Do Second Chances Pay Off?

Evidence from a Natural Experiment with Low-Achieving Students*

Aspasia Bizopoulou† Rigissa Megalokonomou‡ Ştefania Simion§

February 2023

Abstract

In several countries, students who fail end-of-high-school high-stakes exams are faced with the choice of retaking or forgoing postsecondary education. We explore exogenous variation generated by a policy that imposed a performance threshold for admission into postsecondary education in Greece to estimate the effect of retaking exams on academic performance and various measures of the quality of received offers. Using a fuzzy regression discontinuity design and novel administrative data, we find that low-achieving students who retake national exams improve their performance by around 0.6 of a standard deviation, and obtain higher quality postsecondary offers.

JEL Codes: I21, I23
Keywords: postsecondary education admission, low-achieving students, exogenous policy, fuzzy regression discontinuity design

*We thank Andres Barrios Fernandez, Emma Duchini, Joshua Goodman, Kristian Koerselman, Ofer Malamud, Tuomas Pekkarinen, Helene Turon, and four anonymous referees for their useful comments. We also acknowledge participants at the EffEE Conference 2022, the Nordic Summer Institute in Labor Economics 2022, the EALE Conference 2021, the ERMAS conference 2021, the IWAEE 2021, the ESPE Conference 2021, the Scottish Economics Society Conference 2021, the VATT Seminar 2021, the 5th IZA Workshop on the Economics of Education 2020, EEA Conference 2020, and the Aalto Lunch seminar 2020 for their constructive suggestions. We are grateful to Marios Tsoukis for excellent research assistance. Finally, we thank Georgia Lioli and the personnel of the Greek Ministry of Education & Religious Affairs for helping us access the administrative data; in particular Sylvia Dremetsika, Georgios Mallios, and Eirini Pittara. This paper uses administrative data containing the universe of applications to postsecondary education institutions in Greece and we are not at liberty to publish it online, however, we can provide the code as well as detailed instructions and assistance on how to apply and access the data in order to enhance the replicability of our analysis. As far as we are aware, there is no material or financial conflict of interest on the part of any of the authors with respect to any of the research in this paper.

†VATT Institute for Economic Research, Arkadiankatu 7, 00101 Helsinki, Finland; *Email:* aspasia.bizopoulou@vatt.fi; *Phone:* +358 503465034

‡Monash University, Business School, Department of Economics and The School of Economics, Level 6, Colin Clark Building (39), The University of Queensland, St Lucia, QLD, 4072, Australia; *Email:* r.megalokonomou@uq.edu.au

§Corresponding author. School of Economics, University of Bristol, 12A Priory Rd, Bristol, BS8 1TU, United Kingdom; *Email:* stefania.simion@bristol.ac.uk

1 Introduction

A growing literature shows that scoring above or below a threshold in high-stakes exams can be a key determinant for lifetime success or failure (Pop-Eleches and Urquiola, 2013; Goodman, Oded and Smith, 2020; Machin, McNally and Ruiz-Valenzuela, 2020). Several countries use such exams to select students for a limited number of admissions slots, including China, Chile, Greece, Israel, India, Korea, and Turkey. Success on these exams is a gateway to postsecondary education and better labor market outcomes (Blundell, Dearden, Goodman and Reed, 2000; Belfield, Britton, Buscha, Dearden, Dickson, Van der Erve, Sibieta, Vignoles, Walker and Zhu, 2019). Failure on these high-stakes exams can be *absolute* — when students fail to obtain a place in higher education due to poor performance on the exam — or it can be *relative*, when students fail to get into their postsecondary university or degree of choice.

Failing on the first attempt can be costly for students, since they may have to wait up to an entire year to retake the exam. If students are not in the labor market while preparing to retake the entrance exam, they forgo earnings. Such students devote a significant amount of effort, money, and time to retaking high-stakes exams, while also incurring mental health costs due to the stigma of having to retake (Krishna, Lychagin and Frisancho 2018). Some benefits are associated with retaking, such as increased familiarity with the exam format, learning gains, and improved scores (Vigdor and Clotfelter, 2003; Frisancho, Krishna, Lychagin and Yavas, 2016; Goodman, Oded and Smith, 2020).

In this paper, we analyse the impact of retaking high-stakes exams (due to absolute failure) for entry into postsecondary education on students' performance in subsequent attempts and on the quality of their postsecondary placement. Our unique contribution is that we address the potential endogeneity related to who retakes using a policy that introduced a minimum exogenous performance threshold in postsecondary admission. A common concern in the literature is that the decision of a student to retake an exam is likely affected by confounding characteristics, including the lack of motivation, financial hardship for the student's family, a supportive environment at home or school, or the availability of outside options such as studying abroad (Krishna, Lychagin and Frisancho, 2018; Machin, McNally and Ruiz-Valenzuela, 2020; Goodman, Oded and Smith, 2020). Existing studies use either a naturally-induced threshold, such as multiples of 100 in the SAT scale in the US or grade C (based on the original mark prior to any remarking) in the GSCE exam in Britain (Goodman, Oded and Smith, 2020; Machin, McNally and Ruiz-Valenzuela, 2020) or structural estimation (Krishna, Lychagin and Frisancho, 2018) in an attempt to deal with endogeneity. However, it may be the case that students, teachers or graders may respond to existing and anticipated naturally-induced thresholds to some extend or that not all students view these round number thresholds as real milestones. In this paper, we overcome existing challenges by using a policy that imposed an exogeonous absolute performance threshold for postsecondary admission for all applicants in Greece.

Starting in 2006, senior year high school students must achieve a minimum average (across subjects) score of 10,000/20,000 (i.e., 50%) on the postsecondary education admission exams. Previously there was no minimum admission threshold. Students whose performance is below this threshold are now refused access to postsecondary education. The only way for students who failed at their first attempt to access postsecondary education is to retake all exams in the next exam session (1 year later) and score above the required performance threshold. Under the new policy, failing the exam is quasi-randomly assigned among low-achieving students, and thus we control for confounding characteristics of students that might be correlated with the decision to retake the exam. The prospect of *absolute* failure is faced by low-achieving students who obtain scores close to the threshold and who are the focus of this paper. This group is of paramount importance for policy purposes, because they are more likely to give up education. To identify causal effects and account for partial compliance[1], we use a fuzzy regression discontinuity design. Comparing students who just failed the threshold to those students who just passed the threshold eliminates selection bias due to observed and unobserved teacher, class, and school characteristics. We mainly focus on the first cohort that was affected by the policy change (the 2006 cohort), since only for this cohort the policy was unanticipated and we can make sure that students, teachers and graders are unlikely to respond to the unexpected introduction of the minimum performance threshold.

We use new administrative data that we obtained from the Ministry of Education in Greece. The data include detailed student-level records of exam performance and postsecondary placement for the universe of students applying to postsecondary education institutions in Greece and we have access to this information for each time a student retakes the exam. This dataset is also linked to rich information about student postsecondary education preferences and choices, such as a ranked order of students' preferences for college-courses, the type of each college-course (academic university compared to vocational school), the quality of each college-course and information about the offered postsecondary education college-course. To complement our main analysis, we also use a smaller comprehensive dataset which includes additional information on student birthdate and other measures of student-level academic performance.

[1]Due to some affirmative action policies that we explain in Section 3.2.

We find that students who retake the exam improve their performance by around 24%, which is equivalent to an improvement of 0.6 of a standard deviation. This is in line with previous results for lower-achieving students retaking SATs (Goodman, Oded and Smith, 2020). We provide evidence that the exam improvement is not explained by teachers inflating students' grades as a response to the policy. We conclude the performance improvement is most likely the effect of learning through studying during the additional year.

We also find that students obtain better postsecondary placements upon retaking, although there is a period of adjustment from the onset of the policy. In all years that the policy is in place retakers are more than one third more likely to obtain an offer from an academic as opposed to a vocational institution. They are also more likely to obtain an offer from an institution that is overall better ranked (above median in quality), although estimates for the first cohort are positive but imprecisely estimated. Overall, we find that the improved performance upon retaking the entrance exam leads to higher quality postsecondary outcomes on average.

We explore a number of mechanisms to explain the students' overall improved performance and improved postsecondary outcomes upon retaking. We first evaluate the role of grade inflation in explaining improved scores, in light of the exogenous imposition of the minimum threshold. We then study whether retakers choose to re-evaluate their application strategies in terms of the quality of their choices and their risk-taking behavior. We do not find support for either of these mechanisms: we show that high-school GPA distributions do not substantially change before and after the reform. Furthermore, we find no evidence that students become more (or less) ambitious or change their attitudes against risk in the way they list their postsecondary choices in their later attempt compared to their first attempt. In particular, students do not change the way they compile their college-course preference list.

We also identify a possible mechanism that could explain why the first cohort of students that we study experience a weaker improvement in postsecondary outcomes, despite a clear improvement in their exam performance. We show that upon the introduction of the policy of the minimum threshold in 2006, the pool of applicants to postsecondary education changes substantially, since low-performing students are now discouraged from applying. Thus, students who retake the exam in 2007 see a slightly weaker relative improvement in their placement. In a counterfactual exercise, we show that had the 2006 cohort of retakers competed against their contemporary peers or a cohort that was not affected by the threshold, they would have improved their ranking by 14-18 percentage points in the distribution and would have been able to access places at better-ranking institutions.

We provide evidence for the validity of our design. First, we show that the distribution of students' scores is continuous around the threshold and there are no jumps. Second, we provide evidence that there is no manipulation of the assignment variable around the threshold using a formal test by Calonico, Cattaneo and Titiunik (2014b). Third, we show that students who just fail to achieve the required achievement threshold on their first attempt have on average the same observed characteristics to those students who just achieved the required achievement threshold.

We also perform a battery of robustness checks. First, we use alternative bandwidths and the estimated effects remain almost unchanged. Second, we use pseudo-thresholds and we find no significant effects. Third, we eliminate observations located very close to the performance threshold (donut method), and the estimated effects remain similar to the main ones. Forth, we show that the estimated effects remain similar when we use different measures of performance.

Our study moves beyond the existing literature in several important ways. We are able to extend and contribute to a previously established result by Goodman, Oded and Smith (2020) regarding the impact of retaking exams on performance. They study the effect of retaking US SATs in the US (exploiting thresholds at multiples of 100) and find a positive effect on test scores and outcomes among low-achieving students, similar to our main results. However, our results stem from an institutional framework that is both extremely common around the world (a high stakes exam at the end of the academic year to enter post-secondary education taking place only once a year, e.g. Brazil, Chile, China, Croatia, India, Spain, Finland and more) and also very different to the SATs. The SATs can be retaken several times throughout the year, while the students are still enrolled at school and with little financial cost (save the SAT fee). It is not immediately clear that this result would survive in the context of an extremely high stakes exam taking place once a year, carrying the significant social cost of spending a year outside of formal education or employment while waiting to retake. This may lead to a loss of human capital for students. Within our context, it is particularly salient to identify whether retaking leads to better outcomes, since the cost of retaking is so large.

Second, our paper contributes to a small but growing literature on retaking high-stakes exams (Goodman, Oded and Smith, 2020; Landaud and Maurin, 2020; Krishna, Lychagin and Frisancho, 2018; Goodman, Hurwitz and Smith, 2017; Frisancho, Krishna, Lychagin and Yavas, 2016; Proud, 2015; Vigdor and Clotfelter, 2003). The data-generating process that we exploit allows us to contribute and add to previous work evaluating the role of the strength of the competition when retaking the exam (Krishna, Lychagin and Frisancho, 2018). Our study highlights the role of the strength of the competition in settings where a minimum threshold is imposed and thus the overall skill level in the applicant pool changes. Selective institutions with limited places, such as degree-granting entities, the civil service, or popular companies, often readjust their selective criteria for entry either to increase the signalling value of accepted individuals or to decrease selection costs. Thus, we highlight the fact that while retaking a high-stakes exam will most likely lead to improved performance, when a limited number of places is at stake the strength of the competition is a relevant factor that should be considered as carefully as the individual student's performance.

We also contribute to the literature that analyzes the effects of passing college admission exams on subsequent outcomes (Avery, Gurantz, Hurwitz and Smith, 2018; Goodman, Hurwitz and Smith, 2017; Zimmerman, 2014). Given our focus on low-achieving students, our study is also related to the literature that focuses on just passing or failing important exams. In line with Goodman, Hurwitz and Smith (2017), we show that failing to meet a relevant minimum threshold increases exam retaking, which shows that despite worse scores low-achieving students value access to post-secondary education. Clark and Martorell (2014) provide evidence that marginal students do not benefit from marginally passing high school exit exams in the US. Canaan and Mouganie (2018) show that students marginally passing the French high school exit exam are more likely to enroll into STEM degrees and into better quality postsecondary institutions and also earn more in their late 20s, relative to those who marginally failed the exam. Machin, McNally and Ruiz-Valenzuela (2020) present findings that students who just fail the high-stakes national examination taken at the end of compulsory schooling in England have a lower likelihood of entering postsecondary education and are more likely to drop out of education before reaching 18. Martorell, McFarlin and Xue (2015) show that having to take remediation courses does not lead to a drop in college enrollment among students in Texas. Andresen and Løkken (2020) show that low-achieving students who fail an exam are more likely to drop out from school in Norway.

Finally, we exploit an education system that is highly centralized. This feature makes our setting ideal for the type of questions that we study. In highly centralized systems, such as the Greek one, admission to postsecondary education depends entirely on well-defined measures of performance (mainly externally marked) and students' preferences. In other systems, in contrast, postsecondary placement depends on merit more loosely (i.e., personal statements, reference letters, extracurricular activities, subjective ranking of students or schools). Another feature that makes the Greek setting appealing is that it relies less on a student's socioeconomic background compared with other settings. Postsecondary education is free of tuition fees in Greece, which means that students express their unconstrained preferences in terms of their desired postsecondary college-courses. This allows them to potentially move up the socioeconomic ladder by enrolling to a prestigious college-course regardless of whether they come from low-income families.

The rest of the paper is structured as follows. Section 2 describes the institutional setting, section 3 presents the data and section 4 discusses the methodology. Section 6 reports the main results, and Section 7 discusses the mechanisms. Section 8 focuses on the robustness checks, and section 9 concludes.

2 Institutional Setting

2.1 Admission to Postsecondary Education in Greece

Most postsecondary institutions in Greece are public and state-funded. Admission to postsecondary education is based on student performance on national high school examinations and is centralized and administered by the Ministry of Education. Students must take these standardized national exams at the end of their final year of senior high school to enter postsecondary education. Each postsecondary education college-course has a specific number of available places every year, and students submit a preference list of their desired postsecondary college-courses in ranked order to the Ministry of Education. These preferences can be over both college-courses in academic universities and vocational schools.

Students are ranked based on their *admissions grade*. The admissions grade is a weighted average grade based on a student's average performance on national and school exams in core and track subjects in the final year of high school.[2] In 2006, the postsecondary education admissions grade was calculated based on a student's national exam performance in six subjects: two core subjects and four track subjects. Students may also take exams in an optional subject.[3]

The admissions score ranges from 0 to 20,000 and is increasing in performance. Core subjects are general education subjects and are compulsory for all students. Students choose a track in the penultimate year of high school. All high schools in Greece offer three tracks. These are classics, science, and exact science. Students take the same subjects within tracks, but subjects differ across tracks. Track subjects are compulsory for students in a given track. All schools that administer these national exams follow the same curriculum and offer courses in core and track subjects in accordance with the material covered on the exam.

[2]National exams take a much higher weight compared with school exams (70% compared with 30%) in the calculation of the admissions grade.

[3]This optional subject is required for entrance to specific fields of study, such as economics or foreign languages.

Exams in each subject are centrally collected and sent to examiners across the country for scoring; the student's name and school are hidden. Each exam is scored by at least two anonymous independent graders. Regrading an exam script is not an option for students in this educational system. All students who desire to apply to tertiary education must take the national exam at least once.

After exam scores have been finalized and released, students rank their preferences for postsecondary college-courses.[4] In general, all postsecondary college-courses are divided into five broad scientific fields of study.[5] Each student can apply to college-courses that belong to up to two scientific fields of study. Each scientific field includes both academic and vocational courses. Depending on the scientific fields of study students are considering, different high school subjects take different weights in calculating the admissions score. Since students can apply to up to a maximum of two postsecondary education specialisation tracks, they may obtain two different admissions scores which they use in their applications. Thus students may have two slightly different admission scores depending on the scientific field of study they apply to.[6]

Once students have expressed their college-course preferences, they submit their ranked preference list to the Ministry of Education, which assigns students to college-courses using a centralized application system for postsecondary education (Goulas and Megalokonomou, 2019). This algorithm considers college-course rankings in the order of students' admissions scores. It starts with the student with the highest admissions score and places her at her top-ranked college-course. It then assigns the student with the second-highest admissions score to their top-ranked college-course and then it moves down to the student with the third-highest admissions score. Once a college-course runs out of places, a student is placed to the college-course ranked second highest, and so on; thus each student will be offered a unique placement. Since there is no limit in terms of how many college-courses students can include in their preference list, the dominant strategy for each student is to submit rankings for college-course preferences that truthfully reveal their preferences (Chade and Smith, 2006). It should also be noted that the contemporaneous college-course-specific threshold is not known to students when they submit their preference list; however, they can easily check the local performance admission thresholds of previous

[4]A college-course includes a university and department configuration.

[5]Those scientific fields of study are determined by the Ministry of Education and include: (1) Humanities, Law and Social Sciences, Theology, Teaching Professions, (2) Hard Sciences, Military Academies, Nautical Academies, Teaching Professions, (3) Medical Schools, Military Medical Schools, (4) Technology and Engineering, Military Academies, Nautical Academies, and (5) Economics and Business, Police Academies, Tourism Academies.

[6]Nevertheless, students' raw average test core in the six subjects in national exams is always the same. Consider a student who is applying for college-courses in the following two scientific fields of study: (a) Technology and Engineering and (b) Economics and Business. Performance in mathematics and physics received a higher weight in the calculation of the admission score for college-courses in technology and engineering, while performance in economics and mathematics receives a higher weight for college-courses in economics and business.

years.

2.2 Decision to Retake the Admissions Exams

Students take the exam in their senior year of high school, hoping to first pass the admissions performance threshold and then to score high enough to meet their desired college-course choices' thresholds. Students who do not pass the admissions performance threshold on their first attempt may retake the national exams 1 year (or more) later if they desire to reattempt to gain access to a postsecondary college-course.[7] In 2005, 19% of applicants failed to secure a postsecondary place.[8] Students who failed to gain access to tertiary education in their first attempt and desire to try again must retake the exam in all subjects and not only specific subjects. Exam questions are the same for retakers and students who are taking the exam for the first time. The curriculum assessed at the national level is usually the same over time. Students who retake the exam usually do not attend any school, pursue any job, or undertake military service between graduation and the next examination period.[9] There is only one examination period per year, around May-June, and schools do not provide any additional classes for retakers. In other words, retaking is a costly decision, in terms of both opportunity costs and preparation costs faced by students who must prepare and study intensively during the intervening year.[10] Despite this, in 2005, 15% of applicants retook the exam and gained access to postsecondary education.[11]

2.3 The 2006 Performance Admissions Threshold Reform

Before 2006, there was no specific admission performance threshold that students had to achieve to gain access to postsecondary education; each postsecondary college-course had its own local performance admissions threshold, and there was no universal performance threshold for admission. That is, any postsecondary education admissions score could—in theory—guarantee a postsecondary education spot, conditional on availability. Starting from 2006, an absolute performance threshold for admissions was imposed, such that students had to obtain an admissions score of at least 10,000/20,000 (equivalent to 50%) to be eligible to apply for a postsecondary college-course.

[7] Alternatively, students might decide to study abroad or not study at all. In 2019, 5% of the tertiary student population was enrolled abroad, while 8% of 18-24 year-olds had been in neither employment nor education for the past 12 months (OECD, 2019).

[8] Authors' calculation based on the administrative dataset used in this paper.

[9] Serving in the military is compulsory for men in Greece. The average time they must serve is 12 months. It is not the norm for students to delay their undergraduate studies in order to serve the military.

[10] It should also be noted that preparatory classes in the form of private tutoring are common, with Greece being among countries with the most time spent per week in private after-school classes (OECD, 2018).

[11] Authors' calculation based on the administrative dataset used in this paper.

Amid persistently high youth unemployment, the primary argument for the introduction of the restrictive minimum threshold was to increase the signalling value of a postsecondary education college-course for graduates in the labor market.[12] As Figure A1 in Appendix A shows, the law that introduced the minimum threshold for entering higher education was submitted to Parliament on 1/9/2005, debated on 21/22 September 2005, and implemented without previous notice for the 2006 high school incoming cohort.[13]

The policy was in place for only 4 years, and the admissions threshold was abolished in 2010. During the years the policy was in effect, lower-achieving students were discouraged from applying to postsecondary education. In particular, on average 22% fewer students applied to enter postsecondary education during the years of the reform relative to the two adjacent years before and after the reform, while the grades of the admitted students were 24% higher.[14] The policy was reintroduced in 2020.

3 Data

3.1 Main Study Sample

In this study we use rich administrative data obtained from the Greek Ministry of Education and Religious Affairs for the universe of students who apply to postsecondary education institutions in Greece. The Ministry of Education and Religious Affairs collects annual data on all senior high school students. We focus on the 2006 cohort of students — i.e., students who took the exam for the first time in 2006. We obtain access to students' entire applications for postsecondary education. In particular, we gained access to a variety of variables reported on students' applications, including a unique student identifier, student gender, high school graduation year, name of high school attended, postsecondary education admissions score, the combination of colleges and courses students apply, the order of preference in which students rank their college-course preferences on their applications, whether they obtain an offer and, if so, at which college and course. Importantly, we have access to this information every time a student takes the exam.

[12]In August 2005, a leading newspaper published an article that highlighted the record low admissions scores obtained by the outgoing cohort of 2005. The article criticised the government for letting students with such low grades enter higher education. Two days later, the Minister for Education at that time was interviewed by the same newspaper and promised to look into the matter and introduce changes. It is not entirely clear whether it was the newspaper article that prompted the discussion of introducing an admissions performance threshold for low-achieving students or the government was already planning to introduce the minimum threshold.

[13]Law 3404/2005, Article 13 states that students have to achieve at least half of the possible maximum postsecondary education admissions score to gain access to tertiary education. More information can be found in Figure C1 in Appendix B.

[14]Authors' calculations based on the administrative data used in this paper.

