

Swiss Leading House

Economics of Education • Firm Behaviour • Training Policies

Working Paper No. 241

**Do High-Stakes matter? Evidence from a
Probation Policy in Higher Education**

Enzo Brox and Michael Dörsam



Universität Zürich
IBW – Institut für Betriebswirtschaftslehre

u^b

^b
UNIVERSITÄT
BERN

Working Paper No. 241

Do High-Stakes matter? Evidence from a Probation Policy in Higher Education

Enzo Brox and Michael Dörsam

May 2025

Please cite as:
"Do High-Stakes matter? Evidence from a Probation Policy in Higher Education."
Swiss Leading House "Economics of Education" Working Paper No. 241, 2025. By
Enzo Brox and Michael Dörsam.

Die Discussion Papers dienen einer möglichst schnellen Verbreitung von neueren Forschungsarbeiten des Leading Houses und seiner Konferenzen und Workshops. Die Beiträge liegen in alleiniger Verantwortung der Autoren und stellen nicht notwendigerweise die Meinung des Leading House dar.

Discussion Papers are intended to make results of the Leading House research or its conferences and workshops promptly available to other economists in order to encourage discussion and suggestions for revisions. The authors are solely responsible for the contents which do not necessarily represent the opinion of the Leading House.

The Swiss Leading House on Economics of Education, Firm Behavior and Training Policies is a Research Program of the Swiss State Secretariat for Education, Research, and Innovation (SERI).

www.economics-of-education.ch

Do High-Stakes matter? Evidence from a Probation Policy in Higher Education*

Enzo Brox
University of Bern[†]

Michael Dörsam
BIBB

May 5, 2025

Abstract

Universities in many countries have introduced early requirements that students need to fulfill to be allowed to proceed. However, their effectiveness is disputed because of concerns that they harm graduation chances. We study the impact of a last-chance exam mechanism applying a quasi-experimental design. We exploit the discontinuity in treatment status around the promotion cutoff and show that exposure to a last-chance exam increases early dropout. Graduation chances are on average not affected. However, we find small negative effects on the graduation chances of female students. Further performance measures are largely unaffected.

Keywords: PROBATION, HIGH-STAKES, HIGHER EDUCATION, DROPOUT, REGRESSION DISCONTINUITY DESIGN

JEL Classification: I21, I23, J24

*We are very grateful for the constructive comments from Luna Bellani, Marco Caliendo, Beatrix Eugster, Moritz Janas, Tommy Krieger, Michael Lechner, Stephan Maurer, Guido Schwerdt, Maurizio Strazzeri, Heinrich Ursprung and Ulf Zölitz. We also thank several seminar and conference participants for their helpful feedback. We are appreciative of the financial supports from the Graduate School of Decision Science (University of Konstanz). Enzo Brox is grateful to the State Secretariat for Education, Research and Innovation for financial support through the Leading House ECON-VPET.

[†]University of Bern, Center for Research in Economics of Education and Swiss Institute for Empirical Economic Research at the University of St. Gallen; enzo.brox@unibe.ch.

1 Introduction

Misinformation about graduation chances ([Stinebrickner and Stinebrickner, 2012, 2014](#)) and procrastination in response to new information about these chances ([DellaVigna, 2009](#)) can harm students' investment decisions in higher education. To address this, universities worldwide, including the United States, Canada, France, Switzerland, Austria, the Netherlands, and Germany, have introduced probation policies. These policies typically pair an early performance requirement with a rule that forces students to leave the program after repeatedly failing to meet the standard. While probation policies potentially reduce the costs of inevitable dropout decisions or may serve as an early warning signal helping to increase student effort, they also constitute an additional barrier to graduation and may harm long-run success or cause lower-graduation rates ([Lindo et al., 2010](#)).

Increasing the number of university graduates is a key policy goal in advanced economies, and enrollment rates have increased significantly over the past two decades ([Altbach et al., 2019](#); [UNESCO, 2022](#)). However, concerns about low graduation rates persist, with more than 30% of undergraduates failing to graduate ([Vossensteyn et al., 2015](#); [Bound et al., 2010](#)). As a result, many policymakers focus on programs and policies that aim to improve the success of low-performing students. Two prominent interventions are remedial education and academic probation. Remedial education assigns students to additional classes to help them acquire the necessary skills, while academic probation serves as a signal of unsatisfactory performance, coupled with the negative incentive of suspension after repeated failure. Despite the widespread implementation of these policies, there is little consensus on their effectiveness or optimal design. This lack of agreement is reflected in the wide variation in timing, requirements, and penalties between institutions ([Oreopoulos, 2021](#)).

We offer new insights on the effectiveness of probation policies by studying the consequences of a probation policy that exposes low-performing students to a last-chance exam in the first year of a university program. A major obstacle to identifying the consequences of academic probation is the potentially endogenous nature of the decision whether to introduce a probation policy. Furthermore, students on probation are likely to differ from other students in terms of important determinants that influence their subsequent performance. Thus, comparisons of students on probation with other students are likely contaminated by omitted variable bias.

Our testing ground is Germany. In 2000, the federal government of Baden-Württemberg, a big federal state in the South of Germany, introduced a probation policy by implementing the so-called orientation exams. Unlike regular compulsory exams, orientation exams must be passed during the early stages of the program and typically allow only a limited number of retakes. Due to the reform, students who do not pass the first attempt of an orientation exam are forced to attend a last-chance exam.

We first exploit regional variation in the implementation of the probation policy and document a significant increase in early dropout rates after the introduction of the reform. To understand whether the increase in early dropout constitutes a concern for graduation chances or a reduction in delayed dropout, we use detailed student records and self-collected data on exam performances of a medium-sized public university. The data comprises several cohorts of undergraduate students and allows us to exploit the promotion rule, which generates a sharp discontinuity in the probability of being exposed to the last-chance exam.

We estimate the causal effect of an alternative form of academic probation that exposes low-performing students to a last-chance exam on 1) the probability to dropout early, 2) the probability to graduate and, 3) further performance measures such as the final GPA, the time to degree, and the performance in follow-up courses. To identify the mechanism that leads to our results we differentiate between the effect of i) exposure to an additional exam with a 'natural' risk of failure and ii) the early pressure that makes orientation exams different from non-orientation exams. We use data on non-orientation exams that take place in the first year and apply a difference-in-discontinuity design to difference out the effect of exam failure. Finally, we follow several arguments of the behavior economic literature that mostly theoretically and experimentally document gender differences in response to high-stake situations and provide further evidence from the field ([Gneezy et al., 2003](#); [Niederle and Vesterlund, 2007, 2010](#)).

Our results show that exposure to a last-chance exam significantly increases early dropout. Marginal students exposed to the last-chance exam are more than 10 percentage points more likely to drop out before the next semester begins, which is equivalent to a more than doubled risk of dropping out at this stage. We investigate whether students voluntarily decide to leave the program after being on probation, or whether they ultimately fail. We show that the effect is driven by the ultimate failure to meet the performance requirement, which is noteworthy given the high costs of such failures, as students are not allowed to study the same subject at any other

public university in Germany. Despite this substantial increase in early dropout, we do not find that exposure to a last-chance exam has on average a substantial negative impact on graduation chances.

However, when investigating heterogeneous effects for demographic subgroups, we find two interesting results. First, while male students do not face any penalty in terms of graduation chances, female students' graduation chances are negatively affected. Second, we find that the policy has a negative impact on the graduation chances of students with a below-average high-school GPA, signaling a lower degree of academic preparedness.

We then turn towards studying the mechanism behind our results and differentiate between the effect of exposure to an additional exam and exposure to a high-stakes situation. Our main results are unchanged. However, we find that the mechanisms differ for the subgroups mentioned before. While our results suggest that for lower ability students, the effect is explained by the exposure to an additional exam, the effect for female students is likely driven by exposure to the high-stakes situation. This result is in line with previous studies documenting a negative effect of exposure to high-stakes settings for women. Finally, we investigate the impact on other important determinants of success in higher education and find that the policy has a negative impact on graduation in time, while leaving other learning outcomes mostly unaffected.

Our results provide interesting insights for policy makers. Policies targeting low-performing students are widespread. Several studies have analyzed the impact of remedial courses on students' academic success. In general, empirical studies widely fail to report positive graduation effects and thus remedial education is often criticized for its high costs (Duchini, 2017). In contrast, probation policies are less costly. Our results suggest that the policy is effective in reducing the number of late dropouts without adversely affecting graduation chances, and may lower the individual and societal costs of procrastination (Berens et al., 2018; Bowen et al., 2009; Bound et al., 2010). However, there are two concerns. First, we show that there are no positive effects in terms of learning benefits, and second, we observe, in line with some previous literature, a negative effect for female students. These results should be carefully considered.

1.1 Relation to Literature

This paper connects to three main areas of the literature. First, it contributes to research on the organization of higher education and policies aimed at improving the performance and success

of low-performing students. Although some studies focus on financial aid (Scott-Clayton, 2011; Montalbán, 2022) or college costs (Garibaldi et al., 2012; Ketel et al., 2016; Bietenbeck et al., 2023), others examine the design of the program or the effectiveness of remedial courses (Jacob and Lefgren, 2004; Bettinger and Long, 2009; Martorell and McFarlin, 2011; Duchini, 2017; Malacrino et al., 2024).

Closest to our study are a number of studies examining the impact of academic probation on college success. Since the effectiveness of these policies depends on their design, it is crucial to analyze the mechanisms involved to identify the best way to support low-performing students (Oreopoulos, 2021). So far, research has basically studied two different mechanisms. First, some studies have analyzed the effects of probation, in a set-up where repeated failure to meet a GPA threshold results in part-time or full-time suspension. The results are mixed. Lindo et al. (2010) find that being placed on probation increases early dropout, especially for men and students with above median high-school GPA. They also find some evidence for decreased graduation rates, driven mainly by the latter group of students. Similar results are found by Ost et al. (2018).¹ In contrast, Casey et al. (2018) find an increase in short run persistence, but fail to find significant effects on graduation, which is consistent with the findings presented by Fletcher and Tokmouline (2017). Second, Tafreschi and Thiemann (2016) examine a set-up where academic probation is combined with a retention mechanism. Unlike the previous setup, where remedial courses are offered, students on probation here are required to repeat the entire first-year curriculum. Although this delays graduation, it also helps students gain the necessary skills. Tafreschi and Thiemann (2016) find that forcing students to repeat the first year increases early dropout rates. However, they also report positive effects on GPA at graduation and time to degree for students who stay and pass the first year.

Our study adds to this literature by evaluating the effects of a different type of probation policy—one that does not mechanically delay graduation and can be implemented at low cost. To our knowledge, this is the first paper to examine the consequences of early exposure to a last-chance exam. More broadly, as limiting exam attempts is common in universities, our article explores how students respond to the threat of suspension, as repeated failure after probation leads to suspension. Thus, our findings also contribute to the wider literature on higher education

¹Ost et al. (2018) furthermore examine labor market outcomes and find lower earnings for students placed on probation.

policies, including the impact of restricting exam attempts, or in more general the consequences of test retaking (Bratti et al., 2024; Goodman et al., 2020; Frisancho et al., 2016; Nijenkamp et al., 2022).²

Second, we contribute to the literature on the effects of high-stakes testing on student achievement at different educational levels. Although studied primarily in primary and secondary education—particularly after the No Child Left Behind initiative in the US—it remains controversial among policymakers. Critics argue that high-stakes exams can cause negative effects, such as test anxiety, fear of separation from peers, and reduced motivation. Early, mainly correlational, studies suggest a link between high-stakes testing and higher dropout rates (AERA, 1999; Shriberg and Shriberg, 2006). Supporters claim that it increases motivation and performance. For example, Hanushek and Raymond (2004) finds a positive link between high-stakes testing and student performance, with some demographic variation, while Jacob (2005) shows performance improvements, limited to the subjects tested. We extend this literature by examining the effects of high-stakes testing in higher education. The impact may differ, as our subjects are older and better equipped to handle exam pressure. In addition, in higher education, program choice is voluntary, which may increase motivation but also offer more outside options. To our knowledge, this is the first paper to provide causal evidence on the effects of increased stakes in higher education. We also explore how exposure to high-stakes exams affects key labor market outcomes, such as GPA, time to degree, and performance in follow-up exams (Altonji et al., 2011; Arcidiacono, 2004).

Third, our results also speak towards a more general literature investigating gender differences in response to increases in stress, pressure, and stakes. Most laboratory studies examine the effect of increases in competitive pressure and generally find that female students tend to respond negatively compared to male students (Gneezy et al., 2003; Niederle and Vesterlund, 2007, 2010). Other studies use an increase in the stakes associated with exams to study gender differences in the effect of increased pressure (Azmat et al., 2016; Montolio and Taberner, 2021). They also find that female performance declines as the stakes increase. We view our study as complementary to these findings. Our results suggest higher costs for female students when exposed to

²Goodman et al. (2020) exploit students' increased likelihood of retaking the SAT due to left-digit bias and find positive effects on both SAT scores and the chances of gaining admission to a four-year college in the US. Frisancho et al. (2016) observe comparable results in the context of Turkey's college entry exam, noting that retaking the exam leads to cumulative learning and, consequently, improved scores, especially among disadvantaged students.

high-stake situations.

