal effort and education.

have any explanatory power in the expected direction. There is also 16.0% of the gender gap accounted for by gender differences in the distribution across schools, grades, and tracks.

To conclude, our analysis based on self-collected survey data reveals that gender differences in psychological costs of retaking the NCEE and parental education expectations are important mechanisms behind the gender differences in reactions to failure. Gender differences in noncognitive traits, such as confidence, competitiveness, and causal attribution, also explain a substantial part of our results. By contrast, gender differences in returns to retaking and other psychological traits do not appear to elucidate our main findings.

E. *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. Conditional on year-track fixed effects, females have a 0.021-unit lower standardized score in terms of the final outcome over the 2-year period than males in the sample. If the gender gap in retake probabilities vanishes, and assuming the returns to retake can be extrapolated to this sample, then females would now instead have a 0.016-unit advantage in terms of the final standardized score in this counterfactual case. Similarly, conditional on year-track fixed effects, females have a 2.5 percentage point lower probability of finally being eligible to apply for higher-quality tier-1 universities over the 2-year period than males in this sample. If the gender gap in retake probabilities vanishes, then females would now instead have a 2.3 percentage point higher probability of finally being eligible to apply for higher-quality tier-1 universities over the 2-year period than males. Therefore, females would be substantially more represented in high-quality universities. These effects may have important implications for the gender disparities in the labor market.

The policy implications for the gender differences in reactions to failure would depend on the underlying mechanisms of the findings. For example, if gender differences in labor market returns are the main drivers of the results, then policy interventions reducing gender inequality in labor market would also reduce the gender differences in retake behavior. If gender differences in psychological costs are the main drivers of the results, then policy interventions that could alleviate psychological concerns may be effective. If gender differences in confidence and causal attribution are the main drivers of the results, then providing more information and feedback on students' exam performance and academic ability may reduce such behavioral gender gaps and reduce the gender differences in retake behavior. Our analysis based on survey data suggests that the first case is of a smaller likelihood, while future research is still needed in this area for clearer policy implications.

V. 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 8 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 deviations, 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 performance improvement. Our supplementary analysis based on self-collected survey data suggests that gender differences in psychological costs of retaking the NCEE and parental education expectations are important mechanisms behind the gender differences in reactions to failure, and gender differences in some noncognitive traits, such as confidence, competitiveness, and causal attribution, also explain an important part of our results. By contrast, gender differences in returns to retake and other psychological traits do not explain our main results. Our estimates suggest that if females were equally likely to retake as males, females would have better final exam performance and be substantially more represented in high-quality universities, which may in turn have important implications for gender equality in the labor market.

Future research is needed to examine the effects of retaking NCEE on long-term outcomes, such as labor market and marital outcomes, and to further disentangle the potential explanations for the gender differences in reactions to failure in the NCEE. Answering these questions could be important for understanding the implications of the gender gap in the long run, and for effective policy designs to address the gender gap.

Data Availability

Code replicating the tables and figures in this article can be found in Kang et al. (2023) in the Harvard Dataverse, https://doi.org/10.7910/DVN/TPVTAW.

References

Astorne-Figari, Carmen, and Jamin D. Speer. 2019. "Are Changes of Major Major Changes? The Roles of Grades, Gender, and Preferences in College Major Switching." *Econ. Educ. Rev.* 70:75–93.

Barsbai, Toman, Victoria Licuanan, Andreas Steinmayr, Erwin Tiongson, and Dean Yang. 2020. "Information and the Acquisition of Social Network Connections." Working Paper no. 27346 (June), NBER, Cambridge, MA.

Berlin, Noémi, and Marie-Pierre Dargnies. 2016. "Gender Differences in Reactions to Feedback and Willingness to Compete." *J. Econ. Behavior and Org.* 130:320–36.

Beyer, Sylvia. 1998. "Gender Differences in Causal Attributions by College Students of Performance on Course Examinations." *Current Psychology* 17 (4): 346–58.

Boring, Anne, and Jennifer Brown. 2016. "Gender, Competition and Choices in Higher Education." Working paper, Sciences Po, Paris.

Buser, Thomas, Muriel Niederle, and Hessel Oosterbeek. 2014. "Gender, Competitiveness, and Career Choices." *Q.J.E.* 129 (3): 1409–47.

Buser, Thomas, Noemi Peter, and Stefan C. Wolter. 2017. "Gender, Competitiveness, and Study Choices in High School: Evidence from Switzerland." *A.E.R.* 107 (5): 125–30.

Buser, Thomas, and Huaiping Yuan. 2019. "Do Women Give Up Competing More Easily? Evidence from the Lab and the Dutch Math Olympiad." *American Econ. J. Appl. Econ.* 11 (3): 225–52.

Cai, Xiqian, Yi Lu, Jessica Pan, and Songfa Zhong. 2019. "Gender Gap under Pressure: Evidence from China's National College Entrance Examination." *Rev. Econ. and Statis.* 101 (2): 249–63.

Calonico, Sebastian, Matias D. Cattaneo, Max H. Farrell, and Rocío Titiunik. 2017. "rdrobust: Software for Regression-Discontinuity Designs." *Stata J.* 17 (2): 372–404.

Calonico, Sebastian, Matias D. Cattaneo, and Rocío Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica* 82 (6): 2295–326.

———. 2015. "Optimal Data-Driven Regression Discontinuity Plots." *J. American Statis. Assoc.* 110 (512): 1753–69.

Cattaneo, Matias D., Nicolás Idrobo, and Rocío Titiunik. 2018. *A Practical Introduction to Regression Discontinuity Designs*, vol. 2. Cambridge: Cambridge Univ. Press.