Our dataset contains detailed information and allows us to track students and observe the postsecondary education admissions score every time a student takes the exam, as well as the college-course choices a student lists each time they take the exam. In this system, students get an offer from a unique college-course depending on their choices and the course's availability. We have access to the unique college-course students enroll in.[15]

Students in Greece are assigned to public schools through zoning based on their residential address and residential proximity to the school (Goulas, Megalokonomou and Zhang, 2018). Using the student's school address, we deduce whether they reside in an urban or rural area using information on the number of inhabitants per postcode.[16] We also use the postcode level annual average household income (in euro) to define high- (above median) and low- (below median) income neighborhoods. Information on inhabitants and household income is obtained from the National Statistical Authority.[17]

3.2 Sample for Additional Analysis

For a small subsample of the population of schools and students, we have access to additional information. For students in the smaller sample, our data include information on their senior year GPA and birthdate. The senior year GPA includes the average performance on all first- and second- semester school exams in grade 12.[18] These are not standardized exams across the country, but they are representative of the entire student population and are not restricted only to students applying to postsecondary education (as is the case with the national entrance exam). This dataset is also used in other studies (Lavy and Megalokonomou, 2019; Dinerstein, Megalokonomou and Yannelis, 2020; Kedagni, Krishna, Megalokonomou and Zhao, 2021; Lavy and Megalokonomou, 2021) and includes comprehensive data from 23 high schools. As shown in those studies, these schools are representative of the school population with respect to several observed variables, are distributed throughout the country, and cover a diverse set of areas (Figure C2 in Appendix C). These additional data were obtained by visiting each high school in person, obtaining all student-level records, and digitizing them. We have these additional data for the years 2006 and 2007.

[15]The government indicates that only in very specific cases, a student's enrolled college-course may be different from the one they got an offer from, such as students from families with more than 3 children or students with a major disability. In those cases, students may be allowed to enroll in a college-course in the same field of study as the one they got an offer from, with the only difference being that the enrolled one is closer to their parents' residential address.

[16]Urban areas are those with more than 20,000 inhabitants.

[17]We obtained this information at postcode level and aggregated it at neighborhood level.

[18]Note that the school-level GPA units run from 0-20, while the units for the national exams run from 0-20,000.

3.3 Summary Statistics

Table 1 shows a summary of the novel administrative data used in the main analysis. Column (1) includes the entire student population who took the exam in 2006 in Greece, which is almost 70,000 students. Column (2) includes students within a +/-1,500-point bandwidth around the threshold (i.e., admissions scores ranging from 8,500 to 11,500). This sample includes around 14,500 students. The sample size in the analysis changes with each optimal bandwidth calculated for each outcome, and thus we present summary statistics within a +/-1,500-point bandwidth from the threshold. This encompasses all bandwidths that are calculated using a linear polynomial of the running variable.[19]

Column 1 of Table 1 includes students in the entire sample of the 2006 cohort — i.e., students who were eligible to take the exam for the first time in 2006. The admissions score tends to be higher for the entire sample, with a mean of 12,344 (column (1)), relative to the sample closer to the threshold with a mean of 10,045 (column (2)). This is expected, since the sample included in the analysis focuses on students who marginally pass or fail the admissions threshold, i.e, lower-achieving students. Since students' admissions score is calculated differently for each of the five knowledge fields, and students choose up to two knowledge fields, each student gets at most two admission scores. Thus, in our main analysis we define the admissions score as the admissions score used for the offer if a student gets an offer on first attempt, or the highest of the two admissions scores if a student does not obtain an offer on the first attempt.[20]

Around 34% of students failed to secure a postsecondary education offer on their first attempt in 2006 (column (1)). Around the threshold, failure rates are significantly larger - around 52% (column (2)). We then rank college-courses in terms of quality using the admissions score of the last admitted student in 2005. In the entire sample 43% of students receive an offer from a college-course above the median quality (column (1)), while only 6% get offers from high-quality college-courses in the lower-ability sample (column (2)). "Offer from Academic Institution" is a binary indicator that takes the value 1 if the offered postsecondary institution is an academic university and 0 if it is a vocational school or the student is not admitted to any postsecondary institution. In the entire sample, around 51% of students get offers from an academic institution (column (1)), while only 20% of the lower-ability sample get offers from academic institutions (column (2)).

Overall, 21% of students in the entire sample retake the exam at some point after their first attempt,

[19]Certain bandwidths are calculated using a quadratic polynomial, and they are usually $>$1,500.

[20]We show in a robustness exercise that our results remain similar if we use the highest of the two admissions score for everyone (Table 8).

and most (19%) do so in the next consecutive year, i.e., in 2007 (column (1)). Of the sample within the +/-1,500-point bandwidth from the threshold, 43% retake the exam at some point after their first attempt, while 41% retake the exam in the next consecutive year (column (2)). Moreover, most students who retake tend to do so only once and in the following year after their first attempt. Regarding those who do not fail in absolute terms (but failed in relative terms, which means that they still scored an admissions score >10,000), 12% of them retake the exam in the entire sample (column (1)) and 25% of students retake the exam in the +/-1,500-point bandwidth sample (column (2)). The retake rate is much higher among those who fail the exam in absolute terms (admissions score <10,000) in both samples.

We then show summary statistics for students' postsecondary applications. Students with admissions scores around the threshold express more college-course choices (around 32 college-course choices) than the entire sample (around 26 college-course choices). For the entire sample, 56% of students opt for academic tracks and 49% choose top college-courses (column (1)). In the +/-1,500-point bandwidth sample, only 24% of students choose academic college-courses and only 12% focus on a high-quality college-course (column (2)).

A similar proportion of students who apply for a postsecondary college-course have graduated from urban high schools in the two samples (columns (1) and (2)). Students in the +/-1,500-point bandwidth sample are slightly less likely to have graduated from a top-performing school. We measure school-level performance by the average admission score of all applicants in that school the year before (2005). We define as top-performing schools those whose school-level performance is above the median performance. We also notice that students in the low-performing sample have a slightly smaller household income than those in the entire sample, and are slightly less likely to be female; however, these differences are relatively small and not statistically different from zero.

4 Empirical Strategy

4.1 Baseline Model and Specification

We identify the relationship between retaking the exam and resulting outcomes by exploiting a regression discontinuity design (RDD). In particular, we rely on the change in the postsecondary education admissions rule that took effect in Greece in 2006. This policy imposed an important threshold at a score of 10,000 out of 20,000. Students gain access to postsecondary education only if they score at least 10,000; in previous years, anyone could get into postsecondary education, subject to vacant slots.

This new admission rule generated a discontinuity in the treatment assignment (i.e., receiving a post-secondary education offer) around the threshold. We are interested in establishing the causal effect of passing the performance threshold on later outcomes for similar students who fell on either side of the performance threshold simply because they performed relatively better or worse than the performance threshold.

We note that not everyone below the threshold is refused access to tertiary education, and thus some students below the threshold may not have to retake the exam. This is because the government applies some affirmative action policies, and means that students who do not achieve the required performance threshold might still be admitted to postsecondary education and will not have to retake the exam, simply because they belong to a group that receives beneficial treatment. Since there is only partial compliance around the threshold, the identification strategy we use is a fuzzy RDD (Angrist and Lavy, 1999; Hahn, Todd and der Klaauw, 2001). The imposed postsecondary performance threshold serves as an exogenous source of variation for retaking decisions. In particular, a binary indicator that shows whether a student fails to achieve a score of 10,000/20,000 is used to instrument for whether a student has to retake the exam.

To estimate the effects of retaking the exam a year later[21] on subsequent outcomes, we employ a fuzzy RDD to account for potential confounding factors. In particular, we instrument the probability of retaking the exam in 2007 for student i in school s ($Retake_{is}$) with an indicator for whether the admission score was below 10,000 (T_i):

$$Retake_{is} = \alpha_1 + \alpha_2 T_i + \gamma_1 f(score_i) + \gamma_2 T_i f(score_i) + \mu_{is}, \tag{1}$$

where $f(score_i)$ is a linear or quadratic polynomial of the distance between student $i's$ score and the 10,000 threshold. Our main regression is thus given by

$$Y_{is} = \beta_1 + \beta_2 Retake_{is} + \theta_1 f(score_i) + \theta_2 T_i f(score_i) + e_{is}, \tag{2}$$

where Y_{is} is the outcome of interest for student i in school s. Outcomes of interest include (for example) a student's most recent performance on the exam, the quality of the most recent offer received or an indicator of whether the student received an academic offer or not. The above specifications produce first- and second-stage 2SLS estimates. Standard errors are clustered at the school level to allow for heteroskedasticity and serial correlation at that level, since students who attend the same school may

[21] Our main analysis examines the effects of retaking the exam in 2007, because the vast majority of students retake the exam only once (in 2007). We conduct a robustness check in Section 8 where we focus on students' outcomes in the latest attempt they take the exam.

share some error patterns.[22] To further check the robustness of our findings and in addition to the 2SLS estimates, we apply a nonparametric method that relies on local polynomial (fuzzy) regression discontinuity point estimators with robust-bias corrected (RBC) confidence intervals and we call these CCT estimates given that the procedure follows Calonico, Cattaneo and Titiunik (2014a,b).

We use the mean squared error (MSE) optimal bandwidths computed according to the method proposed by Calonico, Cattaneo and Titiunik (2014b).[23] Our nonparametric estimates come from a local polynomial regression that uses a triangular kernel. This procedure assigns higher weights to observations closer to the threshold, following the approach proposed by Calonico, Cattaneo and Titiunik (2014a) and Calonico, Cattaneo, Farrell and Titiunik (2019).

4.2 First-Stage Results

Figure 1 illustrates the first stage. The x-axis describes the distance of the student's score from the threshold of 10,000/20,000 (indicated by the distance from the vertical line at 0); the y-axis describes the probability to retake the exam at different distances from the threshold. The sample is restricted to students who obtain a score in the +/- 1,500-point sample (i.e., student performance between 8,500 and 11,500). First, we see that retaking the exam is quite common, with an average of 40% for the sample in the figure.[24] There is a clear discontinuity in the probability to retake the exam for students around the threshold. In particular, the probability to retake among below-threshold students is 0.5. For students above-threshold students the probability to retake ranges from 0.4 (just above the threshold) to around 0.27, depending on their distance from the threshold.

As a robustness check, we show that there is no first stage in the years prior to and after the reform. In particular, we use information on 2004 and 2005 (2 consecutive years before introduction of the performance threshold) and 2010 and 2011 (2 consecutive years after lifting of the performance threshold). Figure 2 presents the first stage for these placebo years. The x-axis describes the distance of the student's score from the threshold of 10,000/20,000; the y-axis describes the probability to retake the exam

[22]To check the validity of our results we also run the regressions with robust standard errors. Results, available upon request, show the robustness of our main findings.

[23]According to Calonico, Cattaneo and Titiunik (2014b), when researchers apply the MSE optimal bandwidth, it yields the MSE-optimal regression discontinuity treatment effect estimator, which is invalid for inference. RBC inference methods are valid when using the MSE-optimal bandwidth, but they yield suboptimal confidence intervals in terms of coverage error. In this paper, the authors establish valid coverage error expansions for RBC confidence interval estimators and use these results to establish new inference-optimal bandwidth choices for forming these intervals. Thus, we follow their work and use RBC confidence intervals and MSE optimal bandwidth throughout the paper.

[24]Figure A2 in Appendix A illustrates the first-stage estimates for the entire distribution of scores, and not only for the +/-1,500-point sample. We notice that there is a considerable discontinuity of retake rates around the performance threshold. We also notice that for students far to the left of the threshold, the retake rate jumps to almost the frequency of retake rates for students just to the left of the threshold.

at different distances from the threshold. Panels A and B of Figure 2 presents the first stage for cohorts 2004 and 2005, before introduction of the threshold. Panels C and D present the probability to retake the exam the first year after lifting the threshold (2010) and a year later (2011), respectively. In all these years, we see that the probability of retaking the exam is continuous around the threshold, as expected. Consistent with the institutional setting, this evidence reassures us that we are indeed estimating the effects of the 2006 reform.

The exclusion restriction requires the instrument to affect student outcomes only through the effect it has on retaking and not through other channels (such as university eligibility). If this assumption is not satisfied, then one would need to focus on the reduced-form parameter and look at the effect of scoring just below the 10,000 threshold on later outcomes. However, we believe that this assumption is likely not violated in our design. For outcomes such as student academic performance in the exam and university offer, the exclusion restriction is likely satisfied because retaking is the only channel through which such outcomes can change (Goodman, Oded and Smith, 2020). Students take the national exams and apply to universities (by submitting a list of preferences of the desired university degrees they would like to apply to) within a couple of weeks time, which means that they have to a large extend pre-decided their university application strategy. In fact, students submit their preference list without even knowing what the admission thresholds in this particular year are. So there is little time to adjust their university applications plans or be affected by the national exams that they just took. Nevertheless, one may still be concerned that effort and other behavioural outcomes may be discontinuous around the performance threshold for retakers.

We empirically examine whether conditional on retaking the national exams, students are likely to be similar around the performance threshold. First, as shown in Table 2, conditional on retaking the exam, students above the threshold are similar to those below the threshold with regards to their observable characteristics. In particular, we show that compliers - those who retake the exams because they missed the threshold in the first attempt - are similar to: a) the pool of all retakers within the +/-1,500 bandwidth around the threshold and b) students who retook the exam although they scored just above the performance threshold in their first attempt.[25] We note that there are slightly fewer girls than boys retaking the exam below the threshold relative to the overall retaker mean. However, retakers below the threshold are overall very similar to the mean retaker within the bandwidth of interest. Though this does not exclude the possibility that compliers are unusual with respect to unobservable characteristics like

[25]For this exercise we choose a +/-1,500 bandwidth since it is the closest to the optimal bandwidth and we use it in our preferred linear specifications for the different outcomes of interest.

motivation, the results from Table 2 suggest that compliers are not outliers along observable dimensions. This is reassuring and provides further evidence that it is students of similar cognitive characteristics, who attend similar schools in postcodes of similar socio-economic backgrounds who retake the exams while they had randomly ended in either side of the threshold in the first attempt. [26]

5 Validity of the Estimation Strategy

5.1 Manipulation Around the Threshold

Our main identifying assumption in the RDD setting is that the density of treated observations just above the threshold should be approximately similar to the density of control observations just below the threshold. In the context of high-stakes exams, the manipulation of the running variable (admission grade) is a legitimate concern and has been shown to be present in the school system, especially close to important performance thresholds (Diamond and Persson, 2016; Dee, Dobbie, Jacob and Rockoff, 2019). In our setting, one may expect manipulation to be present if graders inflate the scores of students who just fail to pass the newly introduced threshold.

Figure 3, Panel A presents a histogram of the running variable around the performance threshold. In Panel B, we follow the approach proposed by Cataneo, Idrobo and Titiunik (2019) and present the actual density estimate with 95% confidence intervals. Both figures provide clear evidence that the density of the admissions score around the threshold displays no systematic differences below and above the performance threshold. Moreover, the overlap of the confidence intervals in Panel B is further supported by the formal test, which yields a p-value of 0.241 for the null that there is no difference in the density of treated and control observations at the threshold. This is not surprising, given that these are standardized exams that are graded by different professional graders under specific rules and in anonymity. The names of the student and the school are also blinded in these exams. Two or more graders (in different geographical areas) grade each exam script and the overall grade is the average of the graders' scores. Additionally, a student's admissions score is made up of an average of six or seven separate subjects. Manipulating the admissions score would require cooperation among several graders in a systematic manner, which is not feasible in this institutional setting.

[26] Further discussion of the assumptions used for our identification strategy is provided in Appendix B.

5.2 Balancing Tests on Observables Around the Admissions Threshold

Another key assumption required in a RDD setting is that the treated and control observations are similar on average around the threshold in terms of observable predetermined characteristics. Figure 4 provides graphical evidence regarding the validity of this assumption. Our results show that students just above and below the newly imposed performance threshold are equally likely to be girls (Panel A), are equally likely to be attending a top-performing high school (Panel B), are equally likely to be attending an urban high school (Panel C) and are equally likely to be located in a high income area (Panel D). These figures provide evidence that students are randomly distributed around the threshold with respect to their observed characteristics.

6 Results

6.1 Effect of Retaking on Subsequent Academic Performance

Table 3 shows the estimates of retaking the exam on students' subsequent academic performance on the exam.[27] Each column corresponds to a different bandwidth and polynomial specification of the running variable. We report 2SLS estimates (columns (1)-(5)), as well as robust CCT estimates (columns (6) and (7)) obtained following Calonico, Cattaneo and Titiunik (2014a) and Calonico, Cattaneo, Farrell and Titiunik (2019). The optimal bandwidth in columns (4)-(7) is obtained following Calonico, Cattaneo and Titiunik (2014b). We use either a linear (columns (1)-(4) and (6)) or a quadratic (columns (5) and (7)) polynomial of the running variable whose slope is allowed to change at the performance threshold. Panel A shows our first-stage estimates across all columns. Students who just fail to pass the performance threshold are 8-12 percentage points more likely to retake the exam. The instrument is strong in all 2SLS specifications, with the Kleinbergen-Paap Wald F-statistic being consistently greater than 10. The magnitude of the coefficient of the first stage is, as expected, not so large, since retaking the exam is common throughout the entire distribution of scores. Nevertheless, all first-stage estimates are positive and statistically different from zero, suggesting that there is a clear causal positive impact of missing the threshold on the probability to retake the entrance exams.

Columns (1)-(3) of Table 3 show 2SLS estimates when the distance from the performance threshold increases from +/-1,000 (column (1)), to +/-1,500 (column (2)) and +/-2,000 (column (3)). Our preferred specifications are those in columns (4) and (6) and use the optimal bandwidth, as defined by Calonico,

[27] We consider retaking a year after because almost everyone who does retake the exam does so in the subsequent year.

Cattaneo and Titiunik (2014a) and a linear polynomial of the running variable. Panel B presents the effects of retaking the exam in 2007 on student academic performance on the exam. The estimated effects are positive and statistically significant across all specifications. Our preferred specifications indicate that retaking the exam leads to improved performance for students of 2,282-2,588 points.[28] This corresponds to a 0.6 of a standard deviation increase in their admissions score. This level is similar to what Goodman, Oded and Smith (2020) find using retaking the SAT in the US for low-scoring students. We report robust confidence intervals in all remaining tables instead of standard error as this is preferred for inference (Calonico, Cattaneo, Farrell and Titiunik, 2019).

6.2 Grade Inflation

The results reported above show that students improve their academic performance when they retake the exam. In this section, we examine grade inflation as a possible mechanism through which students may perform better. We use the additional comprehensive dataset for the smaller sample of schools and students we describe in Section 3.2.[29]

One could worry that teachers responded to the minimum performance threshold by inflating scores of students in an attempt to help them pass the threshold. This is more likely to happen in 2007 when the minimum performance threshold is anticipated and teachers know this in advance of reporting students' first and second semester grades compared with 2006 when the policy was unanticipated. To address this concern, we focus on the distribution of students' high school GPA instead of the admission scores. The reason is that we want to include all students in this exercise and not only those who desire to gain access to tertiary education and thus take the national exams. We present the GPA distribution of lower-achieving students to obtain a picture of whether teachers are likely to inflate the grades of students in 2007 compared with 2006. Figure 5 shows the distributions of GPA for 2006 and 2007 for lower achieving students holding a GPA below 12/20 and above 8/20 and who are our sample of interest. There is no

[28]We also estimate the main specification for different sub-groups: (a) by income, namely for students who reside in regions with a postcode-level average household above and below the median, separately, in Table A2, (b) by gender, in Table A3, (c) by high school quality, in Table A4, (d) by type of high school, namely for students who attend public and private school, separately, in Table A5 and (e) by age of the students, namely for 17- and 18-year-olds, in Table A6. We find that, if anything, the effects are slightly more pronounced and precise for students attending schools in high income regions, boys, high quality high schools, private schools and older students. However, in most of those cases we fail to reject equality of the parameters between the different subgroups or the value of the Kleinbergen-Paap Wald F-statistic takes values below 10, indicating that we cannot draw strong or general conclusions about our heterogeneity exercises.

[29]We replicate the main estimation in the smaller sample and the estimated effects are shown in Table A7 in Appendix A. These effects are almost the same as the ones using the entire sample. In particular, the main effect of retaking on academic performance is 2,282 scores in the entire sample and 2,396 scores in the small sample using the 2SLS estimation (under the preferred specification), with very similar confidence intervals. The main estimate is 2,589 scores in the entire sample and 2,864 scores in the small sample using the CCT estimation (under the preferred specification), with very similar confidence intervals.

evidence that teachers inflated student grades between 2006 and 2007, since the GPA distributions are almost identical for the low-achieving students with the p-value of the Kolmogorov-Smirnov test being equal to 0.978. Overall, we show that grade inflation is not the driving mechanism behind students' observed improvement in their admissions score from retaking. We conclude that it is reasonable to assume that the observed improvement in performance is the result of additional study and learning.

6.3 Effect of Retaking on Offers

Table 4 presents the estimated effects of retaking on students' quality of subsequent offers. We measure the quality of an offer using two measurements: whether the average college-course quality is above or below the median[30], and whether it is an offer from an academic institution or a vocational one. Columns (1)-(5) show 2SLS estimates, while columns (6) and (7) show CCT estimates. Our preferred estimates are in columns (4) and (6), in which linear polynomials of the running variable are used. Panel A shows the results of the effect of retaking the exam on whether the offer is from a college-course with quality above the median. The estimates range from 0.101 to 0.154, but are relatively imprecise. This suggests the lack of sufficient evidence for an effect. This pattern is robust across different bandwidths (+/-1,000, +/-1,500, +/-2,000, and optimal), across different estimations (2SLS and CCT), and when using a quadratic instead of a lineal polynomial of the running variable (columns (5) and (7)). In Appendix A, Table A1, Panel A, we also show the impact of retaking on a continuous measure of quality (defined as the admission threshold in 2005).[31] The estimates are in line with those from Panel A in Table 4, showing a positive but imprecisely estimated effect of retaking on the quality of the received offers.

Panel B in Table 4 shows the effect of retaking the exam on whether the student is more likely to obtain an academic as opposed to a vocational offer (or no offer). On average, college-courses from academic institutions require higher grades and are more prestigious. The outcome variable here is a binary indicator that takes the value of 1 if the student obtains an offer from an academic college-course and 0 if they obtain an offer from a vocational college-course or no offer. Our preferred specifications (columns (4) and (6)) show that there is a positive impact of retaking on the students' probability to obtain an academic offer. The estimated coefficient is not equally precise across different bandwidths, estimations, and polynomials.