We proceed as follows. Section 2 informs about the institutional framework and shows how the introduction of orientation exams affected early dropout rates. Section 3 describes our data set. Section 4 describes our identification strategy. Section 5 presents our results. Section 6 concludes.

2 Orientation exams

2.1 Institutional Framework

In September 1999, the federal government of Baden-Württemberg passed a law that obliged all universities to implement orientation exams, from the winter term 2000 onward, as part of the first year curriculum.³ The main objective of the reform was to have a strict performance requirement, which allows to identify in an early stage of a program whether a student has the required skills and motivation to successfully proceed in a program (Bölke et al., 2020).

Importantly, the reform did not enforce any content or structural changes in the curriculum, but it enforced all programs to define some (core) exams as orientation exams for which the examination regulations changed. The examination regulations of orientation exams differed from non-orientation exams for three reasons: First, a student must pass all orientation exams by the end of the third term. If a student does not satisfy this requirement, he/she will be forced to drop out of the program. Second, a student who fails an orientation exam once, has only one more chance to pass the exam. If he/she fails again, he/she is automatically forced to drop out of the program and loses the eligibility to study the program at another German university.⁴ Third, a student is automatically registered for any orientation exam, while he/she registers himself for non-orientation exams.

In summary, in contrast to the prior examination regulations, students were now forced to write the exams and receive a strong performance signal in an early program phase. Although the implementation of orientation exams was mandatory, the reform gave some leeway to universities and departments with regard to the design of the policy. Thus, the probation policies implemented in Baden-Württemberg differ slightly between and within universities.

³Baden-Württemberg is the third largest federal state in Germany. It is located in the South of Germany with borders to France and Switzerland.

⁴In non-orientation exams there is usually at least a second retake exam possible.

2.2 Orientation exams and early dropout

To provide some evidence on the meaningfulness of the probation policy, we now show that the reform had a visible effect on early dropout rates. We use data from the German Student Register from 1996 to 2006 (RDC, 2017). The Student Register encompasses the full population of students enrolled at universities and universities of applied sciences in Germany during a given year. It compiles data from the administrative records of these institutions, including individual-level details about current enrollment (such as program, degree type), the institution and year of initial enrollment, and demographic information (e.g., gender, nationality, and high school county). Due to strict data protection regulations, the register does not contain personal student identifiers. As a result, students cannot be tracked over time. However, we can construct a panel at the university-cohort level and we can observe how the size of the cohort develops over time.⁵

We follow this idea and define the early dropout rate of cohort c at university u . In line with the implemented reform, we define dropout as early dropout if it occurs within the first two years. First, we calculate the number of first-year, second-year and third-year students at university u for each cohort c in every year. Second, we calculate the difference between the number of first-year students of cohort c in year $t - 2$ and the number of third-year students of cohort c in year t . Third, the dropout rate in the first two years is then calculated by dividing the result from step two by the number of first-year students of cohort c in year $t - 2$. Because of the two-year time lag that is necessary to observe students in the third-year, the 2004 cohort is the last cohort we consider in our analysis.

We then use a simple difference-in-differences specification to compare the dropout rates of starting cohorts in the treated state (Baden-Württemberg) with those of starting cohorts in untreated states before and after the introduction of the reform. We begin with the following basic model:

$$early\ dropout_{cs} = \alpha + \beta BW_s + \gamma Post_c + \delta(BW \times Post)_{cs} + \eta_{cs} \quad (1)$$

where $early\ dropout_{cs}$ denotes the early dropout rate of cohort c in state s . The dummy variable BW_s indicates whether the cohort was a cohort at a university in Baden-Württemberg.

⁵To have a balanced panel in our baseline specification we exclude universities in Hamburg from our sample as they were not sampled in 1997/1998. Our final sample comprises about 3.75 million student-year observations from 72 universities. For a list of universities, see Table 1.

$Post_c$ is 1 if the cohort started after 1999 and 0 if the cohort started before. We are interested in the interaction of these two dummy variables $(BW \times Post)_{cs}$. η_{cs} is the error term. We cluster standard errors at the university level and therefore allow for arbitrary correlation of these error terms across universities.⁶

Our results are reported in Table 2. We start with a reduced sample of two years before and after the reform. In Column 2, we use the full period from 1996-2004. In Column 3 we replace the BW_s dummy with state fixed effects and the $Post_c$ dummy with cohort fixed effects. In Column 4 we add a set of university control variables and in Column 5, we add group-specific linear time trends. The effect of the introduction of orientation exams on early dropout is roughly 3 p.p. This corresponds to a 10% increase relative to the pre-reform mean. The effect is significant at the five percent level across all specifications.

The main assumption underlying our approach is the parallel trend assumption. A typical approach to verifying this assumption is to generate placebo treatments in order to test whether the policy had an effect on the outcome before the policy has been introduced.

The results are plotted in Figure 1. The coefficients plot shows that before the policy was implemented the average conditional evolution of the outcome variable over time was parallel across treated and non-treated cohorts. The coefficients of the interaction of the state dummy with the post-reform years ($k > 1999$) are similar to the effect estimated in Table 2. The effect remains stable and persists in the four years after the introduction of the policy.⁷

We provide a couple of further robustness checks to verify our result from Table 2. In Table 3 we show that our results are not driven by a particular university treated. To show this, we provide results from specifications in which we one after the other exclude one treated university from our sample. Our results remain unaffected. In Table 4 we summarize several other robustness checks. First, we show results in which we do not treat each university cohort as an equally weighted unit, but do weight each cell according to its size. Furthermore, we conduct a robustness check in which we define the cohort at the department level rather than the university level. Columns 1 and 2 of Table 4 confirm our results. Second, we exclude universities that

⁶In the appendix, we present alternative methods of inference.

⁷A further concern is the existence of unobserved shocks that correlate with the reform. We are not aware of any other changes in the legislation. In addition, no comparable policy has been introduced in other federal states during our observation period. However, universities or departments outside of Baden-Württemberg may also have voluntarily implemented a comparable policy at the same time as Baden-Württemberg. Our estimates should therefore be interpreted as a lower bound of the treatment effect.

during our observation period report in some years unusually high or low dropout rates (Column 3). In Column 4 and 5, we show that the majority of the observed effect is caused by dropout in the second year, when the administrative rules bite. However, we also find a small positive effect in the first year. In the remaining columns, we show specifications with alternative methods to compute standard errors. In our previous specification, we used clustered standard errors at the university level. We provide additional tests using clustered standard errors at the state level, the state-year level, or we use a two-way clustering approach following [Cameron et al. \(2011\)](#) (Columns 6-8). Finally, we also show that our results are robust to a wild-cluster bootstrap procedure (Column 9).

In this section, we have shown that the probation policy significantly affected student progress in higher education. In the remainder of this paper we turn towards detailed administrative student records of a university that implemented the policy. This allows us to track students over time and to see whether, beyond an effect on early dropout, there are costly consequences for graduation chances or other student performance measures.

3 Data

In the remainder of the study, we use data from one medium-sized public university to understand the consequences of exposure to a last-chance exam for graduation chances and other measures of educational performance. In this section, we explain the institutional framework and the data in more detail. The university offers a three-year (six-term) undergraduate program in economics that consists of 180 ECTS. In the first term, students have compulsory courses in business administration, economics, and mathematics. The assessment in all these courses is a written exam. In each term, there are two examination periods: the main examination period is right after the lecture period, and the retake examination period takes place right before the next term begins. Among the five compulsory courses, students have to write two orientation exams in the first semester ('Mathematics I' and 'Introduction to Economics').⁸ Both orientation exams are designed separately and are independently graded. In addition to the settings explained in Section 2, a student who does not pass the orientation exam on the first attempt has to repeat the orientation exam in the retake period of the same term. The student can only proceed to the

⁸There are two more orientation exams in the second semester. In this paper, we focus on the first semester courses to avoid selection issues.

second term if he/she passes the retake exam and will be exmatriculated otherwise.

In the second and third term, students have to take additional compulsory courses. Afterwards, they choose courses related to their field of specialization. Students graduate on time if they complete the program by the end of the sixth term. The final grade is a weighted average of all course grades.

We compile administrative data on six cohorts of undergraduate students in economics between 2007 and 2012.⁹ The data include some background characteristics (age, gender, high school GPA, school duration) and a unique identifier that allows tracing students over time. We also have detailed student-term specific information such as enrollment status, course choice and exam performance in all courses. Furthermore, the data includes administrative graduation information.

In our baseline analysis we impose two sample restriction. First, we only include economics undergraduate students who started their program between 2007 and 2012. The reason for this is censoring in 2018. Thus, we can track all undergraduates for at least six-years. Second, our identification strategy requires that first term students show up in at least one orientation exams. We therefore exclude all students who dropped out without showing up for any orientation exam. In total, our baseline sample includes 1562 students.

Panel A of Table 5 presents summary statistics on students' background characteristics. We observe that students are on average between 20 and 21 years old when they enroll. The share of male students is 56% and thus almost identical to the share of male students in economics in Germany (BIBB, 2018). The average high-school GPA of our sample (2.37) also closely resembles the German average (2.39). Panel C of Table 5 focuses on our main outcome variables. We focus on two outcome variables. First, early dropout (at the end of semester 1). The dropout information is constructed using the enrollment information in our data. The detailed student-exam records and the knowledge of the precise promotion policy also allow us to identify those students who are forced to leave the program, because of repeatedly failing to meet the performance requirement and students who voluntarily decide to leave the program. Second, we use our administrative information on the graduation status to understand long-run consequences of the probation policy. This allows us to assess whether the increase in early dropout has costly

⁹Because of administrative changes and lack of data we cannot take earlier cohorts into account. This data restriction also prohibits exploiting the introduction of the policy.

consequences for graduation rates. Although about 12% of the students drop out after the first semester, around two-thirds finally graduate from the program.¹⁰

Our treatment is defined in the following way. Since all students are exposed to orientation exams, we exploit variation in exposure to the strictness of the policy. While some students pass the orientation exam in the first attempt, others fail and are exposed to a last-chance exam with high-stakes attached. To determine the treatment status, we therefore rely on the performance in the first attempt of the orientation exam. The performance cutoff is sharp such that students are exposed to the last-chance exam if and only if they fail to meet the performance cutoff in the first attempt.

Typically, performance in an exam is documented by grades. University grades range from 1.0 to 5.0.¹¹ The performance requirement to pass an exam is a minimum grade of 4.0. Between 1.0 and 4.0 grades are set in steps of 0.3/0.4. Below the minimum passing grade, there is only one grade: 5.0. Due to the grading scheme, observing the grade in an orientation exam does not allow to isolate the effect of exposure to a last-chance exam, since students with the minimum grade that suffices to pass and students who fail are likely to differ in several characteristics.

To rely on a set of comparable students, we collected the exact number of points students achieved at the first attempt of an orientation exam from the university archive. We digitized this information and added it to our student record data.

4 Empirical strategy

4.1 Estimation

Estimating the effect of exposure to a last chance exam using OLS will fail to provide unbiased estimates if students are selected into treatment based on unobserved factors that influence educational outcomes. In particular, lower ability, lack of motivation or parental background may generate a non-zero correlation between unobserved characteristics u_i and treatment:

$$\text{cov}(Treat, u) \neq 0. \tag{2}$$

¹⁰In Section 5.5 we also look at further important student outcomes, such as final GPA, time to degree and grades in follow-up exams.

¹¹1.0 is the highest grade and 5.0 the lowest grade a student can achieve.

We address this concern by taking advantage of the promotion policy, that forces all students who do not meet the promotion cutoff to retake the exam. This leads to a highly non-linear relationship between the number of points a student achieves in an exam and the probability that he is forced to retake the exam. Assuming that unobserved factors do not vary discontinuously around the promotion cutoff, this generates plausible exogenous variation in the treatment status. We can therefore identify the causal effect of exposure to a last-chance exam, by using a regression discontinuity design that relies on a comparison of students' academic achievement who scored just below and just above the promotion cutoff in the first exam attempt of an orientation exam.

The promotion cutoff is strictly enforced, such that all students scoring below the cutoff in the first attempt of the orientation exam are exposed to the last-chance exam. Thus, we estimate a sharp RDD that can be described by the following equation:

$$Y_{ic} = \alpha + \beta \text{probation}_{ic} + \gamma f(\text{score}_{ic}) + \delta X_{ic} + \nu_c + \eta_{ic} \quad (3)$$

where Y_{ic} represents the educational outcome, that is, a binary indicator equal to one if the student i of cohort c drops out or graduates from the program. probation_{ic} is a dummy variable equal to one if a student fails the first attempt of an orientation exam. Our running variable is score_{ic} which is the score relative to the promotion cutoff.¹² X_{ic} is a vector of demographic covariates, including the age of the student at the time of enrollment, nationality, gender, the time they spend at school, and their high school GPA. To account for time-varying conditions such as course contents and grading schemes we control for cohort fixed effects ν_c in several specifications. The error term η_{ic} is clustered at the level of the assignment variable.