Cattaneo, Matias D., Michael Jansson, and Xinwei Ma. 2018. "Manipulation Testing Based on Density Discontinuity." *Stata J.* 18 (1): 234–61.

———. 2020. "Simple Local Polynomial Density Estimators." *J. American Statis. Assoc.* 115 (531): 1449–55.

Delaney, Judith M., and Paul J. Devereux. 2021. "Gender and Educational Achievement: Stylized Facts and Causal Evidence." IZA Discussion Paper no. 14074, Inst. Labor Econ., Bonn.

Dweck, Carol S., William Davidson, Sharon Nelson, and Bradley Enna. 1978. "Sex Differences in Learned Helplessness: II. The Contingencies of Evaluative Feedback in the Classroom and III. An Experimental Analysis." *Developmental Psychology* 14 (3): 268–76.

Ellison, Glenn, and Ashley Swanson. 2018. "Dynamics of the Gender Gap in High Math Achievement." Working Paper no. 24910 (August), NBER, Cambridge, MA.

Fang, Chao, Ernest Zhang, and Junfu Zhang. 2021. "Do Women Give Up Competing More Easily? Evidence from Speedcubers." *Econ. Letters* 205:109943.

Flory, Jeffrey A., Andreas Leibbrandt, and John A. List. 2015. "Do Competitive Workplaces Deter Female Workers? A Large-Scale Natural Field Experiment on Job Entry Decisions." *Rev. Econ. Studies* 82 (1): 122–55.

Goodman, Joshua, Oded Gurantz, and Jonathan Smith. 2020. "Take Two! SAT Retaking and College Enrollment Gaps." *American Econ. J. Econ. Policy* 12 (2): 115–58.

Ha, Wei, Le Kang, and Yang Song. 2020. "College Matching Mechanisms and Matching Stability: Evidence from a Natural Experiment in China." *J. Econ. Behavior and Org.* 175:206–26.

Holz, Justin E., John A. List, Alejandro Zentner, Marvin Cardoza, and Joaquin Zentner. 2020. "The $100 Million Nudge: Increasing Tax Compliance of Businesses and the Self-Employed Using a Natural Field Experiment." Working Paper no. 27666 (August), NBER, Cambridge, MA.

Imbens, Guido W., and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62 (2): 467–75.

Jia, Ruixue, and Hongbin Li. 2021. "Just Above the Exam Cutoff Score: Elite College Admission and Wages in China." *J. Public Econ.* 196:104371.

Kang, Le, Ziteng Lei, Yang Song, and Peng Zhang. 2023. "Replication Data for: 'Gender Differences in Reactions to Failure in High-Stakes Competition: Evidence from the National College Entrance Exam Retakes.'" Harvard Dataverse. https://doi.org/10.7910/DVN/TPVTAW.

Kleibergen, Frank, and Richard Paap. 2006. "Generalized Reduced Rank Tests Using the Singular Value Decomposition." *J. Econometrics* 133 (1): 97–126.

Kolesár, Michal, and Christoph Rothe. 2018. "Inference in Regression Discontinuity Designs with a Discrete Running Variable." *A.E.R.* 108 (8): 2277–304.

Krishna, Kala, Sergey Lychagin, and Veronica Frisancho. 2018. "Retaking in High Stakes Exams: Is Less More?" *Internat. Econ. Rev.* 59 (2): 449–77.

Landaud, Fanny, and Éric Maurin. 2020. "Aim High and Persevere! Competitive Pressure and Access Gaps in Top Science Graduate Programs." Working paper, Dept. Econ., Paris School Econ.

Lee, David S., and David Card. 2008. "Regression Discontinuity Inference with Specification Error." *J. Econometrics* 142 (2): 655–74.

Li, Hongbin, Lingsheng Meng, Xinzheng Shi, and Binzhen Wu. 2012. "Does Attending Elite Colleges Pay in China?" *J. Comparative Econ.* 40 (1): 78–88.

List, John A. 2020. "Non Est Disputandum de Generalizability? A Glimpse into the External Validity Trial." Working Paper no. 27535, NBER, Cambridge, MA.

List, John A., Azeem M. Shaikh, and Yang Xu. 2019. "Multiple Hypothesis Testing in Experimental Economics." *Experimental Econ.* 22 (4): 773–93.

Niederle, Muriel, and Lise Vesterlund. 2007. "Do Women Shy Away from Competition? Do Men Compete Too Much?" *Q.J.E.* 122 (3): 1067–101.

Reuben, Ernesto, Matthew Wiswall, and Basit Zafar. 2017. "Preferences and Biases in Educational Choices and Labour Market Expectations: Shrinking the Black Box of Gender." *Econ. J.* 127 (604): 2153–86.

Ryckman, David B., and Percy D. Peckham. 1987. "Gender Differences in Attributions for Success and Failure." *J. Early Adolescence* 7 (1): 47–63.

Saygin, Perihan Ozge. 2016. "Gender Differences in Preferences for Taking Risk in College Applications." *Econ. Educ. Rev.* 52:120–33.

Wasserman, Melanie. 2023. "Gender Differences in Politician Persistence." *Rev. Econ. and Statis.* 105 (2): 275–91.

Zhang, Y. Jane. 2019. "Culture, Institutions and the Gender Gap in Competitive Inclination: Evidence from the Communist Experiment in China." *Econ. J.* 129 (617): 509–52.

Zhang, Yan, Yanqing Ding, Honghui Li, Wei Li, Yujie Ma, and Xiaoyang Ye. 2019. "Retaking the College Entrance Exam: Evidence from a Regression Discontinuity Design." *Global Educ.* 48 (7): 111–28 (in Chinese).