Overall, we notice that despite the significant improvement in their exam performance due to retak-

[30] We use a standardized index of quality that is measured based on the 2005 admission thresholds for each college-course.

[31] We define the measure using the local minimum admission threshold in 2005, with higher value indicating a college-course of higher quality.

ing, the evidence on whether students manage to receive offers from highly ranked college-courses is positive but sometimes imprecisely estimated.[32] [33] Retakers improve their scores and are more likely to obtain offers from academic as opposed to vocational. The coefficient on improving the quality of their offers is positive but imprecisely estimated.

As additional evidence, we present the impact of retaking the exam in all the years during which the policy of the minimum threshold is in place (i.e. 2006-2009) to see if we get more precise estimates on the improvement in the quality of offers. In Table 6 we can more clearly see that students improve their academic performance in all years by a similar magnitude (Panel B), as well as the results for their offer-related outcomes (panels C and D).[34] The first affected cohort (2006) has the least precisely estimated outcomes in Panels C and D, but in years 2007 and 2008 the students improve not only their academic performance but also the quality of their offers and estimates are precise.[35] We next study the mechanisms behind students' improved outcomes and the apparent delay in the improvement for the first cohort (2006) affected by the reform of the minimum threshold.

7 Mechanisms

7.1 Changes in Students' Application Choices

Changes in students' application choices could be one of the mechanisms through which students improve their outcomes upon retaking, in addition to their improved scores. Two available possibilities

[32] In Appendix A, Table A1, Panel B, we also show the impact of retaking on entering a college-course with better employment prospects and find that the impact is also not precisely estimated. We classify college-courses based on an employability index that relates to the job prospects and job insecurity associated with each college-course. This index reflects the fear of involuntary job loss and is collected from a series of long-term surveys of university graduates in Greece discussed by Goulas and Megalokonomou (2019). For each university department, this index indicates whether the employment prospects for students after graduation are good, mediocre, poor, or very poor. More information about how this index is constructed is provided in the note for Figure C3 in Appendix B. We classify each college-course as a high or low employment prospect: college-courses with high employment prospects those that have an initial employability between 1 and 1.5, or in other words a college-courses degree with good employment prospects. Following the same logic, we define as college-courses with low employment prospects those that have an initial employability index between 1.5 and 3, or in other words a college-course degree with mediocre, poor or very poor employment prospects.

[33] We also evaluate the impact of retaking on students' field of study offers and choices. We categorize the college-courses into fields of study and use the following groups: Health, Engineering, Science, Humanities and Business. Results are presented in Appendix A in Table A8. Panel A shows that while there are no statistically significant effects of retaking the exam on the probability to get an offer for a college-course in a health-, engineering-, sciences-, or humanities-related college-course, there is a large drop in the likelihood to get an offer from a business-related college-course. Panel B indicates that there are no statistically significant effects of retaking on the probability for a student to report at least 50% of their college-courses choices in a health-, engineering-, sciences-, or humanities-related college-courses. However, students are less likely to report at least 50% of their college-course choices on a business-related college-course. Although there seems to be a drop in the likelihood of a business-related offer and choices, we are unable to draw strong conclusions about any systematic pattern in preferences for college-courses from retaking.

[34] We note that cohort 2009 improves their academic performance and outcomes significantly less than the remaining cohorts. This is partly due to the impact of the policy being lifted in the year 2010, i.e. the year that the 2009 retakers retake the exam.

[35] The results are robust to the use of a quadratic polynomial instead of a linear polynomial.

emerge: students improve their admission scores but also become a more ambitious in their application choices by listing higher quality college-courses. Alternatively students keep application choices constant and their improved outcomes are a result climbing up the ranks of their existing applications. An additional alternative, in particular for the first cohort of students who retake in the presence of the newly-imposed threshold, is that students possibly become overly ambitious and optimistic in their application forms, and list college-courses that are not achievable given their exam scores, explaining why our estimates are imprecise for the first cohort. In Table 7, we examine the effects of retaking on a series of outcomes related to student college-course choices: the number of choices students list on their application form (Panel A),[36] whether students report more than 50% of their college-course choices being above median quality (Panel B), whether more than 50% of their choices are academic compared with vocational (Panel C).

Table 7, Panel A presents estimated effects which are small and statistically insignificant. This indicates that students do not change the number of college-courses they list on their preference list upon retaking the exam. The intuition behind looking at the number of choices is that the more choices a student lists, the more risk-averse they may be. The structure of the table is similar to the one in the main results, with the only difference being the use of different outcome variables. Panel B of Table 7 shows that students do not report more top college-courses upon retaking. This indicates that students do not become more ambitious when resubmitting their college-course choices upon retaking. The estimated effects are small and not statistically different from zero across all specifications.

Academic college-courses on average require higher exam performance than vocational courses and are perceived to be more prestigious. Since students obtain a higher score upon retaking, we then examine whether students are more likely to apply to academic college-courses. In Table 7, Panel C does not provide any evidence that students list more academic than vocational choices on their preference lists, since all estimates are close to zero and statistically insignificant.

These estimates clearly suggest that a change in student application choices is not associated with better students' better outcomes upon retaking. Retakers do not seem to change their application choices to become more (or less) ambitious compared with their first-attempt applications.

[36]There is no limit on the number of choices students can list on their application forms in the Greek application system.

7.2 Changes in the Overall Skill Level of the Applicant Pool

We next examine the role of the change in the skill level of the pool of applicants as a potential reason why the improved performance of the first affected cohort of retakers does not immediately translate into an improved postsecondary placement. Figure 6 shows the distribution of admissions scores before, during, and after lifting the minimum threshold.[37] The reform only targets low-achieving students and thus discourages them from applying. Panel A shows the distribution of scores for the pre-reform years, i.e., 2004 and 2005. We notice that the two distributions are very similar. Panel B in Figure 6 shows the distribution of admissions scores for years before the reform (2004, 2005) and while the reform was in place (2006-2009). We notice that the distribution of admissions scores shifts to the right during the years of the reform, relative to before, which indicates the presence of fewer low-achieving applicants. Panel C in Figure 6 shows the distribution of admissions scores for years in which the minimum threshold policy was in place (2006-2009) and after it was lifted (2010-2012). We notice that after lifting the minimum threshold, the distribution of scores slowly shifts to the left, close to what it was before the introduction of the policy. Overall, this pattern indicates that retakers who took the exam for the first time in 2006 and are retaking in 2007 are now faced with a different pool of competitors. Their competitors are now students who are taking the exam for the first time in 2007 and, as shown in Figure 6, the distribution of admissions scores of the 2007 cohort is to the right of the 2006 applicant cohort.

We examine the effects of the change in the competitors' skill-level on a series of counterfactual possible outcomes for the retakers of 2006 in Table 5. The idea is to evaluate whether if the first cohort of retakers had retaken during a year where there wasn't a large shift in the skill-distribution of their peers. We divide students into 100 bins based on their exam score and examine potential changes in their relative position in the distribution upon retaking. Panel A in Table 5 shows the counterfactual change in the 2006 retakers' rank had they retaken the exam competing against their own cohort — i.e., the 2006 cohort.[38] All estimated effects are positive and statistically different from zero. Columns (4) and (6) show our preferred estimates — namely, the linear 2SLS and CCT estimates using optimal bandwidths according to Calonico, Cattaneo and Titiunik (2014a). The estimates indicate that students would have

[37]We observe a significant drop in the number of applicants during the 4 years the policy was in place compared with the years pre- and post-policy. Figure A3 in Appendix A shows the drop in the number of applicants over the years. The policy of the minimum performance threshold is introduced in 2006 and is in place until (and including) 2009. The drop in number of applicants is largest the first two years of the reform after which the number of applicants stabilises. The first year the reform is lifted (2010), the number of applicants jumps again close to the pre-reform number of applicants.

[38]For this we use students' exam performance in 2007 and calculate what their relative rank would be in 2006 with their 2007 exam performance.

improved their rank position by 11 to 14 percentage points. The improvement in student performance is stable across the different specifications. The F-statistic for the instrument remains above 10 across all specifications.

Panel B of Table 5 shows the actual change in the rank of the retakers in the year they retook the exam, i.e., in 2007.[39] Students from 2006 who retook in 2007 are competing against the 2007 student cohort. The estimated effects are small and imprecise, following the pattern we observe in Table 4 Panel C, indicating no improvement in their rank when they are competing against the 2007 cohort. This suggests that despite significantly improving their exam performance, students' relative position in the distribution does not change significantly, possibly due to an increase in the overall skill level of their competitors.

Finally, Panel C in Table 5 shows the counterfactual change in 2006 retakers' rank had they been competing against the 2011 cohort — i.e., the first cohort after the minimum threshold is lifted. We use the 2011 cohort because the skill level of the applicant pool changes in 2011 (as shown in Figure 6) and becomes similar to that of the 2005 cohort, which means that the applicant pool again includes students with lower scores. In Panel C, we notice that all estimates are positive and statistically different from zero, similar to Panel A. Based on our preferred specifications, in columns (4) and (6), retakers would have improved their position by 15 to 18 percentage points, similar to Panel A.

Moreover, Table 5 shows that the first cohort of retakers obtain less presicely estimated postsecondary admission offers, partly because the competition in the year they retake the exam (in 2007) changes as a response to introduction of the minimum performance threshold. In particular, there are fewer low-achieving students now, and thus the competition to gain access to postsecondary education becomes more intense. Retakers are therefore not in a relatively better position with respect to other first-attempt applicants, and thus may not obtain better quality of offers. Retakers would have improved the quality of their postsecondary offers if they had retaken the exam in a year in which the minimum threshold to enter postsecondary education was not in place.

[39]For this we use students exam performance in 2007 and calculate what their relative rank would be in 2007.

8 Robustness Checks

8.1 Alternative Bandwidths

We then examine the robustness of our main results. In this subsection, we show the sensitivity of our results to using different bandwidths. In our main specification, we use the MSE optimal bandwidth, which we also present here as a benchmark.

In Figure 7, we change the bandwidth used for local polynomial estimation, and we show results using four bandwidths suggested by Cataneo, Idrobo and Titiunik (2019): the MSE optimal bandwidth (used in main results), the double of the MSE bandwidth, the coverage error (CER) optimal bandwidth, and the double of the CER optimal bandwidth. We show results for the three main outcome variables that we used in Tables 3 and 4 while using these different bandwidths. The figure notes detail which numbers correspond to each bandwidth for each of the three different outcomes.

Panel A in Figure 7 shows that the estimated coefficients on academic performance remain stable when changing the bandwidth. As expected, smaller bandwidths are less precisely estimated. In Panel B, we present the estimates of obtaining an offer from a college-course above the median quality. We find that our main (and imprecise) results change little if we use different bandwidths, although, as expected, they do become more precise when we use larger than optimal bandwidths. Panel C shows the effect of retaking on the likelihood of obtaining an academic as opposed to a vocational offer. The coefficient remains stable, and as we restrict the bandwidth the estimates become less precise. The results indicate a similar coefficient when changing the bandwidth.[40]

8.2 Pseudo-Thresholds

Figure 8 presents robustness exercises based on placebo measures of treatment—namely, when we replace the true minimum performance threshold with pseudo-minimum performance thresholds (Crost, Felter and Johnston, 2014; Dahl, Løken and Mogstad, 2014). In particular, we use pseudo-thresholds equal to 9,000, 9,500, 10,500, and 11,000 instead of the 10,000, which is the real one. We show the estimated effects for our main four outcomes in the different panels. All estimates using the pseudo-thresholds are small, have inconsistent signs, and are statistically insignificant. Notable are the dif-

[40] In Figure A4 in Appendix A we repeat this exercise; the outcomes are the different measures of application choices that we used in Table 7. In particular, the outcomes are: (A) the number of choices students list on their preference form, (B) whether a student reports at least 50% of their college-course choices being above median quality, (C) whether at least 50% of student choices are academic compared with vocational. We notice that the estimated effects remain similar when using the alternative bandwidths.

ferences between the estimates from the placebo regressions and from those obtained when the true minimum performance threshold is used, especially for the outcome that was positive and statistically significant before – i.e., academic performance (Panel A). The remaining outcomes are less precisely estimated, with large confidence intervals, and have different signs than the true effects (Panels B, and C). These patterns are confirmed by Table A9 in Appendix A, which shows all the estimates reported in Figure 8. The third row in each panel corresponds to the true performance threshold, while all other rows correspond to pseudo-thresholds as reported in column (1). The lack of any discernible effects when the pseudo-thresholds are used suggests that the estimated treatment effects are not spuriously picking up any effects of unobserved confounders and that there is no misspecification in the relationship between the running variable and the outcome (Crost, Felter and Johnston, 2014).[41]

8.3 Eliminating Observations

One could be concerned that our results may be sensitive to observations that are located very close to the minimum performance threshold. This could be an issue, especially if there is systematic manipulation of the admissions scores around the threshold (Scott-Clayton and Zafar, 2019; Arnold and Freier, 2013; Dahl, Løken and Mogstad, 2014). Figure 9 presents further evidence that our main results remain stable when we eliminate observations around the threshold. As before, we show results for all three main outcome variables. We eliminate observations for students with admission scores within +/-100 and +/-200 points from the threshold. By doing so, we expect our coefficients to remain stable but the precision of the estimates to decrease. In all three figures, as we move to eliminate up to 200 observations around the threshold, we see that the point estimates are quite stable. We notice that as the confidence intervals gradually increase the estimates become less precise, as expected.[42] These findings are also confirmed by Table A10 in Appendix A. The radius of 0—first row in each panel—corresponds to the main estimated effects when the threshold is 10,000. The second and third row in each panel corresponds to when we eliminate +/-100 and +/-200 scores around the threshold, respectively. The estimated effects remain very similar when we eliminate those observations around the threshold, while the confidence intervals slightly increase.

[41]Figure A5 in Appendix A shows that the effect of retaking on the measures of choices used in Section 7 are also robust to using placebo thresholds.

[42]Figure A6 in Appendix A shows that the effect of retaking on the measures of choices used in section 7 are also robust to eliminating observations around the threshold.

8.4 Different Admissions Scores

We conduct two robustness exercises in which we use different admissions scores to check the sensitivity of our estimates. First, as we mention above, students obtain different admissions scores depending on the scientific field they apply to. Students who obtain offers are assigned an admissions score that corresponds to the offered college-course and we use this in all tables. It is not straightforward which admissions score to assign to students who do not obtain an offer, simply because they do not have an offered college-course. In the main analysis, we assigned to them the highest admissions score based on the knowledge fields they listed in their preference list.

Our first check consists in assigning the highest admissions score to all students regardless of whether they have obtained an offer, thus comparing everyone's best possible performance in the exam. The results of this robustness check are shown in Table 8. The estimates for academic performance (Panel A) slightly increase in magnitude, but the pattern is similar to the main results shown in Table 3. Panels B and C show estimates that are smaller in magnitude, but again the main pattern remains the same as in the main results of Table 4.[43]

Second, in the main analysis we focus on the effect of retaking the exam in 2007, which is the first time students retake the exam.[44] We conduct another robustness exercise in which we use the admissions score from a student's latest attempt of retaking the exam. These results are presented in Table A11. The estimates remain positive and slightly drop in magnitude by 5-10%, although they are not statistically different from the ones used in the main analysis. These results suggest that our main results are very robust to using different admissions scores.

9 Conclusion

This is the first paper that measures the causal effect of retaking a high-stakes admissions exam on student subsequent outcomes using an exogenous unexpected policy that introduced a minimum performance threshold for postsecondary admission. We focus on low-achieving students due to the nature of the policy. The policy was unexpected the first year it was implemented in 2006 and changed the decision rule related to obtaining access to tertiary education in Greece. In particular, students who

[43]Note that the number of observations for the alternative admissions score are slightly fewer than the observations in the main analysis of Table 3. This is a result of a small number of students making mistakes in their applications list and choosing college-courses that they are actually not eligible to apply to since they did not take an additional examination required for certain college-courses. In such cases, the student obtains a missing value next to the college-course that they are not eligible for.

[44]The reason is that the vast majority retakes the exam only once.

fail to obtain 10,000/20,000 scores (i.e., 50%) on the postsecondary admission exam at the end of high school fail to gain access to tertiary education. The only way for these students to access postsecondary education is to retake the exam and pass the required performance threshold in a subsequent period. Students who pass the critical threshold were allowed to compile a list of preferences of postsecondary college-courses.

We estimate the impact of retaking the national exam on academic performance and quality of post-secondary offers. We use rich new administrative data for the universe of students applying to enter postsecondary education, which we obtained from the Ministry of Education in Greece. We rely on detailed student records on exam performance, postsecondary applications, and admission for each time a student retakes. We also have access to an additional comprehensive sample for a smaller number of schools and students that we use for some additional analyses. Since there is only partial compliance, we use a fuzzy regression discontinuity design. Comparing students who just failed the performance threshold to those students who just passed the performance threshold eliminates any selection bias due to observed and unobserved teacher, class, and school characteristics. We support this identification approach with evidence that clearly indicates that students who just failed to pass the critical threshold are identical in terms of their observed characteristics to students who just passed the critical threshold, and can thus serve as a proper counterfactual group.

The results suggest that students who retake the exams improve their performance by just over half a standard deviation. The underlying mechanism is through learning and studying, while we provide evidence that the exam improvement is not explained by teachers inflating students' grades as a response to the policy. We also find that retakers improve their outcomes upon retaking, by means of being more likely to obtain offers from academic as opposed to vocational college-courses and by getting accepted in college-courses of higher quality. We evaluate the role of students' application forms in explaining their improved outcomes and find that students do not change their application strategies in terms of the quality of their choices or their risk-taking behavior.

We find that the first cohort of retakers have positive coefficients in our estimates of improved outcomes, but these are imprecisely estimated. In particular, we find that they are not necessarily more likely to obtain an offer from a college-course of higher quality. We highlight a mechanism stemming from the newly imposed minimum threshold that could potentially drive these results: the strength of the competition that the first cohort of affected retakers face is now much higher and affects their probability of obtaining better offers. We show that if they had been competing against their contemporary peers or other cohorts not affected by the imposition of the threshold they would have improved their outcomes.

This is the first study that highlights the role of the strength of the competition in settings where a minimum threshold is imposed and thus the overall skill level in the applicant pool changes. Although retakers experience an increase in their performance, there is a short adjustment period in which they do not find themseleves in a better relative position for obtaining a more prestigious postsecondary placement. The main objective of the policy was to not allow students with particularly low performance to enter postsecondary education. The results presented above suggest that one must carefully assess the costs and benefits of retaking the exam and of minimum threshold policies (Jacob and Lefgren, 2009). In doing so, one should consider the trade-offs between potential benefits — absolute and relative — and potential costs. Understanding how these policies affect students' current and subsequent outcomes is important when designing those policies. Policy makers should not ignore that restricting access to postsecondary education may introduce selection with respect to applicants' characteristics.

The method we use to study the effects of the introduction of a minimum threshold is general and could be applied in other contexts. Any context in which rules or laws create thresholds in the implementation of events, policies and programs — such as competitive exams, merit scholarships, participation programs and poverty programs — could profit from our approach of using detailed individual records to compare those who are just affected by the rule with those who are just not affected by the rule and uncover the mechanisms behind those effects. The benefits of combining exogenous variation with rich administrative data for the universe of students who apply to postsecondary education institutions to study retaking decisions are that (1) we focus on low-achieving students who are very likely to give up education, (2) we rely on a setting that depends on objective definitions of student merit, (3) we are able to evaluate the strength of the competition related to the applicant pool, and (4) we are able to rule out alternative mechanisms which are common threads in the literature such as grade inflation as a potential driver of the effects of interest.

References

Andresen, Martin Eckhoff and Sturla Andreas Løkken, "The Final straw: High school dropout for marginal students," MPRA Paper 106265, University Library of Munich, Germany December 2020.

Angrist, Joshua D and Jörn-Steffen Pischke, *Mostly harmless econometrics: An empiricist's companion*, Princeton university press, 2009.

Angrist, Joshua D. and Victor Lavy, "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement," *Quarterly Journal of Economics*, 1999, *114* (2), pp.533–575.

Arnold, Felix and Ronny Freier, "Signature Requirements and Citizen Initiatives: Quasi-Experimental Evidence from Germany," Technical Report 2013.

Avery, Christopher, Oded Gurantz, Michael Hurwitz, and Jonathan Smith, "Shifting College Majors in Response to Advanced Placement Exam Scores," *Journal of Human Resources*, 2018, *53* (4), pp.918–956.

Belfield, Chris, Jack Britton, Franz Buscha, Lorraine Dearden, Matt Dickson, Laura Van der Erve, Luke Sibieta, Anna Vignoles, Ian Walker, and Yu Zhu, "The Impact of Undergraduate Degrees on Early-Career Earnings," *Institute for Fiscal Studies Research Report*, 2019.

Blundell, Richard, Lorraine Dearden, Alissa Goodman, and Howard Reed, "The Returns to Higher Education in Britain: Evidence from a British Cohort," *Economic Journal*, 2000, *110* (461), pp.F82–99.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik, "Robust Data-Driven Inference in the Regression Discontinuity Design," *Stata Journal*, 2014.

_ , _ , and _ , "Robust Nonparametric Confidence Intervals for Regression Discontinuity Designs," *Econometrica*, 2014, *82* (6), pp.2295–2326.

_ , _ , Max H. Farrell, and Rocío Titiunik, "Regression Discontinuity Designs Using Covariates," *Review of Economics and Statistics*, 2019, *101* (3), pp.442–451.

Canaan, Serena and Pierre Mouganie, "Returns to Education Quality for Low-Skilled Students: Evidence from a Discontinuity," *Journal of Labor Economics*, 2018, *36* (2), pp.395–436.

Cataneo, Matias, Nicolas Idrobo, and Rocio Titiunik, *Elements in Quantitative and Computational Methods for the Social Sciences*, Cambridge University Press, 2019.

Chade, Hector and Lones Smith, "Simultaneous Search," *Econometrica*, 2006, *74* (5), pp.1293–1307.

Clark, Damon and Paco Martorell, "The Signalling Value of a High School Diploma," *Journal of Political Economy*, 2014, *122* (2), pp.282–318.

Crost, Benjamin, Joseph Felter, and Patrick Johnston, "Aid under Fire: Development Projects and Civil Conflict," *American Economic Review*, June 2014, *104* (6), pp.1833–1856.