The coefficient of interest is β . In the absence of discontinuities in observable and unobservable determinants of the educational outcome it can be interpreted as the effect of exposure to a last chance exam on marginal students' subsequent academic achievement. We estimate both, specification with local linear regressions on both sides of the cutoff and parametric specifications with second-order polynomials of our running variable. The optimal bandwidth relies on [Imbens and Kalyanaraman \(2012\)](#). Results using alternative bandwidths are provided as well.

In our baseline specification, a student is on probation if he at least fails to reach the promo-

¹²We use the percentage deviation to achieve comparability among exams with varying total number of points.

tion cutoff in one of the two orientation exams. We also show results that exclude students who have to retake both orientation exams, and results that exclude students who do not fail on any first attempt of an orientation exam.

4.2 Validity of the RD Design

The main identification assumption for our empirical strategy is that the assignment around the cutoff is locally random (Lee and Lemieux, 2010). The assumption implies that students marginally above the performance cutoff are a good counterfactual for those students marginally below the cutoff and would achieve, on average, the same outcomes. If the local continuity assumption holds, the estimated effect measures the causal effect of exposure to a last chance exam on subsequent academic achievement (Imbens and Lemieux, 2008). The primary threat to identification is the possibility of precise manipulation of the assignment variable around the cutoff (Lee and Lemieux, 2010). The local randomness assumption may be violated if the probability of treatment depends on observable and unobservable characteristics that are correlated with educational performance. Before presenting the results, we provide several tests to address this threat.

First, we investigate the density of the assignment variable around the promotion cutoff. In line with several recent studies that find evidence for manipulation of test-scores in high-stake exams (Dee et al., 2019; Diamond and Persson, 2016; Machin et al., 2020), Figure 2 provides visible evidence for clustering to the right of the promotion cutoff. We observe significantly more students just above the cutoff than just below the cutoff. Even though a manipulation check following Cattaneo et al. (2020, 2024) using the full observation window does not confirm manipulation of the assignment variable, it does so when using a more restricted window around the cutoff. Therefore, in the remainder of this section, we discuss how we deal with the potential threat of sorting.

First, in Section 5 we show results from a *donut* approach in which we exclude students in a very narrow range around the cutoff. Our results are robust to this sensitivity check. Second, the observed pattern is problematic for our identification strategy under two potential scenarios: (i) if students can feasibly manipulate the number of points they achieve by exerting extra effort. (ii) if the teaching staff manipulates the cutoff by placing students with different observed or unobserved characteristics, which are correlated with the educational outcome, on either side of

the cutoff. Both scenarios discussed above are highly unlikely. First, students have, if anything, imprecise control over the exact number of points they achieve in the exam. Students may set their effort level over the semester to target a specific grade, but whether they achieve the exact number of points that is required to pass the course or one point less is unrelated to the effort level. Furthermore, in general, the required number of points is not publicly known prior to the exam. Second, in contrast to previous studies that find evidence for test-score manipulation at school, there is fewer prior contact between students and teachers, and exams are written and graded anonymously, using a student ID number. Thus, a systematic manipulation of the cutoff by teachers is unlikely.¹³ In the remark of this section, we provide graphical and analytical evidence justifying our arguments.

First, analogous to a test for balance of background characteristics in an experimental study, we verify that student pretreatment characteristics are smooth around the cutoff. Table 6 and Figure 4 provide evidence that this condition is verified for all background characteristics. The results confirm our assumption that students on both sides of the cutoff do not differ in observable characteristics. Particularly important for our design is that students at the margin do not differ with respect to their prior ability measured by the high-school GPA.¹⁴

However, even though high school GPA is a very strong predictor of study success (Berens et al., 2018), it may not capture early subject-specific skills in economics. Therefore, we also investigate differences with respect to early university performance. In Figure 3a, we show that there is no difference in the average grade obtained by marginal students in compulsory exams in the first examination period of the first semester. A further concern regarding the validity of our results refers to an argument of Holmstrom and Milgrom (1991). Although students are forced to write orientation exams in the first examination period, they are free to choose their workload in addition to these two courses. Therefore, it could be that students who narrowly fail an orientation exam are just worse informed about their own ability and make mistakes in strategically planning the number of compulsory exams they take in the first examination period. We address this concern by graphically investigating the number of registered exams in

¹³A third potential concern might be that students can manipulate the cutoff ex post in the exam inspection, resulting in a potential selection on unobservables around the cutoff. We collect anecdotal evidence and investigate for a recent cohort whether this is likely to be the case. We do not find any evidence for this.

¹⁴The only variable for which we observe a significant effect in some specification is German. Therefore, we run robustness checks for our baseline results in which we exclude all non-german students to verify that this does not affect our results. Furthermore, we show the specification including control variables in Equation 3.

the first examination period for marginal students. As shown in Figure 3b, we find no significant difference at the cutoff.

From our analysis, we conclude that since strategic planning of workload, performance at university, initial ability, gender, and other characteristics cannot explain the observed discontinuity in the assignment variable, the resulting pattern is more likely to be a result of standard grading patterns and is not related to our treatment. This is confirmed by the results of our donut approach (Section 5).

5 Results

Our framework allows us to investigate how early exposure to a last-chance exam affects students' subsequent educational career. We present our estimation results as follows. Section 5.1 and Section 5.2 present results for our two main outcome variables: early dropout and graduation. In Section 5.3 we provide subsample analyses on important demographic characteristics. In Section 5.4 we discuss the specific role of high-stakes. In Section 5.5 we investigate other important educational outcomes namely GPA, time to degree, and performance in follow-up courses and semesters.

5.1 Early dropout

Figure 5 plots a binary variable that is equal to 1 if a student drops out at the end of the first semester as a function of the performance in the first attempt in orientation exams. We obtain two main results: First, the probability of dropping out early decreases in performance in the orientation exams. Second, we observe a clear discontinuity in the probability of dropping out in the first semester at the cutoff. Although only 5% of the students who avoid the last chance exam drop out at this stage, around 18% of students who have to write a retake exam drop out in the first semester. In Table 7 we confirm this result. We show results from two different specifications. In columns 1-3 we show results from a non-parametric approach using local linear regressions on both sides of the cutoff and an optimal bandwidth calculated according to [Imbens and Kalyanaraman \(2012\)](#). In Columns 4-6 we show results from a parametric specification estimating Equation 3 using a second-order polynomial of our running variable and allowing for different slopes on both sides of the cutoff. We find a significant increase in the probability of

dropping out at the end of the first semester for marginal students on probation. The effect size is around 11 to 13 percentage points across different specifications, which translates in a more than doubled risk of early dropout. Adding cohort fixed-effects and a set of student demographic characteristics does not affect the result.

We perform several checks to verify the robustness of this result. First, we assess the sensitivity of the result obtained with respect to the chosen bandwidth. Therefore, in Panel A of Table 8 we show that using an alternative procedure for optimal bandwidth selection does not affect the observed result (Calonico et al., 2014). The next concern might be the pattern observed in Section 4, where we have shown evidence for a non-continuous pattern of the running variable around the promotion threshold. We have argued that this pattern likely reflects a mechanical pattern when correcting exams and provided evidence that students around the cutoff do not differ significantly in important characteristics. However, to assess the robustness of our results, we present the results from two donut approaches, in which we drop students exactly at the cutoff (within a very narrow window around the cutoff). Panel B and C in Table 8 confirm our main results. Panel D of Table 8 shows that our results are not affected when restricting our sample to German students. A final concern might be the construction of our assignment variable. In our main analysis, we use the performance in both orientation exams to determine treatment status and students are "treated" when they fail at least one orientation exam. A concern might thus be that our obtained results are not the effect of failing one exam and therefore being exposed to one last-chance exam, but in some cases the result of being exposed to two last-chance exams. We show that this concern does not affect our results by showing in Panel E of Table 8 that the results are quantitatively and qualitatively unaffected when restricting the sample to those students who fail at most in one exam.

In the remainder of this section, we dive deeper into the observed dropout decisions. Early dropout can be the result of two different decisions. First, students can decide to leave the program without being expelled (*voluntary dropout*). This can be the case for treated and untreated students and would allow them to continue with the same subject at another German university. While this might constitute a strategic choice, it could, however, also be the result of discouragement effects caused by probation. Second, students can be expelled from the program due to repeated failure to pass the orientation exam (*forced dropout*). *Forced dropout* is only possible for students on probation at this stage. In Table 9 we shed light on the two different forces

behind our baseline results. Our data allows us to investigate whether students leave the program voluntarily in response to being placed on probation or whether they are forced to leave due to repeated failure to meet the performance requirement. Voluntary dropout decisions do not explain the estimated increase in early dropout. We find that the coefficient is positive in all specifications but not statistically significant. Instead, the increase in early dropout is caused by students being expelled from the program after repeatedly failing to pass the orientation exam. The graphical illustration in Figure 6 and Figure 7 confirm this result.

5.2 Graduation

In the last section, we have shown that early exposure to a last-chance exam increases early dropout. However, the crucial question is whether this increase in early dropout has costly consequences for graduation chances or reduces *late(r) dropout*, and whether probation affects other educational outcomes that can affect labor market performance. Opponents argue that academic probation constitutes an additional barrier to graduate and especially the high-stakes at a very early phase may harm long run success (Credé and Kuncel, 2008; Robbins et al., 2004). However, probation can also be an important signal and encourage students to increase their level of effort and their skills, eventually increasing graduation chances (Jacob, 2005).

Figure 8 plots a dummy variable equal to 1 if a student graduates against the assignment variable. We observe a strong positive correlation between the assignment variable and the probability to graduate from the program throughout the whole distribution which is in line with the high predictive power of the orientation exams for program success. Furthermore, we observe a small discontinuous jump in the outcome variable at the cutoff. Table 10 confirms the graphical illustration in Figure 8. Comparable to Section 5.1, we show results from two different specifications. In Columns 1-3 we show results from a non-parametric approach using local linear regressions on both sides of the cutoff and an optimal bandwidth calculated according to Imbens and Kalyanaraman (2012). In columns 4-6 we show results from a parametric specification estimating Equation 3 using a second-order polynomial of our running variable and allowing for different slopes on both sides of the cutoff. Being on probation decreases graduation chances by 4 to 6 percentage points. However, the results are not statistically significant. Adding cohort fixed-effects and control variables further reduces the coefficient size slightly. In sum, we do not find evidence that graduation chances are significantly negatively affected. To assess the robust-

ness of this result, we conduct the same set of sensitivity analyses as explained in Section 5.1. We summarize the results in Table 11. All sensitivity analysis confirm our main result.

5.3 Heterogeneity

Even though we do find that students graduation chances are on average unaffected by exposure to a last-chance exam, this finding may mask substantial heterogeneity. In this section, we investigate whether our results are driven by particular subgroups of students. We investigate heterogeneity by gender, as previous research suggests substantial heterogeneity by gender in high-stake environments (Niederle and Vesterlund, 2007). Furthermore, we investigate heterogeneity with regard to the high-school GPA as a measure of initial ability and college preparedness, and finally in the number of last-chance exams a student is exposed to.

Exposure to a last-chance exam can have heterogeneous effects for male and female students. Differences between sexes can emerge for at least three reasons: First, the literature on behavioral economics shows that female students perform worse in high-stakes situations (Buser et al., 2014; Niederle and Vesterlund, 2007; Niederle and Yestrumskas, 2008). Second, male and female students have different ex-ante beliefs about graduation chances and therefore may respond differently to new information about their graduation chances (Zafar, 2011). Third, male and female students may differ in their behavior after observing a negative performance signal (Buser and Yuan, 2019).

Columns 1 and 2 of Table 12 show our point estimates for male and female students. In Panel A we show the results for early dropout. Both male and female students are more likely to drop out early when being exposed to a last-chance exam. The point estimates are almost identical, even though the standard errors slightly differ. In Panel B we show the results for Graduation. Although the coefficient for male students is negative, but close to zero, we find that female students are significantly less likely to graduate when being exposed to the last-chance exam. This result is in line with previous results from the psychological literature, documenting that female students perceive test taking as more stressful when the stakes are high (Leiner et al., 2018) and with studies showing that female students' performance decreases when the stakes increase compared to male students' performance (Azmat et al., 2016).

Columns 3 and 4 of Table 12 show results for students with above and below median high-school GPA. The high-school GPA is the best available measure for initial ability of students and

the strongest initial predictor for study success and thus also correlates with the performance in orientation exams. We observe twice as many students with below median high-school GPA close to the cutoff. For above median ability students, we do not observe a significant increase in early dropout. In contrast, the risk to dropout early for below-median students increases by 16 percentage points when being exposed to a last-chance exam. In Panel B we also show that low ability students being exposed to a last-chance exam face a 9 percentage point penalty in their graduation chances, when being exposed to the last-chance exam. The coefficient is around half the size for high-ability students.

Finally, we separately estimate the effect of exposure to a first and a second last-chance exam. We present the results in Columns 5 and 6 of Table 12. The observed effect for exposure to a second last-chance exam is almost three times as large as the effect of exposure to one last-chance exam and highly significant. This also holds for the negative effect observed on graduation chances. Among students who are exposed to at least one last-chance exam, exposure to an additional last-chance exam substantially hurts graduation chances by 17 percentage points. One potential reason for this is that marginal students in this specification have a lower initial ability and therefore face a higher risk of failing the retake exam.

5.4 Mechanisms

In the previous sections, we have shown that exposure to a last-chance exam increases the probability to dropout early. In contrast, we find limited evidence for a negative average effect on graduation chances which masks substantial effect heterogeneity. In this section, we discuss the mechanism of exposure to a high-stake situation as potential explanation behind our results in more detail.

Two potential mechanisms may drive our results from the previous section. First, students fail an exam and have to take the retake exam. This induces a 'natural barrier' to graduation, since they have to pass one exam more to graduate compared to students who meet the promotion cutoff. Second, the probation policy introduces high-stakes, since students are forced to retake the exam before they can proceed to the next semester, and repeated failure leads to immediate expulsion from the program. In this section, we attempt to isolate the effect of the high-stakes associated with the orientation exams from the effect of *having to pass an additional exam*.

To isolate the high-stakes introduced by the probation policy, we would ideally compare the

effect of exposure to a last-chance exam to the effect of having to repeat the exam without high-stakes. Since this is impossible because of insufficient data availability, we use information on non-orientation exams in the first semester. We therefore collect for all students from our baseline sample, the number of points achieved in all non-orientation exams in the first semester and calculate the running variable T_i in the same way as described in Section 3. We then merge this information with the administrative data. Thus, we have two observations for each student: one indicating performance in orientation exams, and one indicating performance in other compulsory exams in the first attempt. We also add a binary variable HS_i indicating whether the running variable was calculated using orientation or non-orientation exams.

We estimate the difference between the effect of having to retake an exam and the effect of having to retake an exam under high pressure. We implement this by estimating the boundary points of four regression functions of Y_{it} on T_i , two on both sides of T_c , for orientation exams and non-orientation exams. We use the following version of a difference-in-discontinuity equation:

$$Y_{ic} = \alpha_0 + \alpha_1 T_{ic} + D_{ic}(\beta_0 + \beta_1 T_{ic}) + HS_{it}(\gamma_0 + \gamma_1 T_{ic} + \gamma_2 D_{ic}) + \nu_c + \eta_{ic} \quad (4)$$

where D_i is a dummy that indicates whether a student failed an exam. HS_i is an indicator for orientation exams and $T = T_i - T^*$ is the normalized performance in the first attempts of the exams. Standard errors are clustered at the level of the running variable. The coefficient γ_2 is the difference-in-discontinuity estimator and identifies the difference in the outcome variable between failing a high-stake and failing a low-stake exam.¹⁵

In Table 13 we show the results of estimating Equation 4. We observe a strong effect of early pressure on early dropout as a result of the examination regulations. In contrast, even though the coefficients are negative, we do not find evidence for a negative impact of high-stakes on graduation chances. Thus, the high pressure introduced by the policy does, on average, not constitute an additional barrier to graduation. Adding controls and cohort fixed-effects confirms our results.

Next, we more carefully investigate the observed effect heterogeneity from Section 5.3. In Table 14 we investigate whether the high-stakes drive the observed gender difference in gradua-

¹⁵To test the validity of our identification strategy, we show that there is no difference in potential manipulative sorting around the cutoff between both types of exam (Figure 9). We observe a discontinuity in the density at the cutoff for both types of exams of similar magnitude. This is reassuring given our arguments in Section 4, because it does not show any evidence of specific sorting around the promotion cutoff in high-stake exams.

tion chances as the result of exposure to a last-chance exam. We estimate equation 4 separately for male and female students. The effect on early dropout is positive and significant for male and female students. However, the effect of high-stakes on graduation chances differs. Although the effect is slightly positive, even though statistically insignificant for male students, we observe a meaningful negative effect for female students, which is in line with previous evidence and therefore cannot be explained by exposure to an additional exam. This strengthens the interpretation that the observed effect can be associated with the high-stake situation.

In Table 15 we investigate the role of high-stakes for students with different ability levels. In Section 5.3 we have shown that students with low ability more likely to drop out and face reduced graduation chances, which is not the case for high-ability students. Our difference-in-discontinuity estimates confirm this pattern for early dropout. The high-stakes substantially increase the risk of early dropout for low ability students, while high ability students are almost unaffected. However, unlike the previous section, we find that the negative effect on graduation chances is not particularly related to the high-stakes attached to the probation policy. Neither low- nor high-ability students are differentially affected from exposure to an additional exam with or without high-stakes attached.

5.5 Further Results

So far, we only focused on effects related to dropout and graduation. Exposure to a last-chance exam may, however, also impact other important outcomes. We investigate two different types of outcomes. First, we look at study pace, and second, we study further performance outcomes, such as GPA and performance in follow-up courses.

In contrast to grade retention, this form of academic probation does not automatically cause a delay of study progress. Students who successfully pass the retake exam continue to study with their cohort. However, exposure to a last-chance exam may have an effect on study pace. It can act as a warning signal that allows students to receive more information about their graduation chances and adjust their workload accordingly to maximize success probabilities.

Table 16 shows that students on probation are less likely to finish the program in time (3 years).¹⁶ But this effect is driven by a small share of students, since less than 30 percent of

¹⁶We observe all students for at least 7 years, which is more than twice the program length. Therefore, censoring in 2019 should not invalidate our results.

graduates, graduate in time.

Exposure to a last-chance exam may affect students' follow-up performance for several reasons. First, it enforces retaking the course material. This may improve skills and result in better grades in later semesters. Second, it may work as a warning signal and increase effort provision to avoid bad grades and probation in later semesters. Third, failing in an important exam may demotivate students and harm follow-up performance (Rosenqvist and Skans, 2015).

Related to the idea that probation may be a useful warning signal and help students to adjust their working efforts, we do not find any positive effects on the probability to avoid another probation period in the second semester (Panel B), neither do we find a positive effect on the final GPA (Panel C).

The final GPA is a composite measure of performance over several years and may thus be too noisy to observe whether probation has, in fact, a positive immediate impact on performance. Therefore, as a second exercise, we investigate the effect of probation on performance in a direct follow-up exam. We only have one orientation exam that has a clear follow-up in the next semester. Thus, we compare the performance in Math II of students close to the cutoff in the first attempt of the Math I exam. In Panel D of Table 16 we show that students who have to retake the exam do not perform significantly better in the follow-up course. The point estimate is negative (which indicates a better performance) in all specifications but is not statistically different from zero.¹⁷

6 Conclusion

Many students have inaccurate perceptions about their chances of graduating when they begin their higher education career (Stinebrickner and Stinebrickner, 2012; Zafar, 2011). This paper investigates the effect of exposure to a last-chance exam in an early phase of a higher education program on subsequent student performance. We exploit a probation policy that forces low-performing students to attend a last-chance exam before being allowed to proceed. Our study offers interesting insights on an alternative mechanism of academic probation, and is therefore directly relevant for public policy.

First, in line with general concerns about probation policies and high-stake testing, we doc-

¹⁷Note also that we do not account here for a potentially positive selection of students in the treatment group and even though do not find a positive effect.

ument an increase in the number of early dropouts after the policy was introduced. Second, we use administrative data from six cohorts of first-year students to study their performance after exposure to the exam. We address the endogeneity of treatment status using a regression discontinuity design, that allows us to estimate the causal effects of exposure to a the last-chance exam on early dropout rates, graduation chances, and later academic performance. Third, we separate the impact of high-stakes testing from the effect of failing an exam in the first semester by using detailed student performance data.

We find that early exposure to a last-chance exam significantly increases the number of early dropouts. The risk of dropping out more than doubles for students on probation compared to the counterfactual group of students. However, probation has only a very small negative effects on graduation chances. Being on probation reduces graduation chances by around 5 percentage points. This effect is not significant across specifications. However, we observe a significant decrease in graduation chances for mainly two groups of students: Female students and low-ability students. We show that the negative effect on graduation for female students is driven by the exposure to a high-stake situation rather than the barrier of having to write an additional exam.

Our study adds to the growing literature on policies aimed at helping low-performing students in higher education. We find that last-chance exams early in a program can reduce the costs of inevitable dropout decisions without lowering graduation rates. This is important for two reasons. First, the use of probation policies is becoming more widespread. Second, last-chance exams have low implementation costs, making them attractive to policy makers. Our findings suggest that last-chance exams could help reduce the negative effects of procrastination and misinformation about graduation chances, which are common, particularly among low-performing students (DellaVigna, 2009; Oreopoulos and Petronijevic, 2013; Kim and Seo, 2015). However, policymakers should be aware that the effects can vary between different demographic groups, especially in high-stakes situations.

References

- AERA (1999): *Standards for educational and psychological testing*, Amer Educational Research Assn.
- ALTBACH, P. G., L. REISBERG, AND L. E. RUMBLEY (2019): *Trends in global higher education: Tracking an academic revolution*, vol. 22, Brill.
- ALTONJI, J., E. BLOM, AND C. MEGHIR (2011): “Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers,” *Annual Review of Economics*, 4, 185–223.
- ARCIDIACONO, P. (2004): “Ability sorting and the returns to college major,” *Journal of Econometrics*, 121, 343–375.
- AZMAT, G., C. CALSAMIGLIA, AND N. IRIBERRI (2016): “Gender differences in response to big stakes,” *Journal of the European Economic Association*, 14, 1372–1400.
- BERENS, J., K. SCHNEIDER, S. GOERTZ, S. OSTER, AND J. BURGHOFF (2018): “Early Detection of Students at Risk – Predicting Student Dropout Using Administrative Student Data and Machine Learning Methods,” *CESifo Working Paper*.
- BETTINGER, E. P. AND B. T. LONG (2009): “Addressing the Needs of Underprepared Students in Higher Education: Does College Remediation Work?” *The Journal of Human Resources*, 44, 736–771.
- BIETENBECK, J., A. LEIBING, J. MARCUS, AND F. WEINHARDT (2023): “Tuition fees and educational attainment,” *European Economic Review*, 154, 104431.
- BÖLKE, L., C. EISELSTEIN, S. FAISST, V. M. HAUG, K. HERBERGER, A. KALOUS, H. MESSER, A. PAUTSCH, G. SANDBERGER, K. SCHILLER, ET AL. (2020): *Das Hochschulrecht in Baden-Württemberg: Systematische Darstellung*, CF Müller GmbH.
- BOUND, J., M. F. LOVENHEIM, AND S. TURNER (2010): “Why Have College Completion Rates Declined? An Analysis of Changing Student Preparation and Collegiate Resources,” *American Economic Journal: Applied Economics*, 2, 129–57.
- BOWEN, W., M. CHINGOS, AND M. MCPHERSON (2009): *Crossing the Finish Line: Completing College at America’s Public Universities*, Princeton University Press.
- BRATTI, M., S. GRANATO, AND E. HAVARI (2024): “Another Chance: Number of Exam Retakes and University Students’ Outcomes,” Tech. rep., IZA Discussion Papers.
- BUSER, T., M. NIEDERLE, AND H. OOSTERBEEK (2014): “Gender, Competitiveness, and Career Choices,” *The Quarterly Journal of Economics*, 129, 1409–1447.
- BUSER, T. AND H. YUAN (2019): “Do Women Give Up Competing More Easily? Evidence from the Lab and the Dutch Math Olympiad,” *American Economic Journal: Applied Economics*, 11, 225–52.

- CALONICO, S., M. D. CATTANEO, AND R. TITIUNIK (2014): “Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs,” *Econometrica*, 82, 2295–2326.
- CAMERON, A. C., J. B. GELBACH, AND D. L. MILLER (2011): “Robust inference with multiway clustering,” *Journal of Business & Economic Statistics*, 29, 238–249.
- CASEY, M. D., J. CLINE, B. OST, AND J. A. QURESHI (2018): “Academic probation, student performance and strategic course taking,” *Economic Inquiry*, 56, 1646–1677.
- CATTANEO, M. D., M. JANSSON, AND X. MA (2018): “Manipulation testing based on density discontinuity,” *The Stata Journal*, 18, 234–261.
- (2020): “Simple local polynomial density estimators,” *Journal of the American Statistical Association*, 115, 1449–1455.
- (2024): “Local regression distribution estimators,” *Journal of Econometrics*, 240, 105074.
- CREDÉ, M. AND N. R. KUNCEL (2008): “Study Habits, Skills, and Attitudes: The Third Pillar Supporting Collegiate Academic Performance,” *Perspectives on Psychological Science*, 3, 425–453.
- DEE, T. S., W. DOBBIE, B. A. JACOB, AND J. ROCKOFF (2019): “The causes and consequences of test score manipulation: Evidence from the new york regents examinations,” *American Economic Journal: Applied Economics*, 11, 382–423.
- DELLAVIGNA, S. (2009): “Psychology and Economics: Evidence from the Field,” *Journal of Economic Literature*, 47, 315–72.
- DIAMOND, R. AND P. PERSSON (2016): “The long-term consequences of teacher discretion in grading of high-stakes tests,” Tech. rep., National Bureau of Economic Research.
- DUCHINI, E. (2017): “Is college remedial education a worthy investment? New evidence from a sharp regression discontinuity design,” *Economics of Education Review*, 60, 36 – 53.
- FLETCHER, J. M. AND M. TOKMOULINE (2017): “The Effects of Academic Probation on College Success: Regression Discontinuity Evidence from Four Texas Universities,” IZA Discussion Papers 11232, Institute of Labor Economics (IZA).
- FRISANCHO, V., K. KRISHNA, S. LYCHAGIN, AND C. YAVAS (2016): “Better luck next time: Learning through retaking,” *Journal of Economic Behavior & Organization*, 125, 120–135.
- GARIBALDI, P., F. GIAVAZZI, A. ICHINO, AND E. RETTORE (2012): “College cost and time to complete a degree: Evidence from tuition discontinuities,” *The Review of Economics and Statistics*, 94, 699–711.
- GNEEZY, U., M. NIEDERLE, AND A. RUSTICHINI (2003): “Performance in Competitive Environments: Gender Differences*,” *The Quarterly Journal of Economics*, 118, 1049–1074.