Dahl, Gordon B., Katrine V. Løken, and Magne Mogstad, "Peer Effects in Program Participation," *American Economic Review*, July 2014, *104* (7), pp.2049–2074.

Dee, Thomas S., Will Dobbie, Brian A. Jacob, and Jonah Rockoff, "The Causes and Consequences of Test Score Manipulation: Evidence from the New York Regents Examinations," *American Economic Journal: Applied Economics*, July 2019, *11* (3), pp.382–423.

Diamond, Rebecca and Petra Persson, "The Long-Term Consequences of Teacher Discretion in Grading of High Stakes Tests," Working Papers No. 22207, National Bureau of Economic Research 2016.

Dinerstein, Michael, Rigissa Megalokonomou, and Constantine Yannelis, "Human Capital Depreciation," *National Bureau of Economic Research*, 2020, (No. 27925).

Frisancho, Veronica, Kala Krishna, Sergey Lychagin, and Cemile Yavas, "Better Luck Next Time: Learning Through Retaking," *Journal of Economic Behavior & Organization*, 2016, *125*, pp.120–135.

Goodman, Joshua, Gurantz Oded, and Jonathan Smith, "Take Two! SAT Retaking and College Enrollment Gaps," *American Economic Journal: Economic Policy*, 2020, *12* (2), pp.115–158.

_ , Michael Hurwitz, and Jonathan Smith, "Access to 4-Year Public Colleges and Degree Completion," *Journal of Labor Economics*, 2017, *35* (3), pp.829–867.

Goulas, Sofoklis and Rigissa Megalokonomou, "Which Degrees do Students Prefer during Recessions?," *Empirical Economics*, 2019, *56*, pp.2093–2125.

_ , _ , and Yi Zhang, "Does the Girl Next Door Affect Your Academic Outcomes and Career Choices?," Discussion Papers No. 11910, Institute of Labor Economics (IZA) 2018.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw, "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, 2001, *69* (1), pp.201–209.

Imbens, Guido W and Joshua D Angrist, "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 1994, *62* (2), pp.467–475.

Jacob, Brian A. and Lars Lefgren, "The Effect of Grade Retention on High School Completion," *American Economic Journal: Applied Economics*, July 2009, *1* (3), pp.33–58.

Kedagni, Desire, Kala Krishna, Rigissa Megalokonomou, and Yingyan Zhao, "Does class size matter? How, and at what cost?," *European Economic Review*, 2021, *133*, 103664.

Krishna, Kala, Sergey Lychagin, and Veronica Frisancho, "Retaking in High Stakes Exams: Is Less More?," *International Economic Review*, 2018, *59* (2), pp.449–477.

Landaud, Fanny and Eric Maurin, "Aim High and Persevere! Competitive Pressure and Access Gaps in Top Science Graduate Programs," Working Papers, HAL 2020.

Lavy, Victor and Rigissa Megalokonomou, "Persistency in Teachers' Grading Biases and Effect on Longer Term Outcomes: University Admission Exams and Choice of Field of Study," Working Papers No. 26021, National Bureau of Economic Research 2019.

_ and _ , "Long-Term Effects of Teachers: Evidence from a Random Assignment of Teachers to Students," *Mimeo*, 2021.

Machin, Stephen, Sandra McNally, and Jenifer Ruiz-Valenzuela, "Entry through the Narrow Door: The Costs of just Failing High-Stakes Exams," *Journal of Public Economics*, 2020, *190*, pp.104224.

Martorell, Paco, Isaac McFarlin, and Yu Xue, "Does Failing a Placement Exam Discourage Underprepared Students from Going to College?," *Education Finance and Policy*, 2015, *10* (1), pp.46–80.

OECD, "Education at a Glance 2018," Technical report, OECD 2018.

_ , "Education at a Glance 2019," Technical report, OECD 2019.

Pop-Eleches, Cristian and Miguel Urquiola, "Going to a Better School: Effects and Behavioral Responses," *American Economic Review*, June 2013, *103* (4), pp.1289–1324.

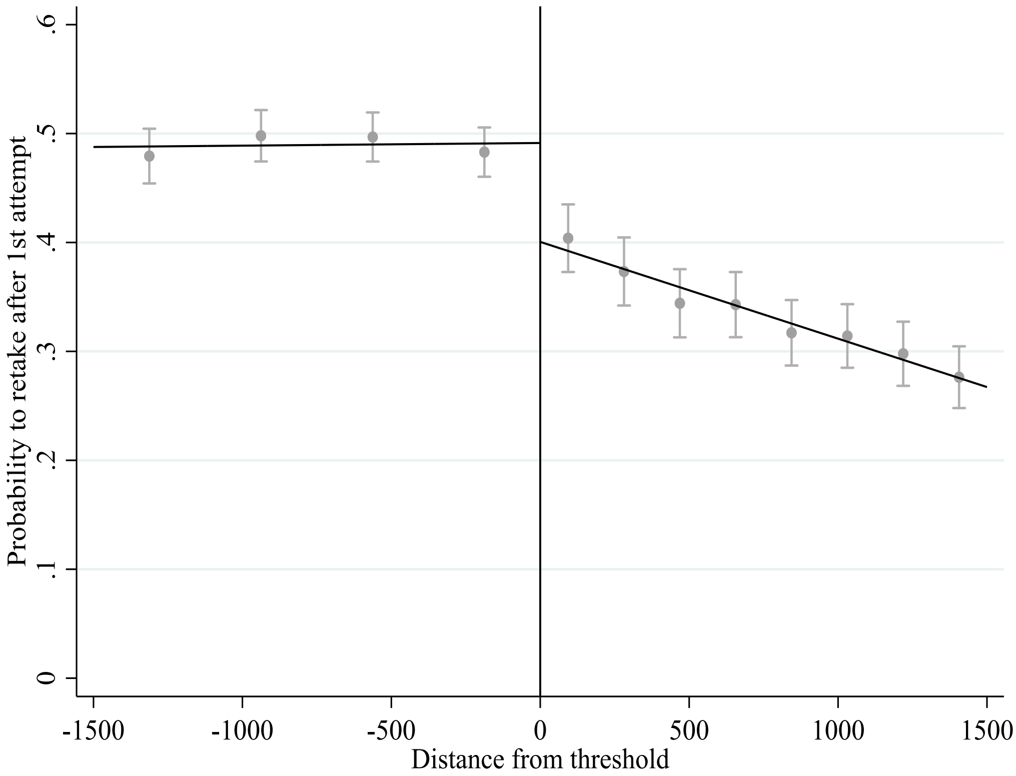
Proud, Steven, "Resits in higher education: merely a bar to jump over, or do they give a pedagogical 'leg up'?," *Assessment & Evaluation in Higher Education*, 2015, *40* (5), pp.681–697.

Scott-Clayton, Judith and Basit Zafar, "Financial Aid, Debt Management, and Socioeconomic Outcomes: Post-college Effects of Merit-based Aid," *Journal of Public Economics*, 2019, *170*, pp.68–82.

Vigdor, Jacob L and Charles T Clotfelter, "Retaking the SAT," *Journal of Human Resources*, 2003, *38* (1), pp.1–33.

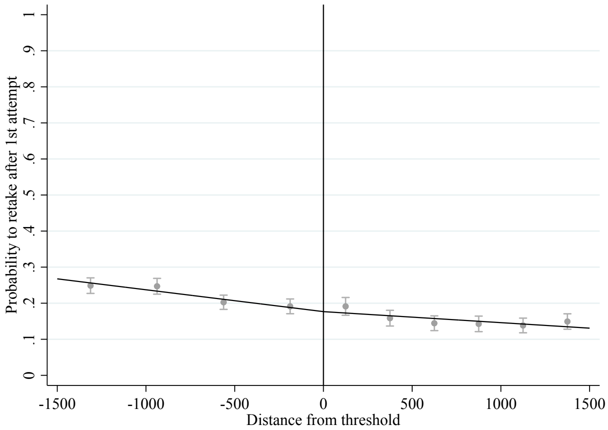
Zimmerman, Seth D., "The Returns to College Admission for Academically Marginal Students," *Journal of Labor Economics*, 2014, *32* (4), pp.711–754.

Figure 1: FIRST STAGE: PROBABILITY TO RETAKE IN 2007

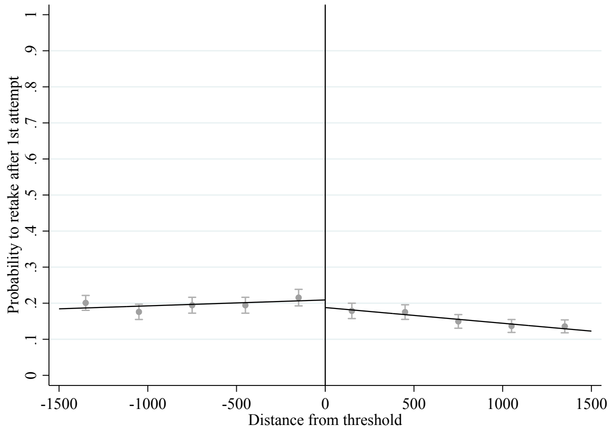


Notes: The figure illustrates the first stage result, i.e. the likelihood of retaking the exam in the subsequent year (2007) after students' 1st attempt for students with admissions scores within +/-1,500 from the threshold of 10,000. The sample is based on 2006 cohort students. We estimate local linear effects using evenly spaced bins, and the lines illustrate the 95% confidence intervals.

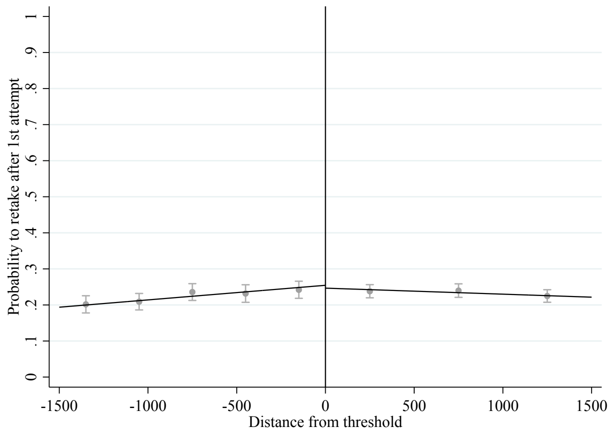
Figure 2: FALSIFICATION EXERCISE FOR FIRST STAGE USING PLACEBO COHORTS INSTEAD OF TREATED COHORT



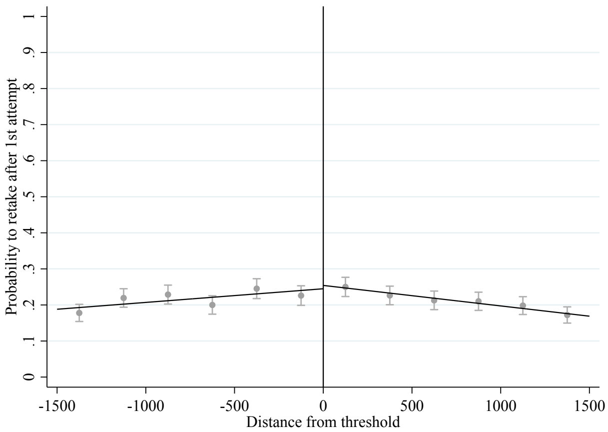
(A) Cohort 2004



(B) Cohort 2005



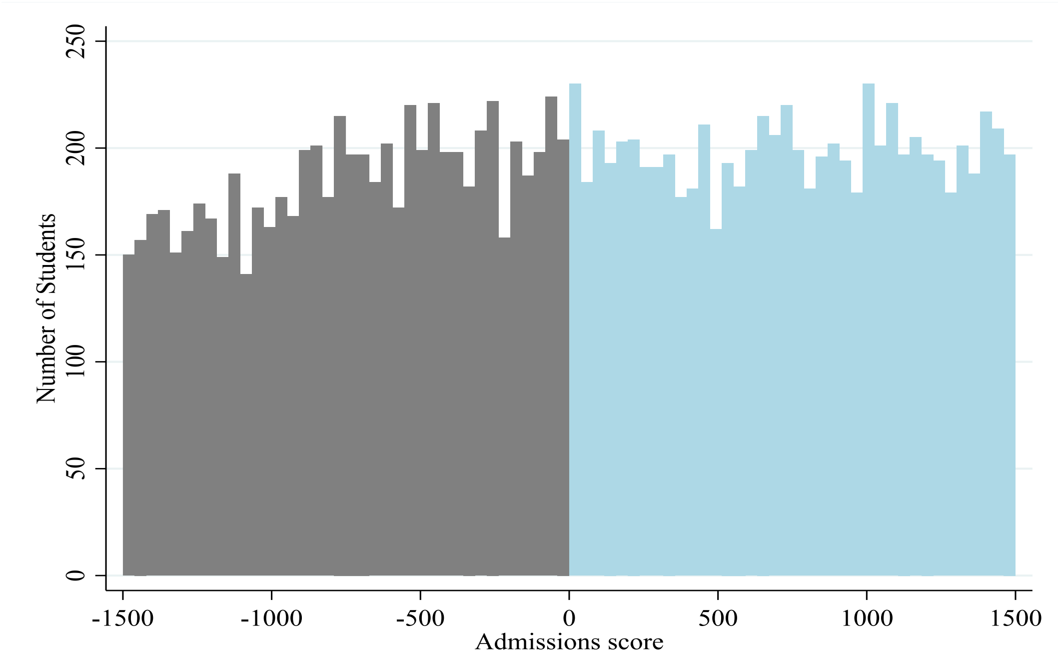
(C) Cohort 2010



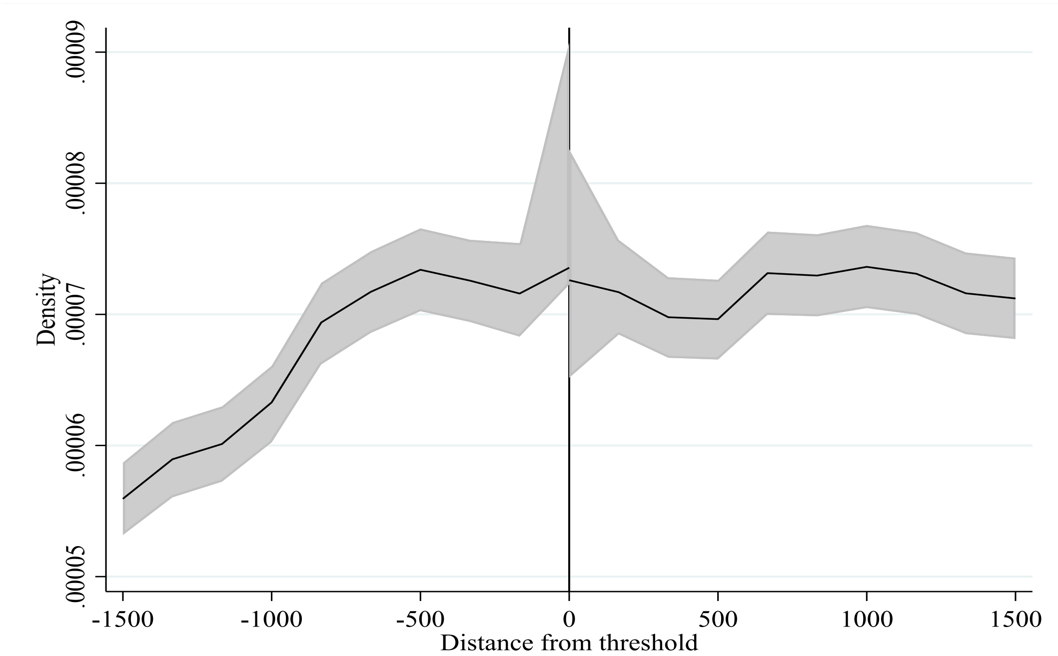
(D) Cohort 2011

Notes: The figure illustrates the likelihood of retaking the exam in the subsequent year for placebo cohorts instead of the treated cohort. Those placebo cohorts are: 2004 and 2005 (the two years before the minimum performance threshold was introduced), and 2010-2011 (the first two years after the minimum performance threshold was lifted). Panel A refers to the 2004 cohort, Panel B to the 2005 cohort, Panel C to the 2010 cohort, and Panel D refers to the 2011 cohort. In those placebo cohorts there is no minimum performance threshold for students to gain access to tertiary education. Each panel refers to a different cohort and the sample is based on the students who obtained scores within +/-1,500 from the threshold of 10,000. We estimate local linear effects using evenly spaced bins, and the lines illustrate the 95% confidence intervals.

Figure 3: HISTOGRAM AND DENSITY OF THE ADMISSIONS SCORE



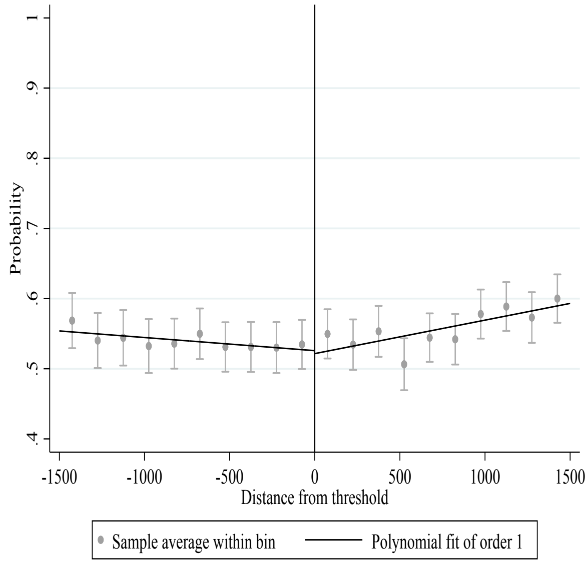
(A) Histogram of the Admissions Score



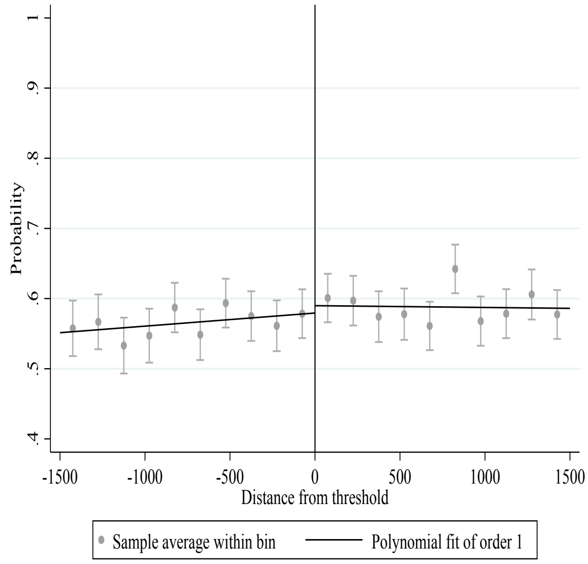
(B) Density of the Admissions Score

Notes: Panel A shows the distribution of admissions score for students with admission scores within +/-1,500 points from the threshold of 10,000. Panel B shows the density of the admissions scores around the 10,000 threshold, for those with scores within +/-1,500 points from the threshold. The density and 95% confidence intervals on each side of the threshold displayed in Panel B are estimated following the approach in Cataneo, Idrobo and Titiunik (2019). We fail to reject the null hypothesis that the density of the running variable is continuous at the threshold of 10,000 scores as proposed by Cataneo, Idrobo and Titiunik (2019) with a p-value equal to 0.241. Overall, there is no statistical evidence of manipulation around the threshold, which provides additional evidence supporting the validity of the RD design.