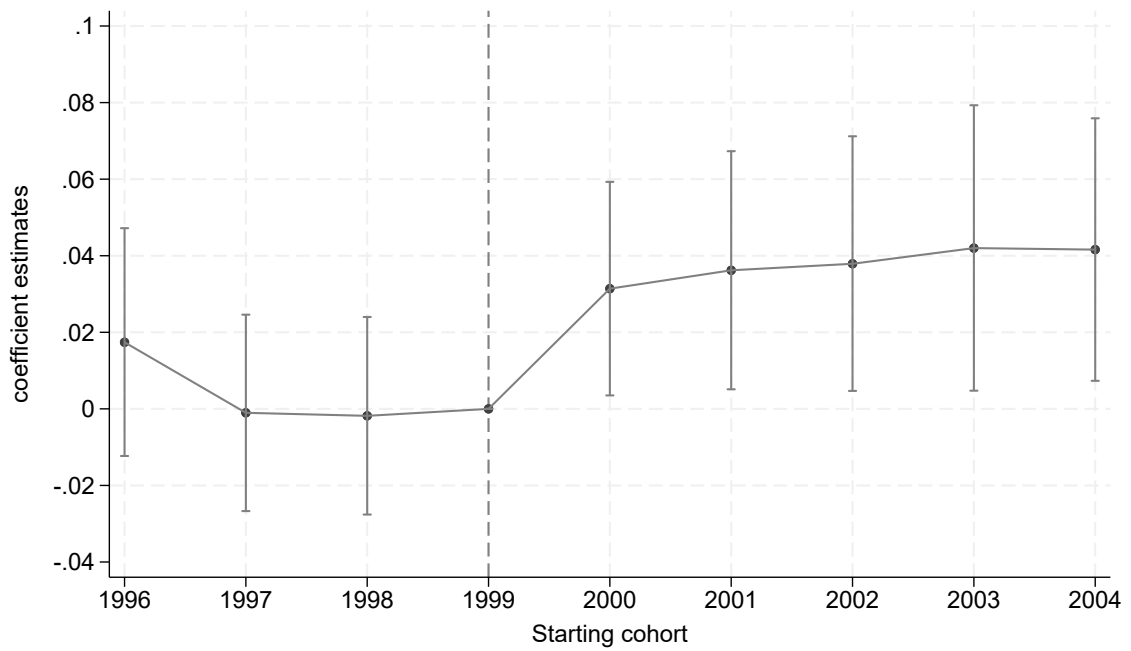
- GOODMAN, J., O. GURANTZ, AND J. SMITH (2020): “Take Two! SAT Retaking and College Enrollment Gaps,” *American Economic Journal: Economic Policy*, 12, 115–58.
- HANUSHEK, E. A. AND M. E. RAYMOND (2004): “The effect of school accountability systems on the level and distribution of student achievement,” *Journal of the European Economic Association*, 2, 406–415.
- HOLMSTROM, B. AND P. MILGROM (1991): “Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design,” *Journal of Law, Economics, and Organization*, 7, 24–52.
- IMBENS, G. AND K. KALYANARAMAN (2012): “Optimal Bandwidth Choice for the Regression Discontinuity Estimator,” *The Review of Economic Studies*, 79, 933–959.
- IMBENS, G. W. AND T. LEMIEUX (2008): “Regression discontinuity designs: A guide to practice,” *Journal of Econometrics*, 142, 615 – 635.
- JACOB, B. A. (2005): “Accountability, Incentives and Behavior: The Impact of High-Stakes Testing in the Chicago Public Schools,” *Journal of Public Economics*, 89, 761–796.
- JACOB, B. A. AND L. LEFGREN (2004): “Remedial Education and Student Achievement: A Regression-Discontinuity Analysis,” *The Review of Economics and Statistics*, 86, 226–244.
- KETEL, N., J. LINDE, H. OOSTERBEEK, AND B. VAN DER KLAUW (2016): “Tuition Fees and Sunk-cost Effects,” *The Economic Journal*, 126, 2342–2362.
- KIM, K. R. AND E. H. SEO (2015): “The relationship between procrastination and academic performance: A meta-analysis,” *Personality and Individual Differences*, 82, 26–33.
- LEE, D. S. AND T. LEMIEUX (2010): “Regression Discontinuity Designs in Economics,” *Journal of Economic Literature*, 48, 281–355.
- LEINER, J. E., T. SCHERNDL, AND T. M. ORTNER (2018): “How do men and women perceive a high-stakes test situation?” *Frontiers in Psychology*, 9, 2216.
- LINDO, J. M., N. J. SANDERS, AND P. OREOPOULOS (2010): “Ability, Gender, and Performance Standards: Evidence from Academic Probation,” *American Economic Journal: Applied Economics*, 2, 95–117.
- MACHIN, S., S. McNALLY, AND J. RUIZ-VALENZUELA (2020): “Entry through the narrow door: The costs of just failing high stakes exams,” *Journal of Public Economics*, 190, 104224.
- MALACRINO, D., S. NOCITO, AND R. SAGGIO (2024): “Do Reforms Aimed at Reducing Time to Graduation Work? Evidence from the Italian Higher Education System,” Working Paper 32659, National Bureau of Economic Research.
- MARTORELL, P. AND I. MCFARLIN (2011): “Help or Hindrance? The Effects of College Remediation on Academic and Labor Market Outcomes,” *The Review of Economics and Statistics*, 93, 436–454.

- MONTALBÁN, J. (2022): “Countering Moral Hazard in Higher Education: The Role of Performance Incentives in Need-Based Grants,” *The Economic Journal*, 133, 355–389.
- MONTOLIO, D. AND P. A. TABERNER (2021): “Gender differences under test pressure and their impact on academic performance: a quasi-experimental design,” *Journal of Economic Behavior & Organization*, 191, 1065–1090.
- NIEDERLE, M. AND L. VESTERLUND (2007): “Do Women Shy Away From Competition? Do Men Compete Too Much?,” *The Quarterly Journal of Economics*, 122, 1067–1101.
- (2010): “Explaining the Gender Gap in Math Test Scores: The Role of Competition,” *Journal of Economic Perspectives*, 24, 129–44.
- NIEDERLE, M. AND A. YESTRUMSKAS (2008): “Gender Differences in Seeking Challenges: The Role of Institutions,” *National Bureau of Economic Research, Inc, NBER Working Papers*.
- NIJENKAMP, R., M. R. NIEUWENSTEIN, R. DE JONG, AND M. M. LORIST (2022): “Second chances in learning: Does a resit prospect lower study-time investments on a first test?” *Journal of Cognition*, 5.
- OREOPOULOS, P. (2021): “What limits college success? a review and further analysis of holzer and baum’s making college work,” *Journal of Economic Literature*, 59, 546–573.
- OREOPOULOS, P. AND U. PETRONIJEVIC (2013): “Making College Worth It: A Review of the Returns to Higher Education,” *The Future of Children*, 23.
- OST, B., W. PAN, AND D. WEBBER (2018): “The Returns to College Persistence for Marginal Students: Regression Discontinuity Evidence from University Dismissal Policies,” *Journal of Labor Economics*, 36, 779 – 805.
- RDC (2017): “Statistik der Studenten (1995–2014),” .
- ROBBINS, S., K. LAUVER, AND D. HUY (2004): “Do Psychological and Study Skill Factors Predict College Outcomes? A Meta-Analysis,” *Psychological Bulletin*, 130, 261 – 288.
- ROODMAN, D., M. Ø. NIELSEN, J. G. MACKINNON, AND M. D. WEBB (2019): “Fast and wild: Bootstrap inference in Stata using boottest,” *The Stata Journal*, 19, 4–60.
- ROSENQVIST, O. AND O. N. SKANS (2015): “Confidence enhanced performance?—The causal effects of success on future performance in professional golf tournaments,” *Journal of Economic Behavior & Organization*, 117, 281–295.
- SCOTT-CLAYTON, J. (2011): “On money and motivation: A quasi-experimental analysis of financial incentives for college achievement,” *Journal of Human Resources*, 46, 614–646.
- SHRIBERG, D. AND A. B. SHRIBERG (2006): “High-stakes testing and dropout rates,” *Dissent*, 53, 76–80.

- STINEBRICKNER, R. AND T. STINEBRICKNER (2014): “Academic Performance and College Dropout: Using Longitudinal Expectations Data to Estimate a Learning Model,” *Journal of Labor Economics*, 32, 601–644.
- STINEBRICKNER, T. AND R. STINEBRICKNER (2012): “Learning about Academic Ability and the College Dropout Decision,” *Journal of Labor Economics*, 30, 707–748.
- TAFRESCHI, D. AND P. THIEMANN (2016): “Doing it twice, getting it right? The effects of grade retention and course repetition in higher education,” *Economics of Education Review*, 55, 198–219.
- UNESCO (2022): *Higher education global data report (Summary). A contribution to the World Higher Education Conference 18-20 May 2022.*
- VOSSENSTEYN, J. J., A. KOTTMANN, B. W. JONGBLOED, F. KAISER, L. CREMONINI, B. STENSAKER, E. HOVDHAUGEN, AND S. WOLLSCHIED (2015): “Dropout and completion in higher education in Europe: Main report,” .
- ZAFAR, B. (2011): “How Do College Students Form Expectations?” *Journal of Labor Economics*, 29, 301 – 348.

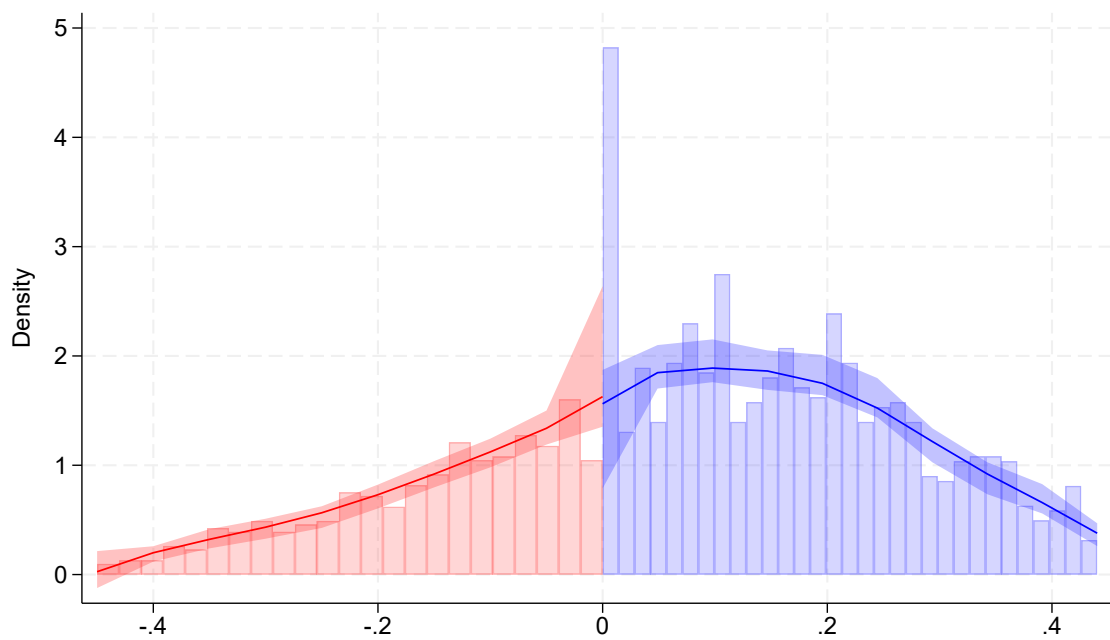
TABLES AND FIGURES

Figure 1:
Difference-in-differences event study



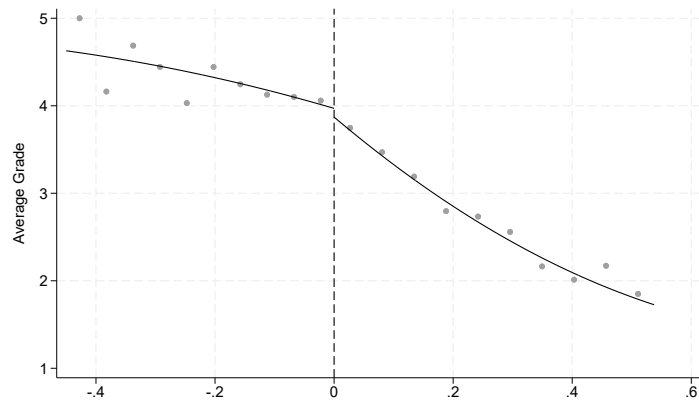
Note: The figure plots point estimates and 90 percent confidence intervals. The point estimates reflect the cohort-specific effects of the introduction of orientation exams on early dropout (within 2 years). The omitted baseline year is 1999, which is the last pre-treatment period.

Figure 2:
Manipulation test

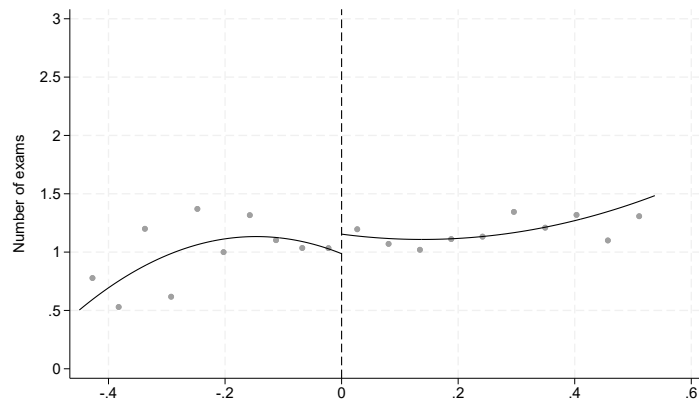


Note: This figure provides graphical evidence for a manipulation check of the assignment variable following an approach by [Cattaneo et al. \(2020, 2024\)](#). The figure is created with the help of the user written Stata command `rddensity` ([Cattaneo et al., 2018](#)).

Figure 3:
RDD: First year performance



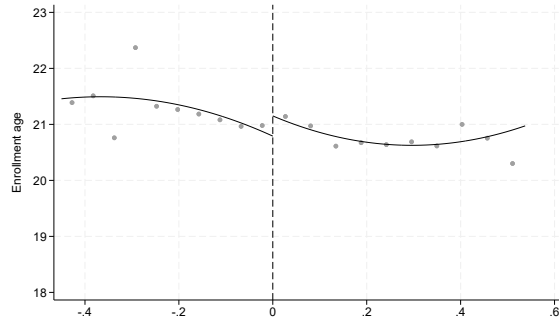
(a) Performance



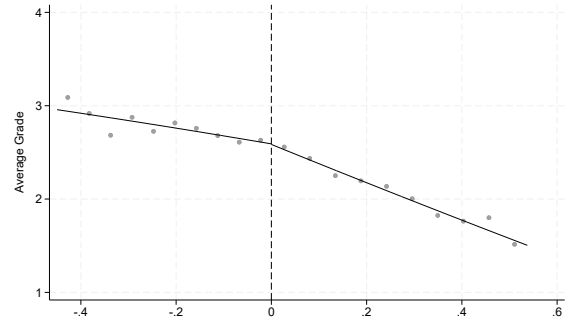
(b) Number of exams

Note: The figure investigates the smoothness of first year performance at the cutoff. The dots indicate means per bin of equal size. The solid lines present a quadratic fit to both sides of the threshold. The upper graph investigates discontinuities in the first semester GPA in non-orientation exams. The lower graph shows the relationship between the number of exams that students chose to attend in the first semester at the first attempt and the assignment variable.

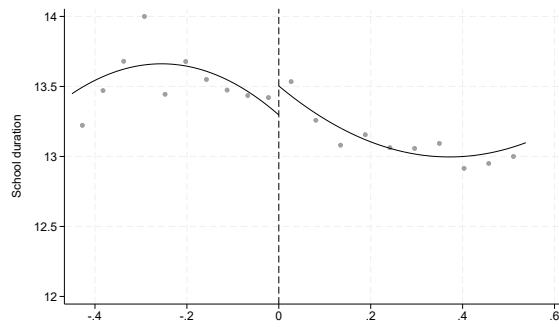
Figure 4:
RDD: Background characteristics



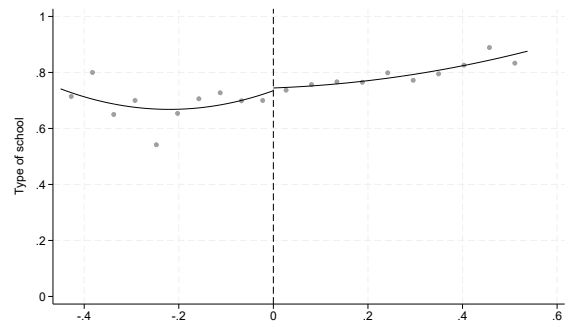
(a) Enrollment age



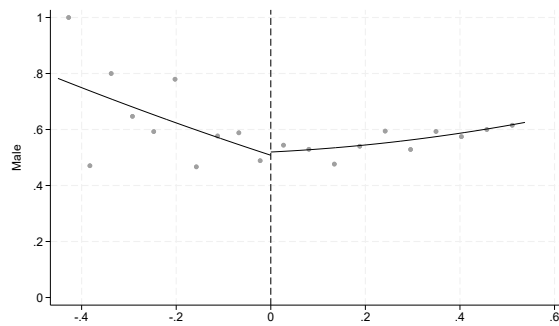
(b) Ability



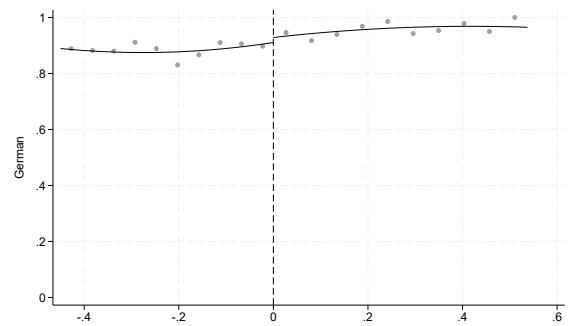
(c) Years attended school



(d) Type of school



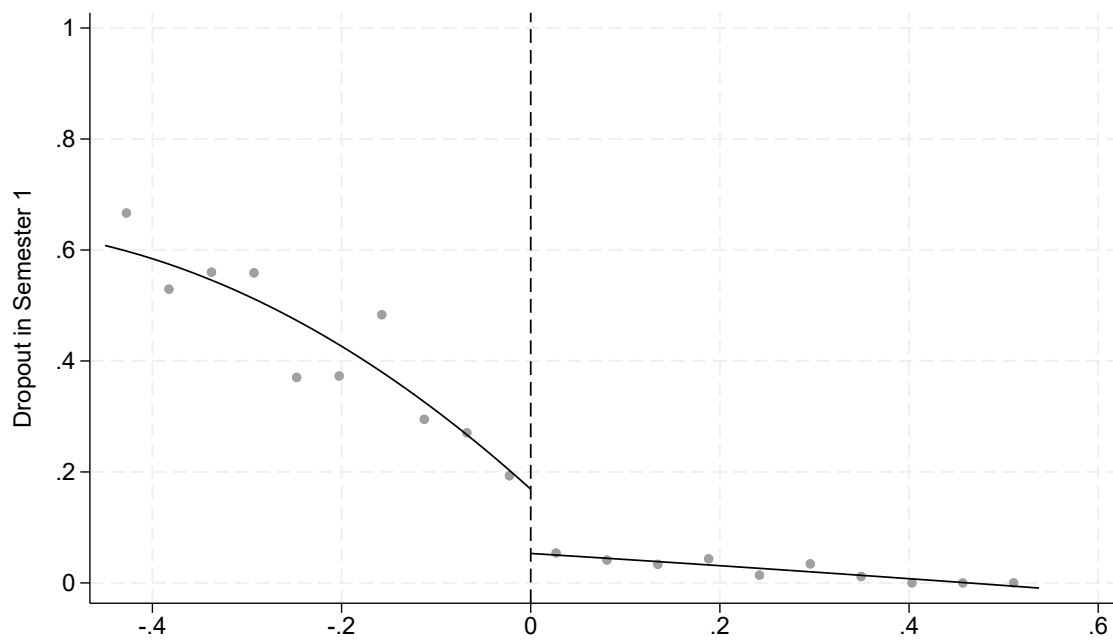
(e) Gender



(f) German

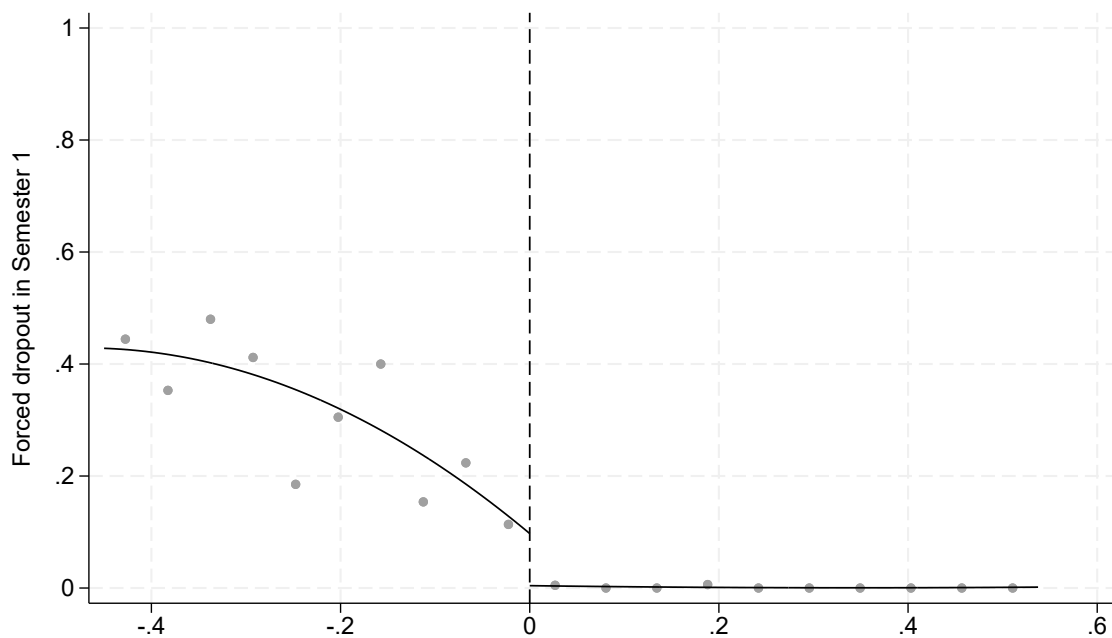
Note: The figures investigate the smoothness of background characteristics at the cutoff. The dots indicate means per bin of equal size. The solid lines present a quadratic fit to both sides of the threshold. Ability is defined by students high-school GPA. Type of school is a binary indicator that is equal to 1 if a student attended a Gymnasium.

Figure 5:
RDD Graph: Early Dropout



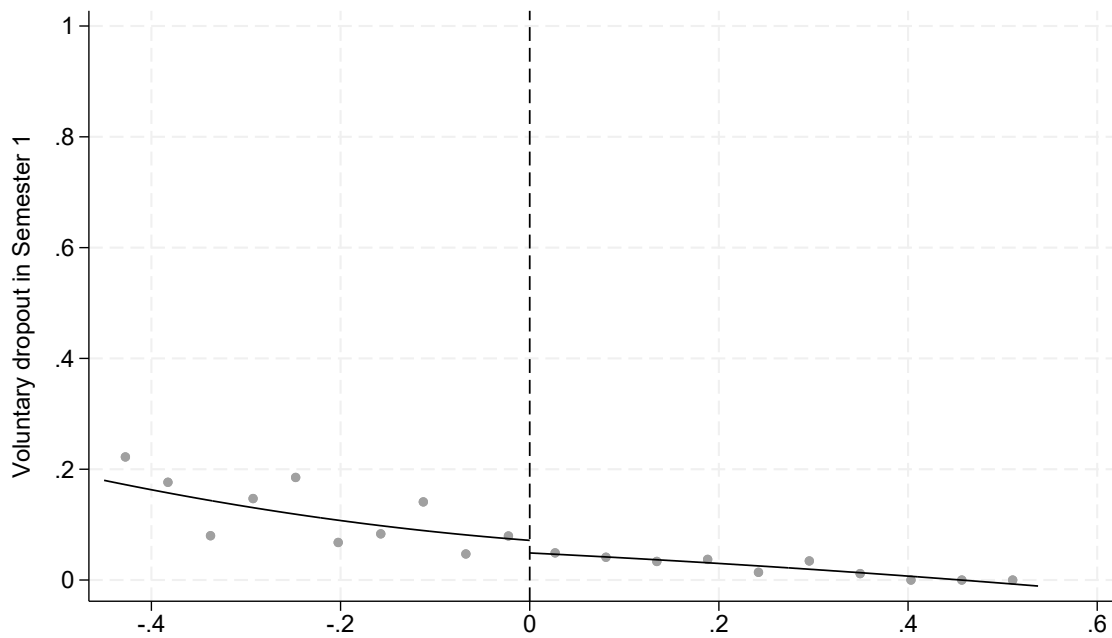
Note: RDD Graph: Early Dropout. The graph shows the effect of exposure to a last-chance exam on immediate dropout. Dots represent means per bin of equal size. The solid line represents a quadratic fit on both sides of the cutoff.