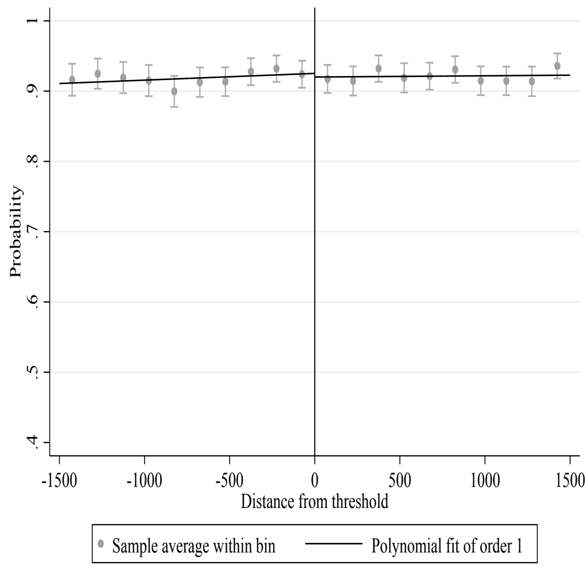
Figure 4: BALANCING TESTS OF PREDETERMINED CHARACTERISTICS



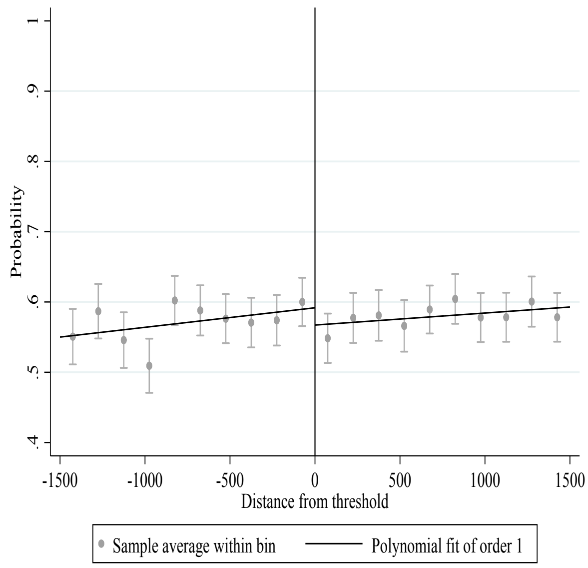
(A) Girl



(B) Attend a Top-Performing High School



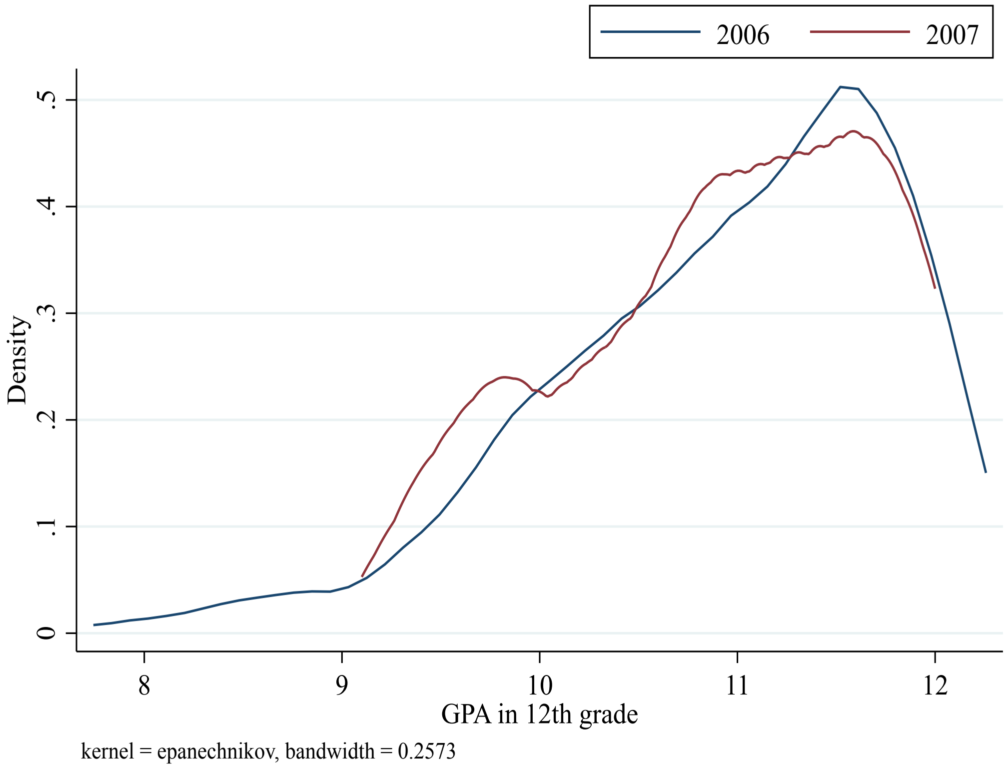
(C) Attend a High School in an Urban Area



(D) Reside in a High-Income Neighborhood

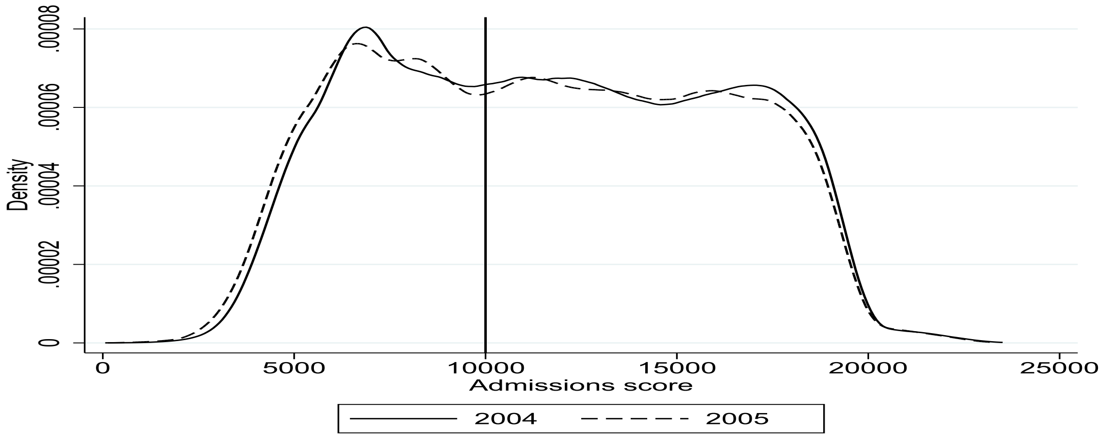
Notes: Each panel shows balancing tests for a different predetermined characteristic, for individuals with admissions scores within +/-1,500 points from the threshold of 10,000. Panel A shows the likelihood that a student is a girl, Panel B shows the likelihood that a student attends a top-performing high school, Panel C shows the likelihood that a student attends an urban high school, and Panel D shows the likelihood that a student resides in a high-income neighborhood. The x-axis shows students' admissions scores as a distance from the admission threshold of 10,000. The lines illustrate the 95% confidence intervals. We estimate local linear effects using evenly spaced bins; the optimal MSE bandwidth is different for each variable and calculated following the approach in Cataneo, Idrobo and Titiunik (2019).

Figure 5: DISTRIBUTION OF HIGH SCHOOL GPA IN THE FIRST AND SECOND YEARS AFTER THE INTRODUCTION OF THE MINIMUM PERFORMANCE THRESHOLD

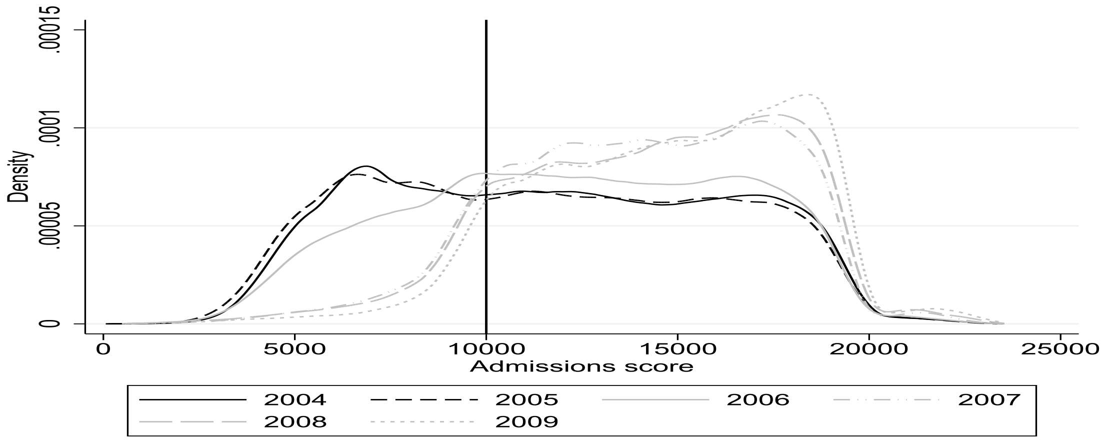


Notes: The figure illustrates the distribution of high school GPA the year the reform was introduced (2006), and the following year (2007). The GPA is the grade the student obtains as the average performance on all first- and second-term examinations in the senior year of high school. The GPA is different from the scores obtained on the entrance exam. All students obtain a GPA at the end of high school. The sample is restricted to students with a GPA below 12/20 and above 8/20. We use the smaller sample of 23 schools to derive these distributions since we have information on students' GPA only in the smaller comprehensive sample that we use for additional analysis.

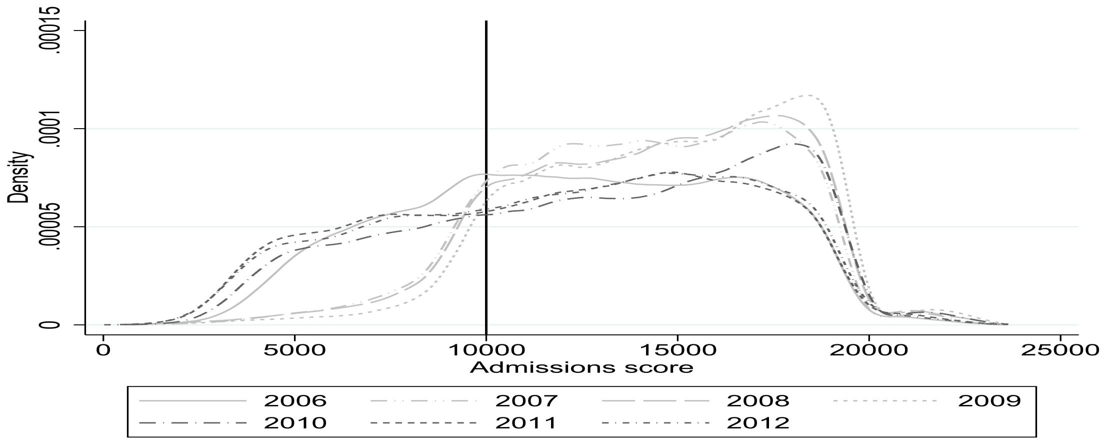
Figure 6: DISTRIBUTION OF ADMISSIONS SCORES BEFORE, DURING, AND AFTER THE REFORM YEARS



(A) Distribution of Admissions Scores Pre-reform



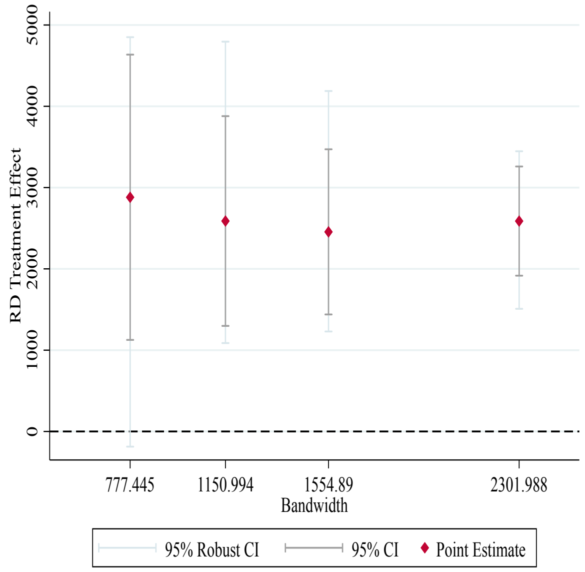
(B) Distribution of Admissions Scores Pre- and During reform



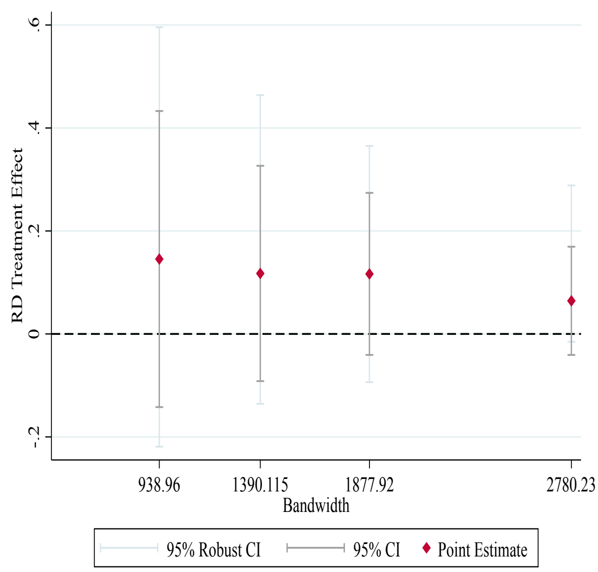
(C) Distribution of Admissions Scores During and Post-reform

Notes: The figure shows the distribution of admissions scores of applicants to postsecondary education in Greece between 2004 and 2012. Panel A shows the distribution of admissions scores before the minimum performance policy was introduced (2004 and 2005). Panel B shows the distribution of admissions scores for the years that the minimum performance policy was in place (2006-2009). Panel C shows the distribution of admissions scores for the years the minimum performance policy was in place (2006-2009) and after it was lifted (2010-2012).

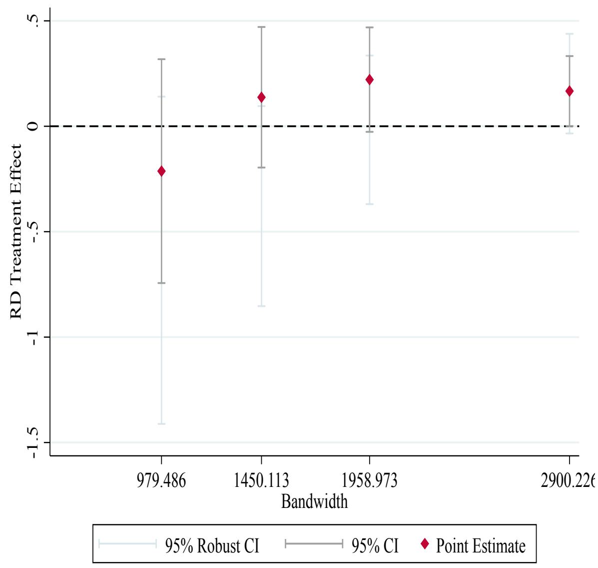
Figure 7: ROBUSTNESS EXERCISE: ALTERNATIVE BANDWIDTHS



(A) Academic Performance



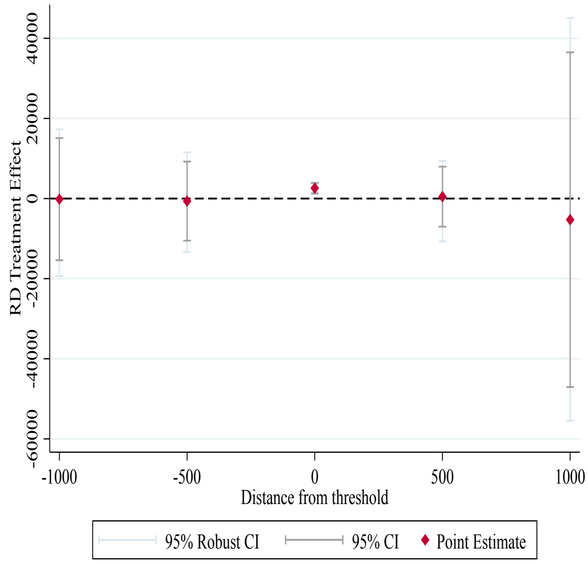
(B) Offer From Top 50% College-course



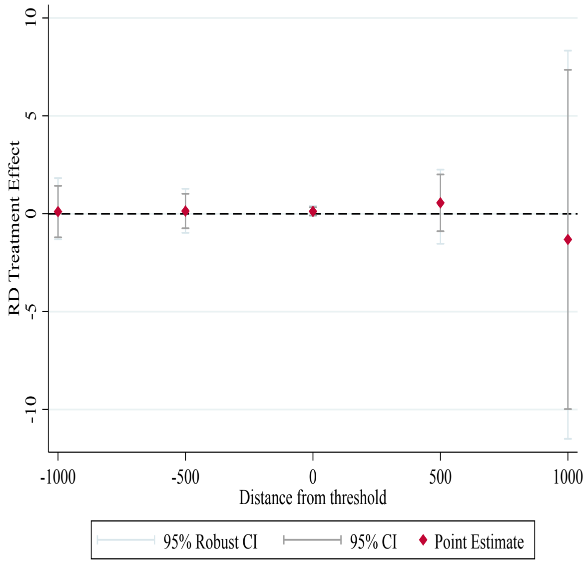
(C) Offer From Academic College-course

Notes: The figure shows the estimates of retaking the exam on a set of outcomes. These outcomes are: academic performance (Panel A), the likelihood to obtain an offer from a college-course that is above the median quality (Panel B), and the likelihood of obtaining an academic offer compared with an offer from a vocational school (or no offer) (Panel C). We show estimated effects using four different bandwidths in each panel, as suggested by Cataneo, Idrobo and Titiunik (2019). These bandwidths are: the MSE optimal bandwidth (the one we use in the main specification as a benchmark), the double of the MSE optimal bandwidth, the CER optimal bandwidth and the double of the CER optimal bandwidth. In Panel A, the MSE and CER optimal bandwidths are 1,150.994, and 777.445, respectively. In Panel B, the MSE and CER optimal bandwidths are 1,390.115, and 938.96, respectively. In Panel C, the MSE and CER optimal bandwidths are 1,450.113, and 979.486, respectively.

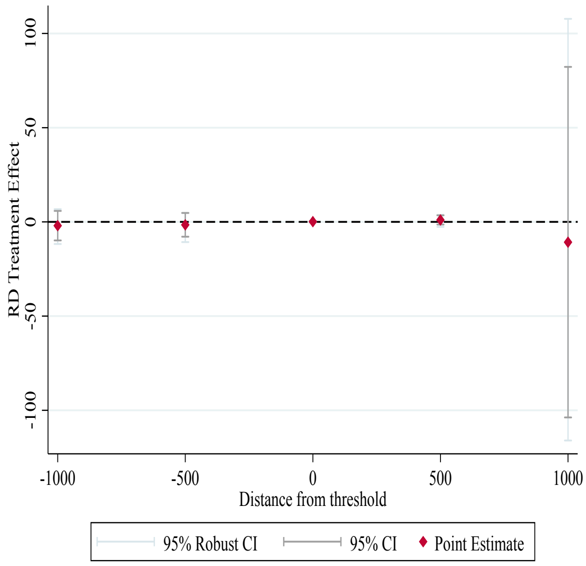
Figure 8: Robustness Exercise: Pseudo-Thresholds



(A) Academic Performance



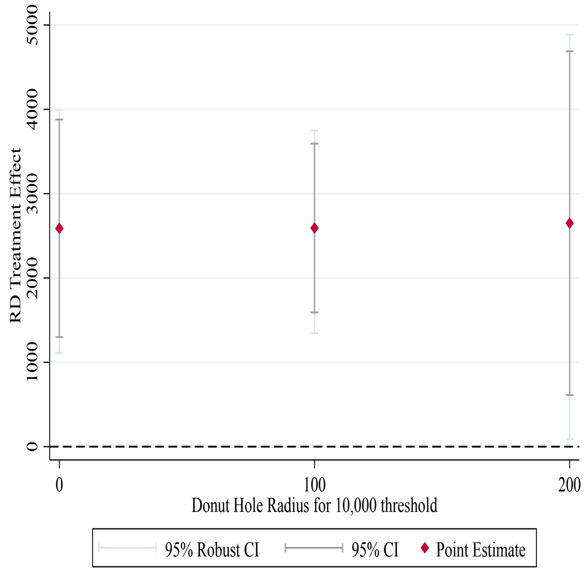
(B) Offer From Top 50% College-course



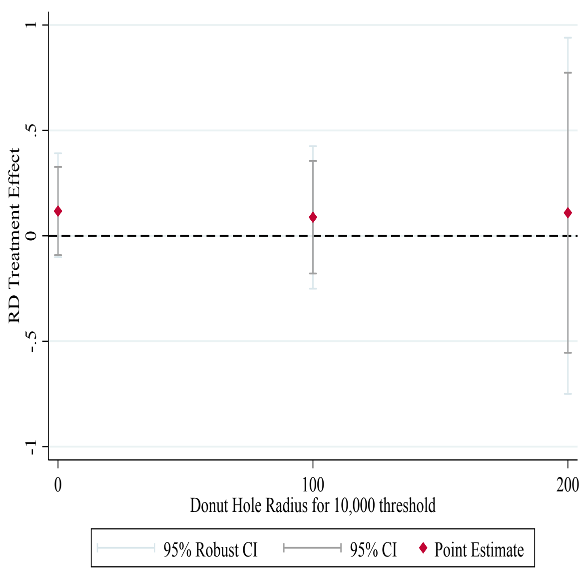
(C) Offer From Academic College-course

Notes: This figure shows the impact of retaking the exam on three different outcomes: academic performance (Panel A), the likelihood to obtain an offer from college-course that is above the median quality (Panel B), and the likelihood of obtaining an academic offer compared to vocational offer (Panel C). The estimated effects on those outcomes are presented for various fake thresholds. The horizontal axis measures the difference from the true threshold (10,000), with the main estimate presented at 0. For each estimation, the optimal bandwidth is calculated using a local linear polynomial following Cataneo, Idrobo and Titiunik (2019).

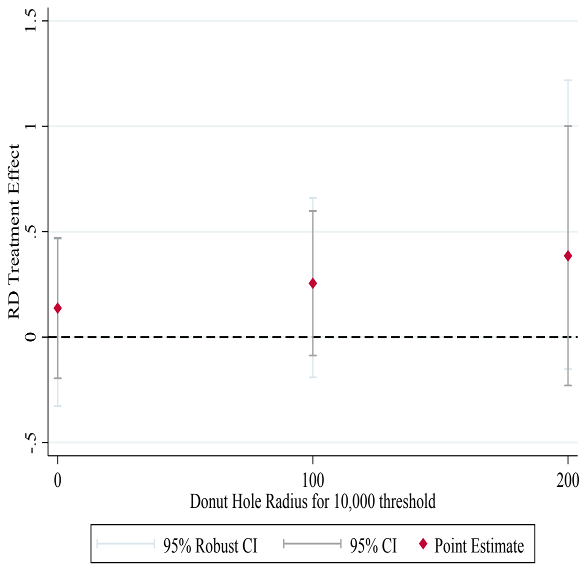
Figure 9: Robustness Exercise: Eliminating Observations Around the Threshold



(A) Academic Performance



(B) Offer From Top 50% College-course



(C) Offer From Academic College-course

Notes: The figure shows the impact of retaking the exam on three different outcomes, when eliminating observations around the threshold. The outcomes are: academic performance (Panel A), the likelihood to obtain an offer from college-course that is above the median quality (Panel B), and the likelihood of obtaining an academic offer compared to vocational offer (Panel C). The horizontal axis measures the donut hole radius, with the main estimate represented at 0. For each estimation the optimal bandwidth is calculated using a local linear polynomial following Cataneo, Idrobo and Titiunik (2019) and as reported in Table A10 in Appendix A.

Table 1: SUMMARY STATISTICS

	Entire Sample (1)	Sample within +/- 1,500 Bandwidth (2)
Admissions Score	12,343.965	10,044.902
	(4,275.356)	(850.357)
Offer Statistics		
No Offers in 2006	0.336	0.521
	(0.472)	(0.500)
Offer From College-course Above Median	0.423	0.063
	(0.494)	(0.242)
Offer From Academic Institution	0.514	0.200
	(0.500)	(0.400)
Retake Statistics		
Retake	0.209	0.427
	(0.407)	(0.495)
Retake in 2007	0.193	0.409
	(0.394)	(0.492)
Retake Once	0.196	0.403
	(0.397)	(0.490)
Retake $\mid$ Fail=0	0.116	0.248
	(0.320)	(0.432)
Retake $\mid$ Fail=1	0.344	0.557
	(0.475)	(0.497)
Application Statistics		
No. of Choices in 2006	26.327	31.514
	(23.096)	(25.778)
$\geq 50\%$ Choices Academic	0.566	0.242
	(0.496)	(0.428)
$\geq 50\%$ Choices Top College-course	0.491	0.120
	(0.500)	(0.325)
High School Statistics		
Urban	0.931	0.920
	(0.253)	(0.271)
Top-performing High School (in 2005)	0.624	0.577
	(0.484)	(0.494)
Demographics		
Average Annual Household Income (euro)	21,962.711	21,025.748
	(10,027.646)	(7,851.361)
Girl	0.561	0.549
	(0.496)	(0.498)
Observations	69,345	14,558

Notes: The sample consists of the student cohort of 2006. The application statistics are from the year 2006, i.e., the first time cohort 2006 takes the exam. Column (1) includes the entire sample of students who take the postsecondary admission exam for the first time in 2006. Column (2) restricts the sample to students who obtained an admissions score within a +/- 1,500 bandwidth around the 10,000 performance cutoff. Application statistics are reported based on the application forms of students' first attempt. The top-performing high schools are the ones with average scores above the median in 2005. Household income is measured in euro and is aggregated at the school postcode level. Standard deviations in parentheses.