Figure 6:
RDD Graph: Forced Early Dropout



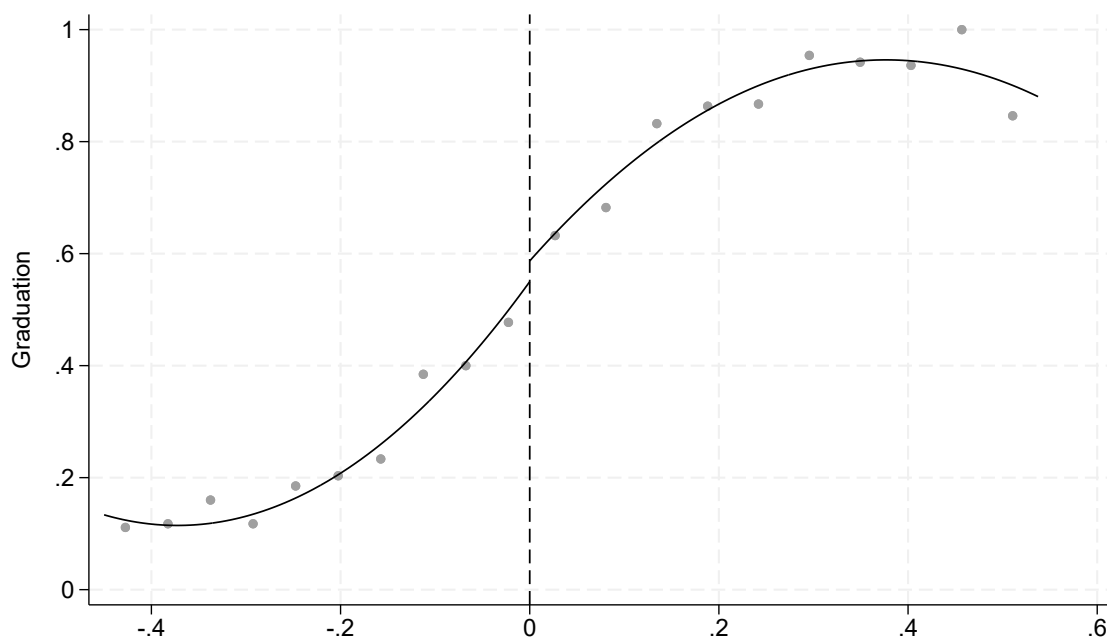
Note: RDD Graph: Forced Dropout. The graph shows the effect of exposure to a last-chance exam on immediate forced dropout. Dots represent means per bin of equal size. The solid line represents a quadratic fit on both sides of the cutoff.

Figure 7:
RDD Graph: Voluntary Early Dropout



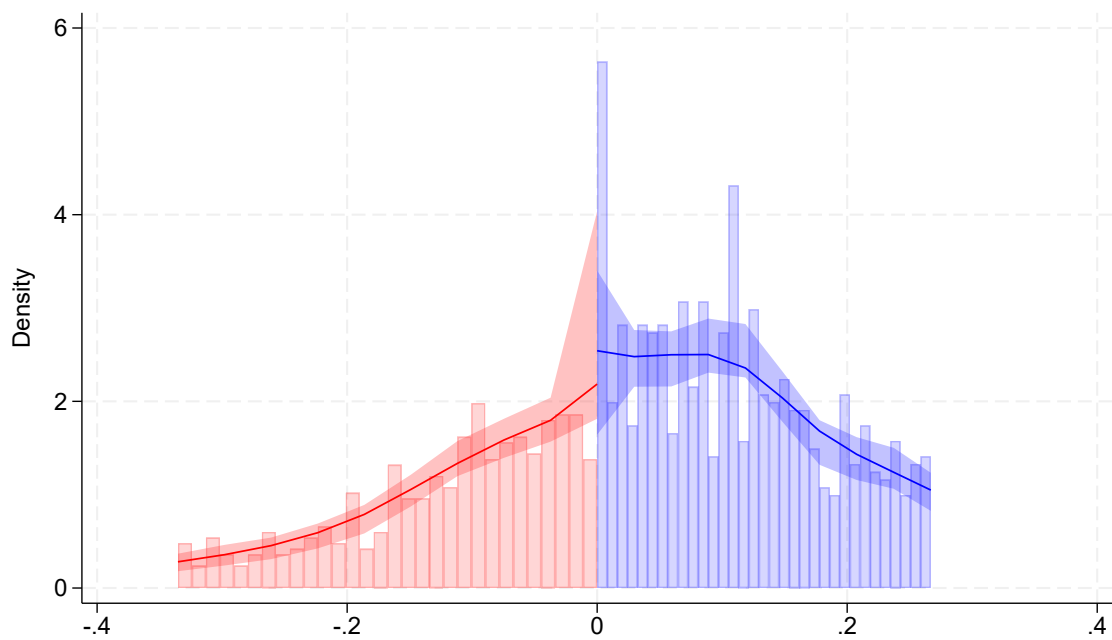
Note: RDD Graph: Voluntary Dropout. The graph shows the effect of exposure to a last-chance exam on immediate voluntary dropout. Dots represent means per bin of equal size. The solid line represents a quadratic fit on both sides of the cutoff.

Figure 8:
RDD Graph: Graduation



Note: RDD Graph: Graduation. The graph shows the effect of exposure to a last-chance exam on graduation. Dots represent means per bin of equal size. The solid line represents a quadratic fit on both sides of the cutoff.

Figure 9:
Manipulation test: Non-orientation exams



Note: This figure provides graphical evidence for a manipulation check of the assignment variable in low stake exams, following an approach by Cattaneo et al. (2020, 2024). The figure is created with the help of the user written Stata command *rddensity* (Cattaneo et al., 2018).

Table 1:
List of universities

University	University	University
University of Paderborn	Bauhaus University Weimar	University of Koblenz-Landau
University of Siegen	TU Ilmenau	University of Potsdam
University of Kassel	University der Bundeswehr München	University of Erlangen-Nürnberg
Europa University Viadrina	University of Vechta	University of München
Humboldt-University Berlin	University of Hildesheim	University of Würzburg
Brandenburgische TU Cottbus	University of Göttingen	University of Regensburg
University of Rostock	University of Bremen	University of Augsburg
University of Greifswald	University of Bochum	University of Saarbrücken
University of Halle	University of Bonn	FU Berlin
University of Magdeburg	University of Düsseldorf	TU Braunschweig
University of Leipzig	University of Cologne	TU Clausthal
TU Dresden	University of Münster	University of Hannover
TU Bergakademie Freiberg	University of Dortmund	RWTH Aachen
University of Jena	University of Bielefeld	TU Darmstadt
University of Bamberg	University of Duisburg-Essen	University of Karlsruhe
University of Bayreuth	University of Frankfurt	University of Stuttgart
University of Oldenburg	University of Gießen	TU München
University of Osnabrück	University of Marburg	TU Berlin
University of Passau	University of Trier	TU Chemnitz
University of Heidelberg	TU Kaiserslautern	University of Ingolstadt
University of Konstanz	University of Mainz	University of Hohenheim
University of Tübingen	University of Freiburg	University of Mannheim
University of Ulm	University of Lüneburg	University of Kiel
University of Lübeck	University of Flensburg	University of Wuppertal

Note: This table lists all universities that are part of the estimation sample in Section 2.

Table 2:
Difference-in-differences results

	(1)	(2)	(3)	(4)	(5)
Orientation exams	0.032** (0.012)	0.028** (0.012)	0.028** (0.012)	0.034*** (0.010)	0.034** (0.017)
University fixed effects	No	No	Yes	Yes	Yes
Cohort fixed effects	No	No	Yes	Yes	Yes
University controls	No	No	No	Yes	Yes
Group specific trends	No	No	No	No	Yes
Universities	72	72	72	72	72
Observation	288	648	648	648	648

Note: This table reports the results from different Difference-in-Difference and TWFE specifications. The outcome variable is the early dropout rate. Columns 1 and 2 show the results of estimating Equation 1. In Column 1 the sample is restricted to two periods before and after the introduction of the policy. Column 2 uses the full observation period. In Column 3 we replace the two dummy variables with university and cohort fixed effects. In Column 4 we add a set of control variables and in Column 5 we add group specific time trends. Standard errors are clustered at the university level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table 3:
Difference-in-differences: Leave one out

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Orientation exams	0.033*** (0.011)	0.030*** (0.010)	0.028*** (0.008)	0.036*** (0.011)	0.035*** (0.011)	0.040*** (0.010)	0.036*** (0.011)	0.036*** (0.011)	0.035*** (0.011)
Universities	71	71	71	71	71	71	71	71	71
Observation	639	639	639	639	639	639	639	639	639

Note: This table reports the results from TWFE specifications based on Column 4 of Table 2. In each column we drop one university from the treatment group. Standard errors are clustered at the university level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table 4:
Difference-in-differences: robustness checks

	(weighted)	(weighted dep.)	(outlier)	(First)	(Second)	(state)	(state*year)	(twoway)	(twoway+boot)
O-exams	0.031*** (0.009)	0.023** (0.011)	0.034*** (0.010)	0.010* (0.006)	0.032*** (0.010)	0.034*** (0.000)	0.034*** (0.000)	0.034*** (0.002)	0.034*** [0.005, 0.065]
Obs.	645	3123	603	645	645	645	645	645	645

Note: This table reports the results from TWFE specifications based on Column 4 of Table 2. If not otherwise stated, standard errors are clustered at the university level. In Column 1 we weight each cell by its size (Starting-cohort size). In Column 2, we conduct our analysis at the university department level instead of the university level and weight each cell by the department size. In Column 3 we exclude outlier with unusual dropout rates. Columns 4 and 5 show results using the first year dropout rate and the second year dropout rates as outcome variables. Columns 6-9 present the results using alternative inference methods. In Column 6, we cluster standard errors at the state level, in Column 7 at the state*year level and in Column 8 we implement a twoway clustering approach at the state and year level following [Cameron et al. \(2011\)](#). Confidence intervals in Column 9 are based on wild cluster bootstrap procedures with state and cohort groups as clusters. Wild cluster bootstrap specifications are estimated using the user-written Stata command *boottest* ([Roodman et al., 2019](#)). *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table 5:
Descriptive statistics

	Full Sample	Pass	Fail	Difference	p-value
<i>Panel A</i>					
Enrollment age	20.92	20.80	21.19	-0.38	0.00
School duration	13.30	13.19	13.53	-0.34	0.00
Male	0.56	0.54	0.60	-0.05	0.07
German	0.93	0.95	0.89	0.06	0.00
High School GPA	2.37	2.22	2.72	-0.49	0.00
School type	0.75	0.77	0.69	0.08	0.00
<i>Panel B</i>					
First semester exams (Number)	1.12	1.15	1.05	0.10	0.04
First semester Performance (Grade)	3.33	2.98	4.21	-1.23	0.00
<i>Panel C</i>					
Early dropout	0.13	0.03	0.36	-0.32	0.00
Graduation	0.65	0.81	0.31	0.50	0.00
Observations	1562	1080	482		

Note: This table shows descriptive statistics for the estimation sample. Column 1 shows means for the full sample, while columns 2 and 3 show means for the sub-samples of treated and untreated students. Column 4 displays the difference in mean characteristics and column 5 shows the p-value of the h_0 of no difference in mean characteristics. In panel B, the first row shows the number of exams students attend in the first examination period. In the second row, we report the average grade in non-orientation exams in the same examination period.

Table 6:
RDD estimates: Background characteristics

	(1)	(2)	(3)	(4)
German	-0.040 (0.025)	-0.034** (0.017)	-0.018 (0.025)	-0.012 (0.022)
Bandwidth	0.19	0.19	–	–
Observation	917	917	1562	1562
Men	-0.010 (0.051)	-0.010 (0.051)	-0.012 (0.071)	-0.012 (0.072)
Bandwidth	0.5	0.5	–	–
Observation	1557	1557	1562	1562
Enrolment age	-0.261 (0.224)	-0.245 (0.193)	-0.361 (0.237)	-0.364 (0.226)
Bandwidth	0.25	0.25	–	–
Observation	1159	1159	1553	1332
High-school GPA	0.003 (0.053)	-0.019 (0.053)	0.006 (0.057)	-0.020 (0.059)
Bandwidth	0.25	0.25	–	–
Observation	1154	1154	1562	1562
School type	-0.028 (0.031)	-0.035 (0.035)	-0.010 (0.039)	-0.013 (0.043)
Bandwidth	0.27	0.27	–	–
Observation	1105	1105	1413	1413
School duration	-0.134 (0.151)	-0.139 (0.138)	-0.207 (0.157)	-0.223 (0.155)
Bandwidth	0.25	0.25	–	–
Observation	1159	1159	1562	1562
Cohort FE	No	Yes	No	Yes

Note: RDD estimates using student background characteristics as outcome variables. The different columns show results for different specifications. In Columns 1 and 2, results are based on a non-parametric approach using local linear regressions on both sides of the cutoff and an optimal bandwidth calculated according to [Imbens and Kalyanaraman \(2012\)](#). In Columns 3 and 4, we use a parametric specification estimating Equation 2 with a second order polynomial of our running variable and allowing for different slopes on both sides of the cutoff using the full sample. In Columns 2 and 4, we add cohort fixed effects to the model. Standard errors are clustered at the level of the running variable. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table 7:
RDD estimates: Early dropout

	(1)	(2)	(3)	(4)	(5)	(6)
Last-chance exam	0.111*** (0.042)	0.111*** (0.042)	0.123*** (0.042)	0.116** (0.044)	0.116** (0.044)	0.129*** (0.044)
Bandwidth	0.25	0.25	0.25	–	–	–
Cohort FE	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes
Observation	1159	1159	1047	1562	1562	1413

Note: RDD estimates: Early dropout. The different columns show results for different specifications. In Columns 1-3, results are based on a non-parametric approach using local linear regressions on both sides of the cutoff and an optimal bandwidth calculated according to [Imbens and Kalyanaraman \(2012\)](#). In Columns 3-6, we use a parametric specification estimating Equation 2 with a second order polynomial of our running variable and allowing for different slopes on both sides of the cutoff using the full sample. In Columns 2 and 4, we add cohort fixed effects to the model, and in Columns 3 and 6, we add student level background characteristics. Standard errors are clustered at the level of the running variable. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table 8:
RDD estimates: Early dropout - robustness checks

	(1)	(2)	(3)
<i>Panel A</i>			
CCT	0.110** (0.054)	0.109** (0.055)	0.116** (0.051)
Bandwidth	0.15	0.15	0.15
Observation	610	610	544
<i>Panel B</i>			
Donut 1	0.103** (0.044)	0.101** (0.046)	0.113** (0.047)
Bandwidth	0.25	0.25	0.25
Observation	1159	1159	1047
<i>Panel C</i>			
Donut 2	0.122** (0.051)	0.122** (0.051)	0.130** (0.056)
Bandwidth	0.25	0.25	0.25
Observation	1159	1159	1047
<i>Panel D</i>			
German	0.119** (0.052)	0.118** (0.053)	0.124** (0.050)
Bandwidth	0.25	0.25	0.25
Observation	1075	1075	1001
<i>Panel E</i>			
At most one last-chance exam	0.108*** (0.035)	0.109*** (0.035)	0.118*** (0.037)
Bandwidth	0.25	0.25	0.25
Observation	1060	1060	960
Cohort FE	No	Yes	Yes
Controls	No	No	Yes

Note: RDD estimates: Early dropout. This table presents the results from several robustness checks described in more detail in Section 5.1. In Columns 1-3, results are based on a non-parametric approach using local linear regressions on both sides of the cutoff and an optimal bandwidth calculated according to [Imbens and Kalyanaraman \(2012\)](#) except from row 1. In Column 2, we add cohort fixed effects to the model, and in Column 3, we add student level background characteristics. Standard errors are clustered at the level of the running variable. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table 9:
RDD estimates: Type of Early dropout

	(1)	(2)	(3)	(4)	(5)	(6)
<i>A: Voluntary dropout</i>						
Last-chance exam	0.022 (0.027)	0.021 (0.027)	0.032 (0.030)	0.023 (0.027)	0.022 (0.028)	0.037 (0.029)
<i>B: Forced dropout</i>						
Last-chance exam	0.089* (0.046)	0.090* (0.046)	0.091** (0.046)	0.093* (0.049)	0.094* (0.049)	0.092* (0.049)
Bandwidth	0.25	0.25	0.25	–	–	–
Cohort FE	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes
Observation	1159	1159	1047	1562	1562	1413

Note: RDD estimates: Type of Early dropout. The outcome variable in Panel A is a binary indicator being 1 if a student leaves the program early without being ultimately expelled. In Panel B, the outcome variable is a binary indicator being 1 if a student is finally suspended from the program in an early phase. The different columns show results for different specifications. In Columns 1-3 results are based on a non-parametric approach using local linear regressions on both sides of the cutoff and an optimal bandwidth calculated according to [Imbens and Kalyanaraman \(2012\)](#). In Columns 3-6 we use a parametric specification estimating Equation 2 with a second order polynomial of our running variable and allowing for different slopes on both sides of the cutoff using the full sample. In Columns 2 and 4, we add cohort fixed effects to the model and in Columns 3 and 6, we add student level background characteristics. Standard errors are clustered at the level of the running variable. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table 10:
RDD estimates: Graduation

	(1)	(2)	(3)	(4)	(5)	(6)
Last-chance exam	-0.062* (0.037)	0.059 (0.039)	-0.049 (0.036)	-0.037 (0.041)	-0.035 (0.042)	-0.028 (0.038)
Bandwidth	0.23	0.23	0.23	–	–	–
Cohort FE	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes
Observation	1018	1018	917	1562	1562	1413

Note: RDD estimates: Graduation. The different columns show results for different specifications. In Columns 1-3, results are based on a non-parametric approach using local linear regressions on both sides of the cutoff and an optimal bandwidth calculated according to [Imbens and Kalyanaraman \(2012\)](#). In Columns 3-6, we use a parametric specification estimating Equation 2 with a second order polynomial of our running variable and allowing for different slopes on both sides of the cutoff using the full sample. In Columns 2 and 4, we add cohort fixed effects to the model and in Columns 3 and 6, we add student level background characteristics. Standard errors are clustered at the level of the running variable. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table 11:
RDD estimates: Graduation - robustness checks

	(1)	(2)	(3)
CCT	-0.056 (0.039)	-0.048 (0.043)	-0.027 (0.037)
Bandwidth	0.16	0.16	0.16
Observation	821	821	738
Donut 1	-0.057 (0.050)	-0.046 (0.052)	-0.045 (0.050)
Bandwidth	0.23	0.23	0.23
Observation	1018	1018	917
Donut 2	0.087 (0.066)	0.078 (0.068)	0.095 (0.070)
Bandwidth	0.23	0.23	0.23
Observation	1018	1018	917
German	-0.078* (0.046)	-0.076* (0.046)	-0.059 (0.041)
Bandwidth	0.23	0.23	0.23
Observation	1016	1016	946
At most one last-chance exam	-0.047 (0.033)	-0.043 (0.035)	-0.026 (0.034)
Bandwidth	0.23	0.23	0.23
Observation	1006	1006	911
Cohort FE	No	Yes	Yes
Controls	No	No	Yes

Note: RDD estimates: Graduation. This table presents the results from several robustness checks described in more detail in Section 5.1. In Columns 1-3, results are based on a non-parametric approach using local linear regressions on both sides of the cutoff and an optimal bandwidth calculated according to [Imbens and Kalyanaraman \(2012\)](#) expect from row 1. In Column 2, we add cohort fixed effects to the model, and in Column 3, we add student level background characteristics. Standard errors are clustered at the level of the running variable. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table 12:
RDD estimates: Heterogeneity analysis

	(1)	(2)	(3)	(4)	(5)	(6)
	men	women	high ability	low ability	Fail 1st	Fail 2nd
<i>A: Dropout</i>						
Last-chance exam	0.109 (0.070)	0.106** (0.051)	0.051 (0.066)	0.155** (0.072)	0.108*** (0.035)	0.290*** (0.086)
Bandwidth	0.25	0.25	0.25	0.25	0.25	0.25
Observation	636	523	584	575	1060	386
<i>B: Graduation</i>						
Last-chance exam	-0.011 (0.051)	-0.122** (0.059)	-0.045 (0.070)	-0.095* (0.052)	-0.047 (0.033)	-0.169*** (0.044)
Bandwidth	0.23	0.23	0.23	0.23	0.23	0.23
Observation	599	498	546	551	1006	370

Note: This table presents results from a heterogeneity analysis with separate regressions for each subsample. In Columns 1 and 2, we split by gender. In Columns 3 and 4, we split by the median ability measured by the high-school GPA. In Column 5, we show results for the sample of students that fails at most one orientation exam in the first attempt. In Column 6 we show results for the sample of Students that fails in at least one orientation exam in the first attempt. Specifications are based on our main specification (Column 1). Standard errors are clustered at the level of the running variable. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table 13:
Diff-in-disc estimates

	(1)	(2)	(3)
<i>A: Dropout</i>			
High-stakes	0.171** (0.073)	0.173** (0.072)	0.192** (0.077)
<i>B: Graduation</i>			
High-stakes	-0.037 (0.071)	-0.047 (0.071)	-0.020 (0.078)
Cohort FE	No	Yes	Yes
Controls	No	No	Yes
Observation	3053	3053	2774

Note: Difference-in-disc estimates for Early dropout (Panel A) and Graduation (Panel B). Results are based on parametric specifications estimating equation 3 with a second order polynomial of our running variable and allowing for different slopes on both sides of the cutoff using the full sample.

Table 14:
Diff-in-disc estimates by gender

	(1)	(2)	(3)	(4)	(5)	(6)
	men			women		
<i>A: Dropout</i>						
High-stakes	0.183* (0.095)	0.184* (0.094)	0.198** (0.096)	0.145** (0.071)	0.150** (0.071)	0.183** (0.079)
<i>B: Graduation</i>						
High-stakes	0.050 (0.119)	0.044 (0.121)	0.093 (0.127)	-0.151 (0.108)	-0.163 (0.106)	-0.187* (0.103)
Cohort FE	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes
Observation	1702	1702	1567	1351	1351	1207

Note: Difference-in-disc estimates for Early dropout (Panel A) and Graduation (Panel B). We present results separately by gender. Results are based on parametric specifications estimating equation 3 with a second order polynomial of our running variable and allowing for different slopes on both sides of the cutoff using the full sample.

Table 15:
Diff-in-disc estimates by ability

	(1)	(2)	(3)	(4)	(5)	(6)
	high ability			low ability		
<i>A: Dropout</i>						
High-stakes	0.040 (0.106)	0.043 (0.106)	0.017 (0.121)	0.238** (0.093)	0.247*** (0.091)	0.272*** (0.095)
<i>B: Graduation</i>						
High-stakes	-0.076 (0.110)	-0.087 (0.107)	-0.005 (0.110)	0.019 (0.150)	0.001 (0.150)	0.030 (0.159)
Cohort FE	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes
Observation	1721	1721	1553	1332	1332	1221

Note: Difference-in-disc estimates for Early dropout (Panel A) and Graduation (Panel B). We present results estimated on subsamples of the data split by the median ability. Results are based on parametric specifications estimating equation 3 with a second order polynomial of our running variable and allowing for different slopes on both sides of the cutoff using the full sample.

Table 16:
RDD estimates: Further outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
<i>A: Graduation in time</i>						
Last-chance exam	-0.066*** (0.018)	-0.062*** (0.020)	-0.076*** (0.019)	-0.045** (0.022)	-0.044* (0.022)	-0.064*** (0.022)
Bandwidth	0.22	0.22	0.22	–	–	–
Observation	1061	1061	955	1562	1562	1413
<i>B: Probation next semester</i>						
Last-chance exam	-0.111 (0.081)	-0.128 (0.080)	-0.105 (0.082)	-0.134 (0.086)	-0.140 (0.085)	-0.121 (0.088)
Bandwidth	0.16	0.16	0.16	–	–	–
Observation	716	716	642	1354	1354	1229
<i>C: GPA</i>						
Last-chance exam	-0.029 (0.050)	-0.008 (0.036)	-0.027 (0.024)	-0.056 (0.057)	-0.047 (0.049)	-0.048 (0.038)
Bandwidth	0.16	0.16	0.16	–	–	–
Observation	457	457	414	974	974	891
<i>D: Grade in follow-up course</i>						
Last-chance exam	-0.200 (0.174)	-0.106 (0.128)	-0.057 (0.222)	-0.158 (0.169)	-0.101 (0.137)	-0.019 (0.221)
Bandwidth	0.21	0.21	0.21	–	–	–
Observation	839	839	758	1281	1281	1165
Cohort FE	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes

Note: RDD estimates: Further outcomes. In Panel A, the outcome variable is a binary indicator stating whether a student graduated within the regular program length (3 years). In Panel B, the outcome variable is a binary indicator stating whether a student, who did not dropout in the first semester is being placed on probation in semester 2. In Panel C, the outcome variable is the final GPA. In Panel D, the outcome variable is the grade a student receives in the follow-up course. The different columns show results for different specifications. In Columns 1-3, results are based on a non-parametric approach using local linear regressions on both sides of the cutoff and an optimal bandwidth calculated according to [Imbens and Kalyanaraman \(2012\)](#). In Columns 3-6, we use a parametric specification estimating Equation 2 with a second order polynomial of our running variable and allowing for different slopes on both sides of the cutoff using the full sample. In Columns 2 and 4, we add cohort fixed effects to the model and in Columns 3 and 6, we add student level background characteristics. Standard errors are clustered at the level of the running variable. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.