Table 2: CHARACTERISTICS OF RETAKERS AROUND THE THRESHOLD

	Compliers	All Retakers	Above Threshold Retakers
	(1)	(2)	(3)
High Income	0.551	0.559	0.571
	(0.497)	(0.497)	(0.495)
Urban High School	0.917	0.919	0.921
	(0.277)	(0.274)	(0.269)
Girl	0.548	0.571	0.603
	(0.498)	(0.495)	(0.489)
Top-performing High School (in 2005)	0.565	0.571	0.579
	(0.496)	(0.495)	(0.494)
Private High School	0.031	0.033	0.035
	(0.174)	(0.179)	(0.185)
Observations	3,444	5,953	2,509

Notes: The table reports the means and standard deviations (in parentheses) of all pre-determined variables conditional on retaking within a +/-1,500 bandwidth around the threshold. Column (1) shows the mean characteristics of those who retake the exam and scored below the threshold in their first attempt (compliers), column (2) shows the mean characteristics of all retakers and column (3) shows the mean characteristics of those students who retake the exam and scored above the threshold in their first attempt.

Table 3: EFFECT OF RETAKING ON ACADEMIC PERFORMANCE

	2SLS					CCT	
	+/-1,000	+/-1,500	+/-2,000	Optimal	Optimal	Optimal	Optimal
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Estimates from the First Stage							
	0.081***	0.091***	0.117***	0.084***	0.079***	0.074***	0.074***
	(0.020)	(0.017)	(0.014)	(0.018)	(0.017)	(0.020)	(0.018)
Panel B: Estimates from the Second Stage							
Retake in 2007	2,336.404***	2,463.050***	2,657.051***	2,282.469***	2,469.717***	2,588.643***	2,461.655***
	(0.628)	(0.646)	(0.652)	(0.610)	(0.522)	(0.690)	(0.521)
	[1,165.105;3,507.702]	[1,642.594;3,283.507]	[2,105.925;3,208.177]	[1,201.463;3,363.475]	[1,501.841;3,437.593]	[1,112.68;3,993.072]	[1,149.744;3,776.778]
Observations	9,948	14,558	18,932	11,378	28,082	11,378	28,082
Kleinbergen-Paap Wald F-stat	16.970	30.110	65.380	19.070	22.250		
Pre-policy outcome mean	10,008.759	10,018.128	10,017.101	10,020.025	10,018.085	10,017.163	10,008.090
Linear polynomial	✓	✓	✓	✓		✓	
Quadratic polynomial					✓		✓

Notes: The table presents the estimated effects of retaking the exam on performance at the entrance exam. Columns (1) to (5) show estimates from a 2SLS estimation, columns (6) and (7) use the robust nonparametric approach of Calonico, Cattaneo and Titiunik (2014a). The bandwidth for the linear polynomial of the running variable is 1,150.99 (columns (4) and (6)) and for the quadratic polynomial of the running variable it is 3,043.9 (columns (5) and (7)). The pre-policy outcome mean refers to cohort 2004. We do not include individuals who retake more than once in this table, since the majority of students retake only once. Panel A shows the estimates from the first stage for different bandwidths, with the standard errors in round brackets. Panel B shows estimates from the second stage; the estimates in standard deviations reported in round brackets; and the standard errors are clustered at school level in squared parentheses and 95% robust confidence intervals are in square brackets. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 4: EFFECT OF RETAKING ON QUALITY OF OFFERS

	2SLS					CCT	
	+/-1,000	+/-1,500	+/-2,000	Optimal	Optimal	Optimal	Optimal
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Offer from Top College-course							
Retake in 2007	0.114	0.123	0.101*	0.110	0.154	0.117	0.140
	[-0.101;0.328]	[-0.037;0.282]	[-0.008;0.209]	[-0.055;0.275]	[-0.174;0.483]	[-0.102;0.391]	[-0.251;0.501]
Observations	9,948	14,558	18,932	13,562	17,258	13,562	17,258
Kleinbergen-Paap Wald F-stat	16.970	30.110	65.380	26.730	6.970		
Panel B: Offer From Academic College-course							
Retake in 2007	-0.018	0.273**	0.219**	0.271**	0.335**	0.138	0.234
	[-0.392;0.355]	[0.017;0.529]	[0.042;0.397]	[0.002;0.540]	[0.017;0.653]	[-0.326;0.464]	[-0.225;0.608]
Observations	9,948	14,558	18,932	14,109	27,377	14,109	27,377
Kleinbergen-Paap Wald F-stat	16.970	30.110	65.380	28.440	21.550		
Linear polynomial	✓	✓	✓	✓		✓	
Quadratic polynomial					✓		✓

Notes: The table presents the estimated effects of retaking the exam on a set of outcomes–each panel refers to a different outcome. Each coefficient comes from a different regression. Columns (1) to (5) show estimates from a 2SLS estimation, columns (6) and (7) use the robust nonparametric approach of Calonico, Cattaneo and Titiunik (2014a). In Panel A, the optimal bandwidth for the linear polynomial is 1,390.12 (columns(4) and (6)) and for the quadratic it is 1,807.17 (columns (5) and (7)). In Panel B, the bandwidths are 1,450.11 (columns (4) and (6)) and 2963.65 (columns (5) and (7)). We do not include individuals who retake more than once in this table, since the majority of students retake only once. Standard errors are clustered at school level and 95% robust confidence intervals are in square brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Effect of Retaking on Student Relative Performance

	2SLS					CCT	
	+/-1,000	+/-1,500	+/-2,000	Optimal	Optimal	Optimal	Optimal
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Counterfactual Relative Rank among 2006 Test-Takers							
Retake in 2007	11.656**	11.719***	12.243***	11.619**	13.633***	13.894**	14.212***
	[2.212;21.100]	[5.037;18.400]	[7.739;16.748]	[2.648;20.590]	[5.618;21.648]	[2.673;26.237]	[3.794;25.326]
Observations	9,948	14,558	18,932	11,090	27,282	11,090	27,282
Kleinbergen-Paap Wald F-stat	16.970	30.110	65.380	17.880	20.760		
Panel B: Relative Rank among 2007 Test-Takers							
Retake in 2007	0.176	0.437	0.839	-0.248	3.692	2.903	3.648
	[-10.320;10.671]	[-6.993;7.868]	[-4.177;5.854]	[-10.259;9.763]	[-4.785;12.168]	[-9.469;16.733]	[-7.245;15.514]
Observations	9,948	14,558	18,932	11,034	28,414	11,034	28,414
Kleinbergen-Paap Wald F-stat	16.970	30.110	65.380	18.080	22.290		
Panel C: Counterfactual Relative Rank among 2011 Test-Takers							
Retake in 2007	15.924***	16.206***	16.769***	15.770***	18.353***	18.189***	18.308***
	[7.039;24.809]	[9.930;22.483]	[12.549;20.989]	[7.374;24.167]	[10.785;25.920]	[7.628;29.721]	[8.261;28.990]
Observations	9,948	14,558	18,932	11,124	26,603	11,124	26,603
Kleinbergen-Paap Wald F-stat	16.970	30.110	65.380	18.120	20.290		
Linear polynomial	✓	✓	✓	✓		✓	
Quadratic polynomial					✓		✓

Notes: The table presents the estimated effects of retaking the exam on a set of outcomes – each panel refers to a different outcome. Each coefficient comes from a different regression. Columns (1) to (5) show estimates from a 2SLS estimation, columns (6) and (7) use the robust non-parametric approach of Calonico, Cattaneo and Titiunik (2014a). Columns (1)-(3) respectively use bandwidths of 1,000, 1,500 and 2,000 marks away from the cutoff. In Panel A, the optimal bandwidth for the linear polynomial is 1,150.99 (columns(4) and (6)) and for the quadratic polynomial it is 3,043.9 (columns (5) and (7)). In Panel B, the bandwidths are 1,119.39 (columns (4) and (6)) and 2,954.94 (columns (5) and (7)). In Panel C, the bandwidths are 1,114.41 (columns (4) and (6)) and 3,081.72 (columns (5) and (7)). We do not include individuals who retake more than once in this table, since the majority of students retake only once. Standard errors are clustered at school level and 95% robust confidence intervals are in square brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Effect of Retaking on Academic Performance and Quality of Offers during Years Reform is in Place

	2006		2007		2008		2009	
	2SLS (1)	CCT (2)	2SLS (3)	CCT (4)	2SLS (5)	CCT (6)	2SLS (7)	CCT (8)
Panel A: Estimates of First Stage								
	0.082***	0.074***	0.074***	0.085***	0.055***	0.067***	0.056***	0.064***
	(0.019)	(0.021)	(0.019)	(0.020)	(0.016)	(0.017)	(0.016)	(0.017)
Panel B: Academic Performance								
Retake in the following year	2,282.469***	2,588.643***	1,924.189***	2,191.586***	2,769.120***	2,485.703***	1,455.153**	1,301.470**
	[1,201.463;3,363.475]	[1,112.68;3,993.072]	[954.299;2,894.079]	1,254.109;3,476.808]	[1,662.256;3875.985]	[1,177.5;3550.408]	[168.098;2,742.208]	[-53.203;2,752.619]
Observations	11,378	11,378	9,397	9,397	12,002	12,002	12,326	12,326
Kleinbergen-Paap Wald F-stat	19.070		14.890		12.550		12.940	
Panel C: Offer from Top College-course								
Retake in the following year	0.110	0.117	0.207**	0.212**	0.163	0.183	-0.096	-0.061
	[-0.055;0.275]	[-0.102;0.391]	[0.014;0.400]	[0.011;0.455]	[-0.084;0.410]	[-0.131;0.414]	[-0.302;0.110]	[-0.243;0.171]
Observations	13,562	13,562	9,716	9,716	7,678	7,678	7,064	7,064
Kleinbergen-Paap Wald F-Stat	26.730		14.840		17.430		13.280	
Panel D: Offer from Academic College-course								
Retake	0.271**	0.138	0.520**	0.571***	0.671***	0.512**	0.950***	0.309
	[0.002;0.540]	[-0.326;0.464]	[0.119;0.921]	[0.156;1.169]	[0.189;1.153]	[-0.014;0.860]	[0.273;1.627]	[-0.250;0.739]
Observations	14,109	14,109	8,955	8,955	8,025	8,025	10,451	10,451
Kleinbergen-Paap Wald F-stat	28.440		15.200		14.640		11.580	
Linear Polynomial	✓	✓	✓	✓	✓	✓	✓	✓

Notes: The table presents the estimated effects of retaking the exam on three outcomes–each panel B-E refers to a different outcome. Panel A shows the first stage for the outcome of Panel B. The first stage for the remaining outcomes is similar in magnitude. Each coefficient comes from a different regression. Odd columns show results from a 2SLS estimation, while even columns use the robust nonparametric approach of Calonico, Cattaneo and Titiunik (2014a). Columns (1)-(2) use the data for cohort 2006, columns (3)-(4) use the data for cohort 2007, columns (5)-(6) use the data for cohort 2008, and columns (7)-(8) use the data for cohort 2009. In Panel B, the optimal bandwidth is 1,150.99 (columns (1) and (2)), 1,399.47 (columns (3) and (4)), 1,903.26 (columns (5) and (6)), and 2,321.56 (columns (7) and (8)). In Panel C, the optimal bandwidth is 1,390.11 (columns (1) and (2)), 1,457.45 (columns (3) and (4)), 1,104.41 (columns (5) and (6)), and 1,202.59 (columns (7) and (8)). In Panel D, the optimal bandwidth is 1,450.11 (columns (1) and (2)), 1,318.36 (columns (3) and (4)), 1,162.15 (columns (5) and (6)), and 1,909.99 (columns (7) and (8)). We do not include individuals who retake more than once in this table, since the majority of students retake only once. Standard errors are clustered at school level and 95% robust confidence intervals are in square brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table 7: EFFECT OF RETAKING ON APPLICATION CHOICES

	2SLS					CCT	
	+/-1,000	+/-1,500	+/-2,000	Optimal	Optimal	Optimal	Optimal
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Number of Choices							
Retake in 2007	0.236	-2.755	2.048	1.870	-5.757	3.991	-4.342
	[-22.648;23.119]	[-20.256;14.745]	[-9.934;14.029]	[-18.757;22.497]	[-25.020;13.507]	[-23.789;33.503]	[-29.825;20.983]
Observations	9,948	14,558	18,932	12,173	29,603	12,173	29,603
Kleinbergen-Paap Wald F-stat	16.970	30.110	65.380	21.380	24.050		
Panel B: At Least 50% Top College-course Choices							
Retake in 2007	-0.125	0.026	0.002	0.002	0.017	-0.051	-0.003
	[-0.505;0.255]	[-0.236;0.289]	[-0.173;0.176]	[-0.280;0.283]	[-0.347;0.380]	[-0.494;0.348]	[-0.464;0.518]
Observations	9,948	14,558	18,932	13,374	22,914	13,374	22,914
Kleinbergen-Paap Wald F-stat	16.970	30.110	65.380	25.960	15.640		
Panel C: At Least 50% of College-course Choices Academic							
Retake in 2007	-0.224	-0.076	-0.003	-0.051	0.072	-0.125	-0.051
	[-0.698;0.249]	[-0.405;0.253]	[-0.217;0.211]	[-0.396;0.295]	[-0.305;0.449]	[-0.713;0.329]	[-0.597;0.469]
Observations	9,948	14,558	18,932	13,687	27,532	13,687	27,532
Kleinbergen-Paap Wald F-stat 16.970	30.110	65.380	26.390	22.060			
Linear polynomial	✓	✓	✓	✓		✓	
Quadratic polynomial					✓		✓

Notes: The table presents the estimated effects of retaking the exam on application choices – each panel refers to a different outcome. Each coefficient comes from a different regression. Columns (1) to (5) show results from a 2SLS estimation, while columns (6) and (7) use the robust nonparametric approach of Calonico, Cattaneo and Titiunik (2014a). Columns (1)-(3) use bandwidths of 1,000, 1,500 and 2,000 marks away from the cutoff, respectively. In Panel A, the optimal bandwidth for the linear polynomial is 1,237.03 (columns (4) and (6)) and 3,171.57 (columns (5) and (7)). In Panel B, the bandwidths are 1,513.86 (columns (4) and (6)) and 3,043.87 (columns (5) and (7)). In Panel C, the bandwidths are 1,193.64 (columns (4) and (6)) and 3,268.14 (columns (5) and (7)). We do not include individuals who retake more than once in this table, since the majority of students retake only once. Standard errors are clustered at school level and 95% robust confidence intervals are in square brackets. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

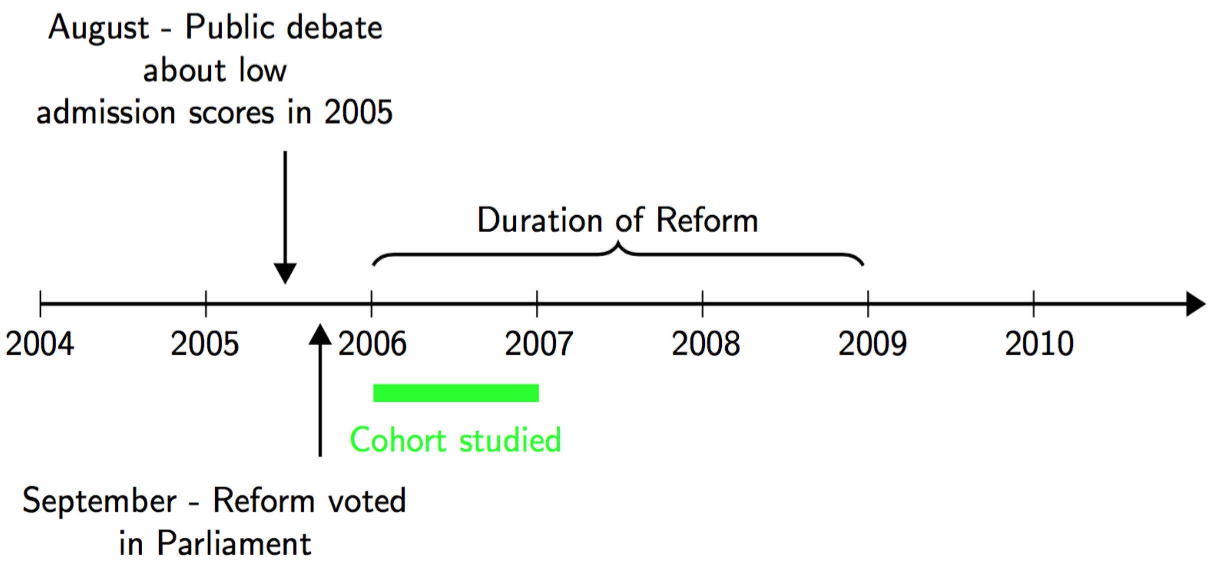
Table 8: ROBUSTNESS EXERCISE: EFFECT OF RETAKING ON ACADEMIC PERFORMANCE USING ALTERNATIVE ADMISSION SCORE

	+/-1,000 (1)	+/-1,500 (2)	+/-2,000 (3)	Optimal (4)	Optimal (5)
Panel A: Academic Performance					
Retake in 2007	2,662.796***	2,815.755***	2,987.511***	2,652.284***	2,781.568***
	[1,693.102;3,632.490]	[2,176.226;3,455.284]	[2,540.603;3,434.419]	[1,773.704;3,530.864]	[1,702.635;3,860.500]
Observations	8,729	13,066	17,253	9,964	9,964
Kleinbergen-Paap Wald F-stat	26.110	52.000	101.220	28.500	.
Panel B: Offer from Top College-course					
Retake in 2007	0.096	0.112*	0.097**	0.080	0.100
	[-0.071;0.263]	[-0.005;0.229]	[0.013;0.182]	[-0.085;0.245]	[-0.103;0.304]
Observations	8,897	13,311	17,572	9,339	9,339
Kleinbergen-Paap Wald F-stat	26.200	51.160	100.720	26.830	.
Panel C: Offer from Academic College-course					
Retake in 2007	-0.091	0.103	0.068	0.075	0.060
	[-0.401;0.219]	[-0.094;0.300]	[-0.076;0.212]	[-0.081;0.232]	[-0.146;0.266]
Observations	8,897	13,311	17,572	16,342	16,342
Kleinbergen-Paap Wald F-stat	26.200	51.160	100.720	85.670	.
Optimal Bandwidth Linear 2SLS	✓	✓	✓	✓	
Optimal Bandwidth Linear CCT					✓

Notes: The table presents the estimated effects of retaking the exam on 3 outcomes - each panel refers to a different outcome. For all students, regardless of whether they have an offer or not, we use the highest of the two possible admissions scores they obtain. Columns (1)-(4) show estimates from a 2SLS estimation and column (5) uses the robust nonparametric approach of Calonico, Cattaneo and Titiunik (2014a). A linear polynomial of the running variable is used in all estimations. The bandwidth for columns (1), (2) and (3) are +/-1,000, +/-1,500, +/-2,000, as indicated in the top row. Columns (4) and (5) use the optimal bandwidth from the CCT estimation, in Panel A it is 1,145.88 in Panel B it is 1,052.37 and in Panel C it is 1,858.37. We do not include individuals who retake more than once in this table, since the majority of students retake only once. Standard errors are clustered at school level and 95% robust confidence intervals are in square brackets. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

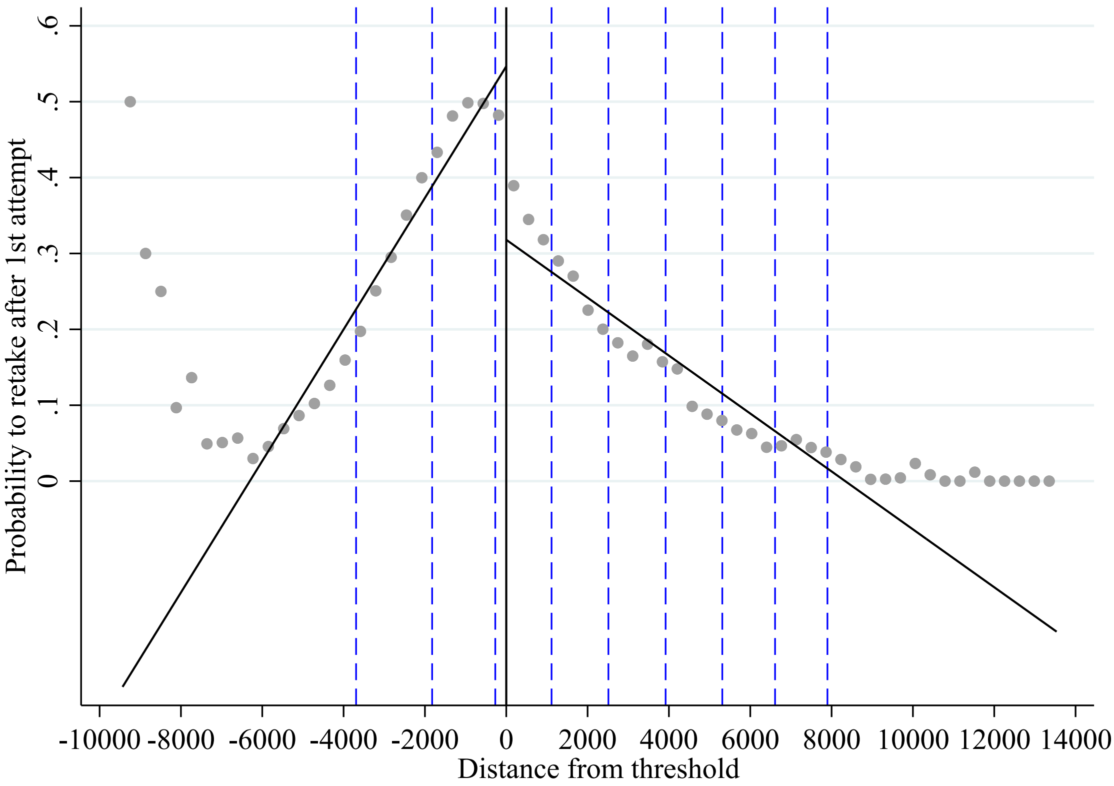
A Appendix

Figure A1: TIMELINE OF THE REFORM



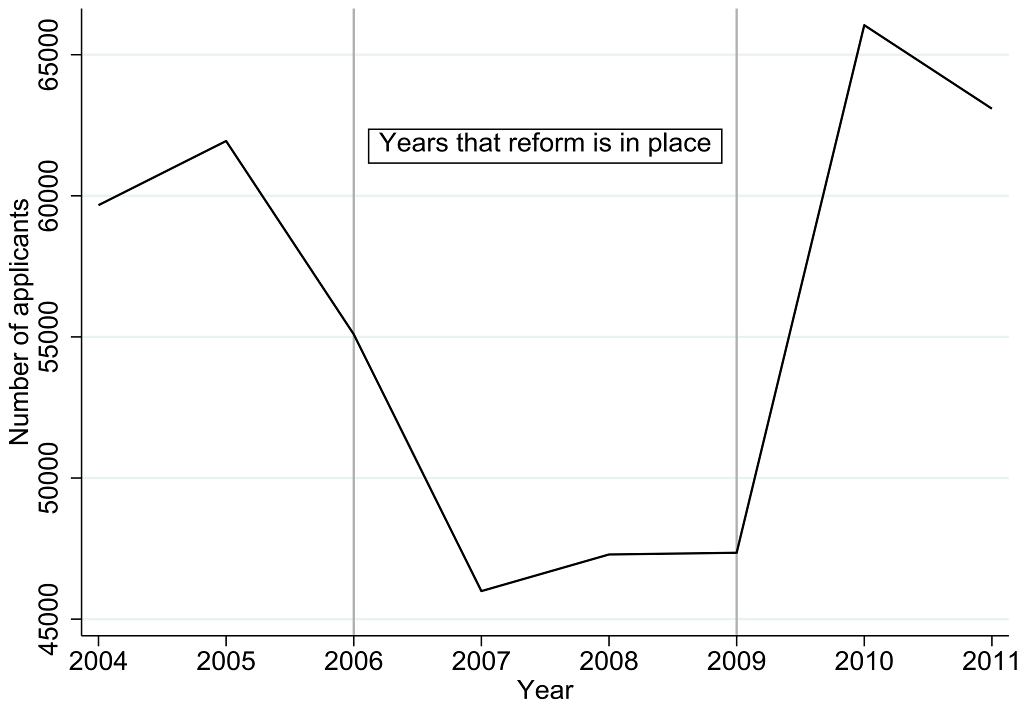
Notes: This figure shows the timing of the reform. In August 2005, there was a public debate and concern that students with very low postsecondary education admissions scores can still get admitted to tertiary education institutions. In September 2005, the reform was voted in Parliament but there was still substantial debate on whether this would be implemented. There were also protests against the implementation of the reform. The first affected cohort is the 2006 cohort, which we use in our main analysis. For this cohort, the reform was unanticipated to a large extent.

Figure A2: FIRST STAGE: PROBABILITY TO RETAKE IN 2007



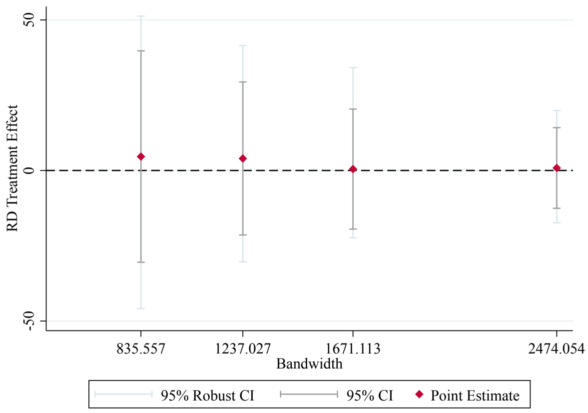
Notes: This figure illustrates how the likelihood of retaking the exam once after the first attempt changes around the score threshold (first-stage estimates). We instrument the probability to retake the exam in 2007 with an indicator of obtaining an admissions score below the 10,000 threshold for all students. Vertical dashed blue lines identify the 10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th, and 90th percentiles of the admissions score.

Figure A3: NUMBER OF FIRST TIME APPLICANTS OVER TIME

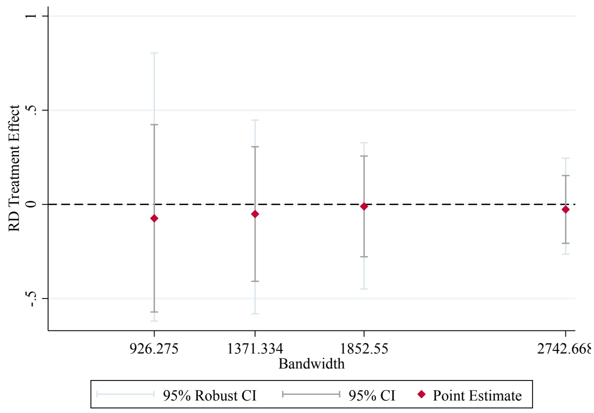


Notes: The figure shows the number of first-time applicants for admission into postsecondary education in Greece from 2005 to 2011. The period within the grey lines indicates the years when the minimum performance threshold for entering postsecondary education was in place.

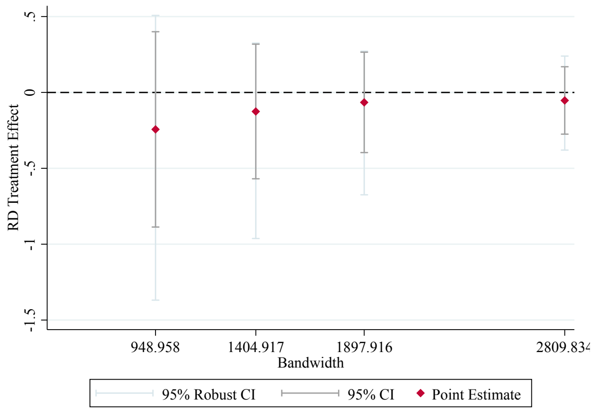
Figure A4: Robustness Exercise: Alternative Bandwidths



(A) Number of Choices



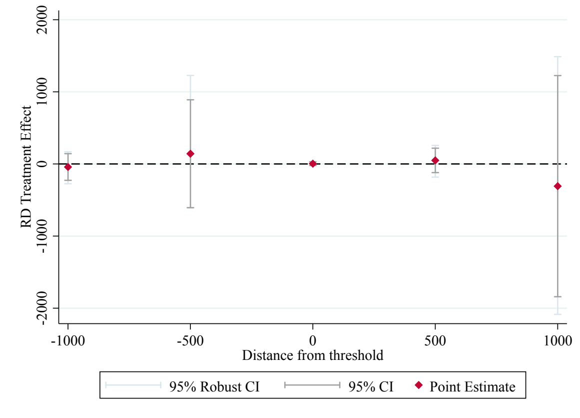
(B) At Least 50% Top College-course Choices



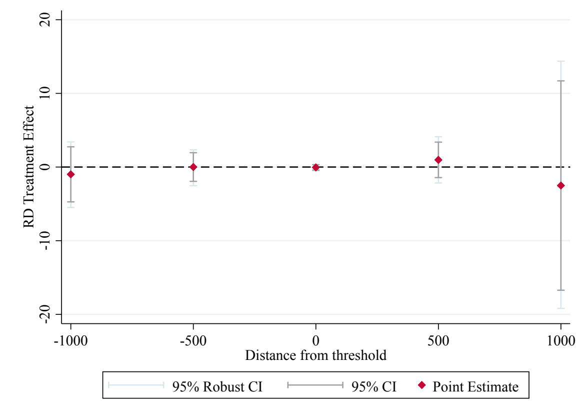
(C) At Least 50% of College-course Choices Academic

Notes: The figure shows the impact of retaking the exam on three outcomes using alternative bandwidths. In Panel A the outcome is number of choices; in Panel B the outcome is the likelihood that at least 50% of the choices are for college-courses ranked above the median; in Panel C the outcome is the likelihood that at least 50% of the choices are for academic college-courses. In Panel A the MSE and CER optimal bandwidths are 1,237.027, and 835.557, respectively. In Panel B the MSE and CER optimal bandwidths are 1,371.334, and 926.275, respectively. In Panel C the MSE and CER optimal bandwidths are 1,404.917, and 948.958 respectively.

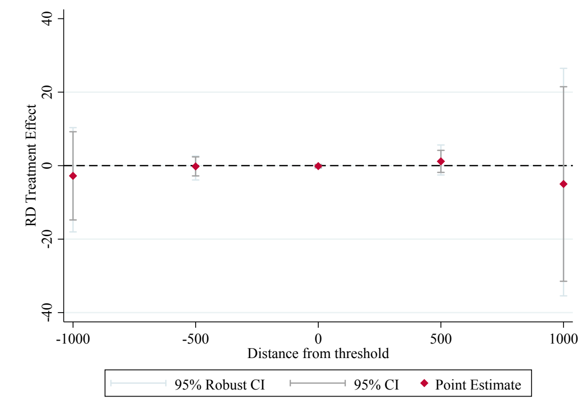
Figure A5: ROBUSTNESS EXERCISE: PSEUDO-THRESHOLDS



(A) Number of Choices



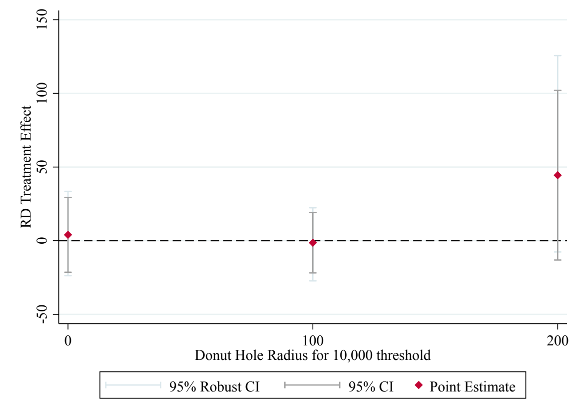
(B) At Least 50% Top College-course Choices



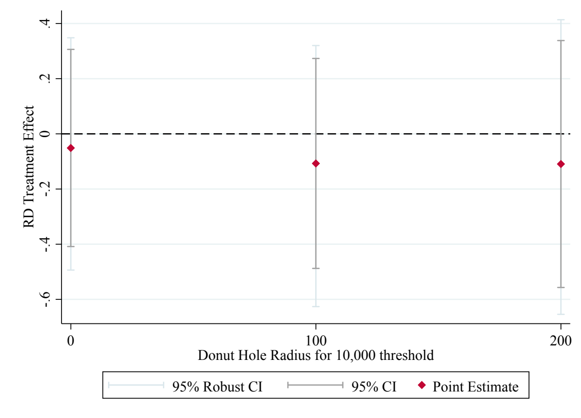
(C) At Least 50% of College-course Choices Academic

Notes: The figure shows the impact of retaking the national exam using alternative thresholds for three outcomes. In Panel A the outcome is number of choices; in Panel B the outcome is the likelihood that at least 50% of the choices are for college-courses ranked above the median in 2005; in Panel C the outcome is the likelihood that at least 50% of the choices are for academic college-courses. The horizontal axis measures the difference from the true threshold (10,000), with the main estimate represented at 0. For each estimation the optimal bandwidth is calculated using a local linear polynomial following Cataneo, Idrobo and Titiunik (2019).

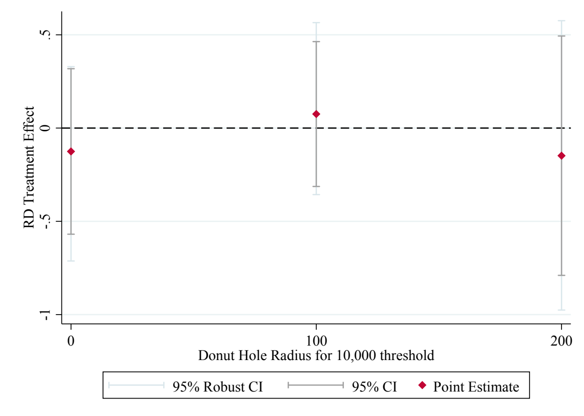
Figure A6: ROBUSTNESS EXERCISE: ELIMINATING OBSERVATIONS AROUND THE THRESHOLD



(A) Number of Choices



(B) At Least 50% Top College-course Choices



(C) At Least 50% of College-course Choices Academic

Notes: The figure shows the impact of retaking the exam when eliminating observations around the threshold for four outcomes. In Panel A the outcome is the number of choices; in Panel B the outcome is the likelihood that at least 50% of the choices are for college-courses ranked above the median; in Panel C the outcome is the likelihood that at least 50% of the choices are for academic college-courses compared to vocational. The horizontal axis measures the donut hole radius, with the main estimate represented at 0. For each estimation the optimal bandwidth is calculated using a local linear polynomial following Cataneo, Idrobo and Titiunik (2019).

Table A1: Effect of Retaking on Quality of Offers: Additional Outcomes

	2SLS					CCT	
	+/-1,000	+/-1,500	+/-2,000	Optimal	Optimal	Optimal	Optimal
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Quality of College-Course Measured as the Threshold of the Admission Score in 2005							
Retake in 2007	-2,611.881	-2,759.224	-3,568.024**	-2,641.719	-1,186.606	-5,175.579	-2,535.994
	[-7,975.196;2,751.435]	[-6,800.152;1,281.704]	[-6,418.503;-717.545]	[-8,142.603;2,859.166]	[-6,113.283;3,740.071]	[-14,146.842;3,489.085]	[-10,786.853;3,898.85]
Observations	9,948	14,558	18,932	10,050	24,340	10,050	24,340
Kleinbergen-Paap Wald F-stat	16.970	30.110	65.380	16.310	16.430		
Panel B: Offer from College-course with High Employment Prospects							
Retake in 2007	0.530**	0.120	-0.039	0.549**	0.352	0.284	0.438
	[0.020;1.040]	[-0.215;0.454]	[-0.269;0.191]	[0.030;1.068]	[-0.111;0.815]	[-0.234;1.017]	[-0.087;1.119]
Observations	9,948	14,558	18,932	9,997	24,926	9,997	24,926
Kleinbergen-Paap Wald F-stat	16.970	30.110	65.380	16.710	17.240		
Linear polynomial	✓	✓	✓	✓		✓	
Quadratic polynomial					✓		✓

Notes: The table presents the estimated effects of retaking the exam on the quality of college-course as measured by the local minimum admissions score threshold in 2005 (Panel A), and the likelihood of obtaining an offer from a college-course with high employment prospects (Panel B). The latter is a dummy equal to 1 if the employment prospect is good and 0 if the employment prospect is mediocre, poor or very poor (see Figure C3 in Appendix B for a full description of the indicator). Each coefficient comes from a different regression. Columns (1) to (5) show estimates from a 2SLS estimation, columns (6) and (7) use the robust nonparametric approach of Calonico, Cattaneo and Titiunik (2014a). The bandwidth in Panel A is 1,011.48 (columns (4) and (6)) and 2,606.31 (columns (5) and (7)). The bandwidth in Panel B is 1,006.86 (columns (4) and (6)) and 2,677.04 (columns (5) and (7)). We do not include individuals who retake more than once in this table, since the majority of students retake only once. Standard errors are clustered at school level and 95% robust confidence intervals are in square brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table A2: EFFECT OF RETAKING ON SUBSEQUENT PERFORMANCE OUTCOMES: BY INCOME

	Academic Performance		Offer from Top College-course		Academic Offer	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: High Income						
Retake in 2007	2,769.337***	3,215.619***	0.054	0.133	0.239*	0.154
	[1,630.035;3,908.639]	[1,720.473;4,962.029]	[-0.186;0.294]	[-0.154;0.501]	[-0.040;0.519]	[-0.279;0.603]
Observations	7,775	7,775	7,019	7,019	9,999	9,999
Kleibergen-Paap Wald F-stat	16.420	.	14.380	.	27.450	.
Panel B: Low Income						
Retake in 2007	1,872.993***	1,358.814	0.207*	0.141	0.301**	0.290
	[520.159;3,225.828]	[-1,260.945;3,210.196]	[-0.016;0.431]	[-0.165;0.536]	[0.008;0.593]	[-0.136;0.787]
Observations	5,946	5,946	6,253	6,253	7,089	7,089
Kleibergen-Paap Wald F-stat	12.450	.	14.450	.	22.370	.
Optimal bandwidth Linear 2SLS	✓		✓		✓	
Optimal bandwidth Linear CCT		✓		✓		✓

Notes: The table presents the estimated effects of retaking the exam on four different outcomes (columns) separately by students from household above the median in terms of income (Panel A) and by students from household below the median (Panel B). Odd columns show estimates using a 2SLS estimation, while in even columns we use the robust non-parametric approach of Calonico, Cattaneo and Titiunik (2014a). We show the impact of retaking the exam on: 1) academic performance in columns (1)-(2), 2) the likelihood the student received an offer from a college-course that is above the median quality in columns (3)-(4), 3) the likelihood the student received an offer from an academic compared to a vocational institution or no offer. Optimal bandwidths are calculated separately for each outcome and for each sub-group (low and high income postcodes). We do not include the results of individuals who retake more than once in this table, the results do not change when we do. Standard errors are clustered at school level and 95% robust CIs are in square brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table A3: EFFECT OF RETAKING ON SUBSEQUENT PERFORMANCE OUTCOMES: BY GENDER

	Academic Performance		Offer from Top College-course		Academic Offer	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Girls						
Retake in 2007	2,693.482***	2,204.675	-0.033	-0.141	0.104	0.036
	[774.937;4,612.026]	[-1,542.353;5,223.381]	[-0.358;0.292]	[-0.789;0.352]	[-0.425;0.633]	[-0.753;0.942]
Observations	7,515	7,515	8,858	8,858	8,612	8,612
Kleibergen-Paap Wald F-stat	5.840	.	8.260	.	7.660	.
Panel B: Boys						
Retake in 2007	2,332.275***	2,599.699***	0.228**	0.257**	0.386***	0.218
	[1,414.245;3,250.304]	[1,433.187;3,830.068]	[0.038;0.417]	[0.027;0.551]	[0.114;0.658]	[-0.195;0.532]
Observations	6,016	6,016	5,670	5,670	6,902	6,902
Kleibergen-Paap Wald F-stat	25.620	.	22.600	.	28.470	.
Optimal bandwidth Linear 2SLS	✓		✓		✓	
Optimal bandwidth Linear CCT		✓		✓		✓

Notes: The table presents the estimated effects of retaking the exam on three different outcomes (shown in columns) separately by girls (Panel A) and boys (Panel B). Odd columns show estimates using a 2SLS estimation, while in even columns we use the robust non-parametric approach of Calonico, Cattaneo and Titiunik (2014a). We show the impact of retaking the exam on: 1) academic performance in columns (1)-(2), 2) the likelihood the student received an offer from a college-course that is above the median quality in columns (3)-(4), and 3) the likelihood the student received an offer from an academic compared to a vocational institution. Optimal bandwidths are calculated separately for each outcome and for each sub-group (boys and girls). We do not include the results of individuals who retake more than once in this table, the results do not change when we do. Standard errors are clustered at school level and 95% robust CIs are in square brackets. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A4: Effect of Retaking on Subsequent Performance Outcomes: by High School Quality

	Academic Performance		Offer from Top College-course		Academic Offer	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: High Quality						
Retake in 2007	2,747.150***	2,774.624***	0.094	0.150	0.322	0.175
	[1,444.146;4,050.153]	[968.812;4,572.550]	[-0.165,0.353]	[-0.233,0.578]	[-0.071,0.715]	[-0.378,0.711]
Observations	7,669	7,669	6,742	6,742	8,849	8,849
Kleibergen-Paap Wald F-stat	12.490	.	11.000	.	12.860	.
Panel B: Low Quality						
Retake in 2007	2,150.289***	2,149.606***	0.107	0.099	0.276*	0.186
	[1,129.373;3,171.206]	[424.044;3,660.989]	[-0.108;0.321]	[-0.168;0.439]	[-0.044;0.596]	[-0.300;0.63]
Observations	6,275	6,275	6,513	6,513	6,733	6,733
Kleibergen-Paap Wald F-stat	19.450	.	21.360	.	20.990	.
Optimal bandwidth Linear 2SLS	✓		✓		✓	
Optimal bandwidth Linear CCT		✓		✓		✓

Notes: The table presents the estimated effects of retaking the exam on three different outcomes (shown in columns) separately by students from high schools above the median in terms of performance (Panel A) and by students from high schools below the median (Panel B). Odd columns show estimates using a 2SLS estimation, while in even columns we use the robust non-parametric approach of Calonico, Cattaneo and Titiunik (2014a). We show the impact of retaking the exam on: 1) academic performance in columns (1)-(2), 2) the likelihood the student received an offer from a college-course that is above the median quality in columns (3)-(4) and 3) the likelihood the student received an offer from an academic compared to a vocational institution or no offer. Optimal bandwidths are calculated separately for each outcome and for each sub-group (high quality schools and low quality schools). We do not include the results of individuals who retake more than once in this table, the results do not change when we do. Standard errors are clustered at school level and 95% robust CIs are in square brackets. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A5: EFFECT OF RETAKING ON SUBSEQUENT PERFORMANCE OUTCOMES: BY TYPE OF SCHOOL

	Academic Performance		Offer From Top College-course		Academic Offer	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Public Schools						
Retake in 2007	2,133.631***	2,341.909***	0.079	0.094	0.224	0.115
	[838.817;3,428.445]	[823.293;3,860.526]	[-0.140;0.298]	[-0.165;0.354]	[-0.048;0.496]	[-0.241;0.470]
Observations	10,798	10,798	11,855	11,855	14,403	14,403
Kleibergen-Paap Wald F-stat	13.630	.	15.840	.	27.000	.
Panel B: Private Schools						
Retake in 2007	3,927.108***	3,845.046***	0.237	0.257	0.179	0.340
	[1,884.743;5,969.473]	[2,064.794;5,625.299]	[-0.158;0.632]	[-0.101;0.615]	[-0.300;0.659]	[-0.104;0.784]
Observations	678	678	702	702	1,138	1,138
Kleibergen-Paap Wald F-stat	7.040	.	8.720	.	11.150	.
Optimal bandwidth Linear 2SLS	✓		✓		✓	
Optimal bandwidth Linear CCT		✓		✓		✓

Notes: The table presents the estimated effects of retaking the exam on four different outcomes (shown in columns) separately for public (Panel A) and private schools (Panel B). Odd columns show estimates using a 2SLS estimation, and even columns show robust nonparametric estimates using Calonico, Cattaneo and Titiunik (2014a). We show the impact of retaking the exam on: 1) academic performance in columns (1)-(2), 2) the likelihood a student received an offer from a college-course that is above the median quality in columns (3)-(4), and 3) the likelihood a student received an offer from an academic compared with a vocational institution or no offer. Optimal bandwidths are calculated separately for each outcome and for each sub-group (public and private schools). We do not include the results of individuals who retake more than once in this table, the results do not change when we do so. Standard errors are clustered at school level and 95% robust confidence intervals are in square brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table A6: EFFECT OF RETAKING ON PERFORMANCE - BY AGE

	2SLS	**CCT**
	(1)	(2)
Panel A: 18-year-old		
Retake	2,510.235***	2,857.931***
	[960.916;4,059.555]	[1,208.345;4,507.516]
Observations	315	1,077
Kleinbergen-Paap Wald F-stat	10.280	.
Panel B: 17-year-old		
Retake	1,935.918*	2,752.977**
	[-108.860;3,980.696]	[294.704;5,211.251]
Observations	331	823
Kleinbergen-Paap Wald F-stat	6.500	.
Linear polynomial	✓	✓

Notes: The table presents the estimated effects of retaking the exam on exam performance by age using our smaller sample of 23 schools. Panel A shows the effects for students who are 18 years old, while Panel B shows the effects for students who are 17 years old. Each coefficient comes from a different regression. Column (1) shows results from a 2SLS estimation and column (2) applies the robust nonparametric (CCT) approach of Calonico, Cattaneo and Titiunik (2014a). In Panel A the optimal bandwidth for the linear polynomial is 3,183.65. and in Panel B is 2,263.41 Optimal bandwidths are calculated following Calonico, Cattaneo and Titiunik (2014a). We do not include individuals who retake more than once in this table, since the majority of students retake only once. Standard errors are clustered at school level and 95% robust confidence intervals are in square brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table A7: EFFECT OF RETAKING ON SUBSEQUENT ACADEMIC PERFORMANCE - RESULTS FROM SMALLER SAMPLE

	(1)	(2)
Retake	2,395.777***	2,863.705***
	[962.484;3,829.069]	[1,213.367;4,514.044]
Observations	312.000	1,077.000
Kleibergen-Paap Wald F-stat	11.170	.
Optimal Bandwidth 2SLS	✓	
Optimal Bandwidth CCT		✓

Notes: The table presents the estimated effects of retaking the exam on academic performance using only schools and students in the smaller sample. The sample used in this table is used to obtain Table A6. The first column shows estimates using a 2SLS estimation, and the second column uses the robust nonparametric approach of Calonico, Cattaneo and Titiunik (2014a). Optimal bandwidths are the same for columns (1) and (2). Standard errors are clustered at school level and 95% robust confidence intervals are in square brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table A8: EFFECT OF RETAKING ON SUBSEQUENT OFFERS AND CHOICES: BY FIELD OF STUDY

	Health		Engineering		Sciences		Humanities		Business	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Offers										
Retake in 2007	0.298**	0.203	0.234	0.149	0.013	-0.007	-0.047	-0.088	-0.902***	-0.837**
	[0.059;0.536]	[-0.021;0.565]	[-0.152;0.621]	[-0.335;0.705]	[-0.058;0.085]	[-0.100;0.120]	[-0.271;0.176]	[-0.399;0.244]	[-1.456;-0.348]	[-1.625;-0.157]
Observations	9,654	9,654	12,050	12,050	12,768	12,768	12,018	12,018	11,441	11,441
Kleinbergen-Paap Wald F-stat	16.780		21.120		25.720		21.390		19.800	
Panel B: Choices										
Retake in 2007	0.094	0.121	0.278	0.284	-0.036	-0.056	0.072	0.024	-0.480***	-0.361*
	[-0.120;0.307]	[-0.207;0.428]	[-0.110;0.665]	[-0.157;0.900]	[-0.087;0.015]	[-0.163;0.007]	[-0.099;0.243]	[-0.272;0.266]	[-0.841;-0.119]	[-0.863;0.103]
Observations	15,521	15,521	11,540	11,540	14,154	14,154	16,229	16,229	12,560	12,560
Kleinbergen-Paap Wald F-stat	32.860		20.450		29.060		38.830		23.610	
Optimal Bandwidth Linear 2SLS	✓		✓		✓		✓		✓	
Optimal Bandwidth Linear CCT		✓		✓		✓		✓		✓

Notes: The table presents the estimated effects of retaking the exam on a set of outcomes. Panel A refers to offers: in columns (1) and (2) the outcome is the probability to get an offer from a health-related college-course; in (3) and (4) it is the probability to get an offer from an engineering-related college-course; in (5) and (6) it is the probability to get an offer from a science-related college-course; in (7) and (8) it is the probability to get an offer from a humanities-related college-course; and in columns (9) and (10) the outcome is the probability to get an offer from a business-related college-course. Panel B refers to choices: in columns (1) and (2) the outcome is the probability to make at least 50% of the choices for a health-related college-course; in columns (3) and (4) the outcome is the probability to make at least 50% of the choices for a engineering-related college-course; in columns (5) and (6) the outcome is the probability to make at least 50% of the choices for a science-related college-course; in columns (7) and (8) the outcome is the probability to make at least 50% of the choices for a humanities-related college-course; in columns (9) and (10) the outcome is the probability to make at least 50% of the choices for a business-related college-course. Odd columns show estimates from a 2SLS estimation and even columns use the robust nonparametric approach of Calonico, Cattaneo and Titiunik (2014a). Optimal bandwidths estimated using the CCT approach are used in all columns. Standard errors are clustered at school level and 95% robust confidence intervals are in square brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table A9: EFFECT OF RETAKING: PSEUDO-THRESHOLDS

Alternative Distance from Threshold	MSE Optimal Bandwidth	RD Estimator	p-value	95% Robust CI	Number Observations Left	Right
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Academic Performance						
-1,000	330.324	-146.748	0.910	[-19,387.410;17,262.110]	1,364	1,598
-500	400.455	-641.542	0.884	[-13,387.880;11,537.740]	1,997	2,002
0	1,150.994	2,588.643	0.001	[1,112.680;3,993.072]	5,597	5,781
500	318.109	466.800	0.896	[-10,724.850;9,379.879]	1,538	1,612
1,000	317.713	-5,280.261	0.840	[-55,468.060;45,085.740]	1,601	1,622
Panel B: Offer From Top College-course						
-1,000	343.129	0.108	0.750	[-1.311;1.820]	1,413	1,654
-500	408.860	0.137	0.793	[-0.974;1.274]	2,035	2,043
0	1,390.115	0.117	0.249	[-0.102;0.391]	6,595	6,967
500	303.722	0.553	0.709	[-1.533;2.252]	1,456	1,549
1,000	285.068	-1.316	0.754	[-11.502;,8.329]	1,431	1,455
Panel C: Academic Offer						
-1,000	358.503	-2.004	0.596	[-11.736;6.744]	1,484	1,725
-500	303.684	-1.559	0.450	[-10.715;4.750]	1,520	1,530
0	1,450.113	0.138	0.732	[-0.326;0.464]	6,836	7,273
500	349.439	1.033	0.733	[-2.628;3.738]	1,687	1,776
1,000	383.305	-10.765	0.942	[-116.040;107.769]	1,950	1,945

Notes: Each panel refers to a different outcome. These outcomes are: academic performance (Panel A), the likelihood to obtain an offer from a college-course that is above median quality (Panel B), and the likelihood of obtaining an academic offer compared to an offer from a vocational school or no offer (Panel C). Each row corresponds to a different threshold. The true threshold (10,000) corresponds to 0 and the estimated effects when the true threshold is used are shown in column 6 in Tables 2 and 4. Results are obtained following the approach of Cataneo, Idrobo and Titiunik (2019).

Table A10: EFFECT OF RETAKING: ELIMINATING OBSERVATIONS AROUND THE THRESHOLD

Donut Hole Radius	MSE Optimal Bandwidth	RD Estimator	p-value	95% Robust CI	Number Observations	Number Observations Excluded Left	Right
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Academic Performance							
0	1,150.994	2,588.643	0.001	[1,112.680;3,993.072]	11,378	0	0
100	1,294.914	2,592.571	0.000	[1,346.185;3,749.882]	11,614	529	532
200	1,011.195	2,649.969	0.042	[91.366;4,884.497]	8,008	1,019	1,023
Panel B: Offer From Top College-course							
0	1,390.115	0.117	0.249	[-0.102;0.391]	13,562	0	0
100	848.731	0.088	0.612	[-0.250;0.425]	7,451	529	532
200	761.186	0.109	0.825	[-0.750;0.940]	5,604	1,019	1,023
Panel C: Academic Offer							
0	1,450.113	0.138	0.732	[-0.326;0.464]	14,109	0	0
100	1,196.320	0.255	0.280	[-0.191;0.660]	10,720	529	532
200	1,028.580	0.386	0.128	[-0.153;1.218]	8,175	1,019	1,023

Notes: Each panel refers to a different outcome.These outcomes are: academic performance (Panel A), the likelihood to obtain an offer from a college-course that is above median quality (Panel B), and the likelihood of obtaining an academic offer compared to an offer from a vocational school or no offer (Panel C). Each row corresponds to a different donut hole radius. The radius of 0 corresponds to the main estimated effects (threshold at 10,000) reported in column 6 in Tables 3 and 4. We also present estimates when we eliminate +/-100 and +/-200 scores around the threshold. Results are obtained following the approach of Cataneo, Idrobo and Titiunik (2019).

Table A11: EFFECT OF RETAKING ON ACADEMIC PERFORMANCE USING ALTERNATIVE ADMISSIONS SCORES FOR THOSE WHO RETAKE MORE THAN ONCE

	+/-1,000	+/-1,500	+/-2,000	Optimal	Optimal
	(1)	(2)	(3)	(4)	(5)
Retake	2,112.036***	2,412.980***	2,627.979***	2,195.038***	2,395.353***
	[838.859;3,385.214]	[1,522.164;3,303.796]	[2,033.190;3,222.768]	[1,043.760;3,346.316]	[1,016.601;3,774.105]
Observations	9,948	14,558	18,932	11,422	11,422
Kleinbergen-Paap Wald F-stat	16.970	30.110	65.380	19.330	.
Optimal Bandwidth Linear 2SLS	✓	✓	✓	✓	
Optimal Bandwidth Linear CCT					✓

Notes: The table presents the estimated effects of retaking the exam on academic performance in the exam. In this table we include students who retake the exam more than once. For students who retake the exam more than once, we use their most updated admissions grade and we take into account up to three retakes. Columns (1)-(4) show estimates from a 2SLS estimation and column (5) uses the robust nonparametric approach of Calonico, Cattaneo and Titiunik (2014a). A linear polynomial of the running variable is used in all estimations. The bandwidth for the (1), (2) and (3) column are +/-1,000, +/-1,500, +/-2,000, as indicated in the top row. Columns (4) and (5) use the optimal bandwidth from the CCT estimation. Standard errors are clustered at school level and 95% robust CIs are in square brackets. *** p<0.01, ** p<0.05, * p<0.1.

B Appendix

B.1 Identification Strategy: Further Discussion

A fuzzy regression discontinuity design can be thought as a particular case of an instrumental variable. Following Imbens and Angrist (1994), under the following assumptions the IV can be consistently estimated: First, we need a relevance condition that missing the threshold of 10,000 marks needs to change the likelihood that students retake the national exams. Second, we assume monotonicity: missing the eligibility 10,000 marks weakly increases students' likelihood to retake the national exams. Third, our instrument needs to be (conditionally) randomly assigned. Fourth, we require an exclusion restriction that just missing the eligibility threshold of 10,000 only affects students' academic performance and quality of placement through its effect on the decision to retake. If these assumptions hold, then we can interpret β_2 (on page 13 of the paper) as the direct local average treatment effect (LATE). Otherwise, we would have to focus on the reduced-form parameter and look at the effect of scoring just below the 10,000 threshold on later outcomes. We elaborate on how the identifying assumptions are likely met in our setting in the following paragraphs.

The institutional setting makes us confident about assumptions 1 and 2. In particular, missing the threshold indeed makes students more likely to retake the national exams so that they increase their chances of gaining access to tertiary education. The relevance assumption is testable and we show in section 4.2 in the paper our first stage results. Specifically, in Figure 1, we find that just missing the threshold leads to a 50% chance of retaking the exam, while being above the threshold makes students 15 percentage points less likely to retake the exam, on average. With regards to the monotonicity assumption, we cannot directly test if students who would not retake the exam when below the 10,000 threshold but would choose to retake it when above this threshold since this is a counterfactual outcome. Nevertheless, given our institutional framework, it is hard to think of reasons why students would want to violate monotonicity.

For the assumptions of random assignment of the instrument and the exclusion restriction, we discuss the likely sources of identification threats below. We initially discuss the (conditional) random assignment of our IV. The institutional setting provides a promising background for such a conditional randomization. In order for β_2 (in our main regression on page 13) to be interpreted as the causal effect of retaking, we need a student's score with regards to the performance threshold to be randomly assigned. We show that our IV is randomly assigned to students around the 10,000 performance threshold. In particular, we provide evidence that the IV is as good as random and students just to the right are identical to those just to the left of the performance threshold in Sections 5.1 and 5.2 in our paper. We show that students have no way to manipulate whether they end up to the right or to the left of this performance threshold using a density test for the admission score, based on Cataneo, Idrobo and Titiunik (2019) (Figure 3B), and balancing exercises for our predetermined covariates (Figure 4). We provide an additional balancing test in Table B1 below. In particular, we show that around the 10,000 threshold (+/-1,500), students on each side of the threshold are statistically indistinguishable with regards to their predetermined characteristics, i.e., the probability of a student scoring below the performance threshold is uncorrelated with any of their predetermined covariates.

Table B1: Additional Balancing Exercise

Outcome: Binary Indicator for Below the Threshold	
High Income	0.006
	(0.005)
Urban High School	0.003
	(0.008)
Girl	0.002
	(0.005)
Top-performing High School (in 2005)	-0.004
	(0.004)
Private High School	0.001
	(0.017)
Observations	13,703
R-squared	0.742

Notes: The table reports estimated coefficients from a regression of a binary indicator that takes the value of one if a student scored below the 10,000 threshold (our IV) in the national exams on all available predetermined variables as indicated in row headers. A constant and the distance of the test score relative to the 10,000 threshold are also included in the regression. Standard errors clustered at the high school level are reported in parentheses. Note that the number of observations within the 1,500 bandwidth in this Table are slightly fewer than the 14,558 observations we state for the 1,500 bandwidth in Tables 1 - 4 in the main paper. This is due to the control on private/public schools for which we have some missing values and which is not part of the main analysis of the paper. * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.

Moving on to the exclusion restriction, we provide some additional robustness exercises. We follow Angrist and Pischke (2009) and estimate the first-stage equation separately for different sub-populations, based on pre-determined characteristics. We stratify the sample based on high vs low socioeconomic background, urban vs rural high school, student gender, high vs low quality high school, and private vs public high school. For each subpopulation, the ratio of the first stage coefficient in the sub-population to the overall first-stage coefficient indicates the relative likelihood that compliers come from a given sub-population. The first stage regression for each sample is shown in Table B2 in the Appendix. Most coefficients are close to the full sample first stage (column 1 below). We see strong first stages for high income, low income, urban postcode, male, high quality high school, low quality high school, private and public. The magnitude is largest for students in private schools, however observations from private schools account for just 7% of the full sample (Table 1 in the paper). The coefficient for females is positive, smaller than the full sample first stage and imprecise. The same applies to rural, although we have only 857 students. All remaining first-stage coefficients are very similar to the full sample first-stage estimate. We see very little difference for students with high or low income postcodes or for students in high vs low quality schools. Generally, the results indicate that there is not vast meaningful heterogeneity in the first stage.

Table B2: FIRST STAGE ESTIMATES FOR DIFFERENT SUB-GROUPS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Full Sample	High Income	Low Income	Urban	Rural	Girls	Boys
First Stage	0.091***	0.089***	0.076***	0.085***	0.035	0.033	0.139***
	(0.017)	(0.024)	(0.029)	(0.020)	(0.064)	(0.025)	(0.028)
Observations	14,558	6,553	4,825	9,900	857	6,167	5,211
	(7)	(8)	(9)	(10)			
	High Quality HS	Low Quality HS	Private	Public			
First Stage	0.078***	0.085***	0.274***	0.073***			
	(0.024)	(0.030)	(0.077)	(0.019)			
Observations	6,557	4,775	471	10,907			

Notes: The table shows the first-stage relationship between students' retaking probability and a binary indicator that indicates whether students missed the performance threshold, stratified by various sub-samples. In column 1 we show the overall first stage within the 1500 bandwidth (as can be found in Table 2 of the paper). For the remaining estimates we use the optimal bandwidths calculated as we describe in the text. F-statistics for the sub-populations can be found in Tables A2-A5 in the Appendix of the paper. Standard errors clustered at the high school level are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.

Furthermore, we note that the cost of retaking the exams is similar for all students around the threshold who retake the exams in this institutional setting. Everyone around the threshold has performed poorly to start with and everyone is going to bear the full cost of retaking in the same way. There is only one exam period per year, and thus retakers have to wait for a whole year to retake the exams in the next exam period, which is usually in May-June. Students usually prepare for these exams during the year, while they do not usually attend any other educational institution or work. Thus, everyone who is retaking would exert a lot of effort and try to achieve the best possible outcome out of this. A considerable share of students (43%) who exceeded the 10,000 performance threshold in their first attempt did not get any university offer, and thus they are in a similar situation as students who did not pass the 10,000 performance threshold in their first attempt and retake. It would be a waste of time and resources for everyone who retakes the national exams to not invest serious effort in it.

One could also think that potential student outcomes may be different due to differences in graders' marking practices or potential discrimination against students who failed in the first attempt. Graders who mark national exams do not know whether a student scored below the threshold in their first attempt. In fact graders do not know whether the script they are marking belongs to a student who is retaking or not the exams, let alone whether they failed or not in the first attempt. The school name, student name, teacher name, and any background information of the student is concealed in those exams Lavy and Megalokonomou (2019). Thus, we think it is unlikely that graders' marking behaviour changes.

C Appendix

Figure C1: EXTRACT FROM THE REFORM'S LAW

ΚΕΦΑΛΑΙΟ Γ΄

ΘΕΜΑΤΑ ΕΙΣΑΓΩΓΗΣ, ΜΕΤΕΓΓΡΑΦΩΝ
ΚΑΙ ΚΑΤΑΤΑΞΕΩΝ
ΣΤΗΝ ΤΡΙΤΟΒΑΘΜΙΑ ΕΚΠΑΙΔΕΥΣΗ

Άρθρο 13

Εισαγωγή στην τριτοβάθμια εκπαίδευση

1. Στο ν. 2525/1997 (ΦΕΚ 188 Α΄) μετά το άρθρο 2 προστίθεται άρθρο 2Α ως εξής:

«Άρθρο 2Α

1. Ως προϋπόθεση συμμετοχής του υποψηφίου στη διαδικασία επιλογής των εισαγομένων στα Πανεπιστήμια, στα Τ.Ε.Ι., στην Α.Σ.ΠΑΙ.Τ.Ε., στις Ανώτερες Σχολές Τουριστικής Εκπαίδευσης (Α.Σ.Τ.Ε.) των Σχολών Τουριστικών Επαγγελμάτων του Υπουργείου Τουριστικής Ανάπτυξης, στα Ανώτατα Στρατιωτικά Εκπαιδευτικά Ιδρύματα και τις Σχολές Υπαξιωματικών των Ενόπλων Δυνάμεων, στις Ακαδημίες Εμπορικού Ναυτικού του Υπουργείου Εμπορικής Ναυτιλίας, στις Σχολές Αστυνομικής Ακαδημίας του Υπουργείου Δημόσιας Τάξης και στη Σχολή Ανθυποπυραγών της Πυροσβεστικής Ακαδημίας του Υπουργείου Δημόσιας Τάξης ορίζεται η επίτευξη από τον υποψήφιο γενικού βαθμού πρόσβασης τουλάχιστον ίσου με το μισό του μέγιστου δυνατού βαθμού πρόσβασης. Ειδικά, υποψήφιος που δεν πληροί την προϋπόθεση αυτή συμμετέχει στη διαδικασία επιλογής των εισαγομένων στα ανωτέρω ιδρύματα εφόσον συγκεντρώνει σύνολο μορίων τουλάχιστον ίσο με το μισό του μέγιστου δυνατού αριθμού μορίων.

Notes: This comes from Law 3404/2005, Article 13 and describes the new requirements for gaining access to tertiary education. In particular, it states that new applicants should obtain an admission score that is equal to at least 50%. The federal document can be accessed here: http://www.pi-schools.gr/special_education_new/ftp/nomoi/Ekp_Themata/N.%203404%20-2005%20-%20FEK.%20260%20-A-%2017-10-2005.pdf

Figure C2: COUNTIES CONTAINING SCHOOLS IN THE SMALLER DATASET

Notes: This figure shows the counties with schools for which we have GPA scores and student birthdate.

Figure C3: DEFINITION OF THE EMPLOYMENT PROSPECTS INDEX

Insecurity index: >= 1 and < 1.5	>= 1.5 and <= 2	> 2 and <= 2.5	> 2.5 and 3
Employment prospects are			
Good	Mediocre	Poor	Very Poor
Economics	Mathematics and Statistics	Education, Greek,	Agriculture and Forestry
Engineering and Computer Science	Business and Management	Foreign Languages and P.E.	Liberal Art and Humanities
Biology	Physics and Earth Science	Social, Political, and European Studies	Home Economics
Nursing and Other Health	Psychology	Other	
Medicine	Law	Journalism	
Pharmacy			
Naval Academies			
Police and Military			
Veterinary Science			

Notes: This table comes from Goulas and Megalokonomou (2019). For each university department, the index takes a value between 1 and 3, and indicates how good the employment prospects are for graduates of those university departments. In particular, if the index is between 1 and 1.5 the employment prospects are good, if the index is between 1.5 and 2 the employment prospects are mediocre, if the index is between 2 and 2.5 the employment prospects are poor, and if the index is between 2.5 and 3 the employment prospects are very poor. As discussed in that paper, the job insecurity index is the result of the amalgamation of information from the career offices of all universities in Greece, the Hellenic Bureau of Statistics, the employment observatory, and various labor unions. The index is intended to represent differences in structural and frictional unemployment among those with available university degrees and time-specific labor market conditions. This index refers to students who apply to university departments in 2006.