A Second Chance at Success? Effects of College Grade Forgiveness Policies on Student Outcomes*

Xuan Jiang† Kelly Chen‡ Zeynep Hansen§ Scott Lowe¶

September 28, 2022

Click here for the latest version

Abstract

The increased popularity of college grade forgiveness policies, allowing students to substitute the new for the previous grades for repeated classes in their GPA calculations, is controversial yet under-studied. Using students' transcript data from a four-year public institution that abolished and reintroduced grade forgiveness, we find that, grade forgiveness increases the probability of repeating by 68%. The policy benefits the students by incentivizing them to take more STEM courses and graduate with STEM degrees. We show that the concerns about students' slacking-off when grades can be forgiven is trivial. However, the policy increases time-to-graduation and widening the existing gaps between demographic groups.

Keywords: Grade forgiveness policies, course repeat, STEM major, graduation, college student success

JEL Codes: I21, I23

*We thank Bruce Weinberg, Audrey Light, Miguel Sarzosa, and seminar and conference participants at Ohio State, Purdue, Jinan, SOLE, AEFP, APPAM, and 2021 Symposium on Contemporary Labor Economics. All errors are our own.

†Assistant Professor, Department of Economics, Jinan University, jiangxuan@jnu.edu.cn

‡Associate Professor, Department of Economics, Boise State University, kellychen@boisestate.edu

§Vice Provost for Academic Planning, Boise State and NBER, zeynephansen@boisestate.edu

¶Associate Dean of Graduate College, Boise State University, scottlowe@boisestate.edu

1 Introduction

The return on college education has been well-documented and federal and state efforts has been made over the recent decades to encourage more college going, yet the 6-year graduation rate for the first-time, full-time college students in public universities barely made it to 60% (NCES, 2020). Among all endeavours to remove barriers for college students, one type of widely adopted academic policies, grade forgiveness (or sometimes called grade replacement), changes the calculation of students' grade point average (GPA) to help students to persist in college.

Grade forgiveness allows college students who retake courses to replace the old grades with the new grades in GPA calculations. The usual alternative for schools that do not have this policy, is to take the average of the new and the old grades in GPA for repeated courses (i.e., grade averaging). The use of this relatively lenient policy of "erasing" the old grade has accelerated and the number of colleges applying this policy has doubled in the past two decades. To date, 88% land grant universities have adopted grade forgiveness policy. However, is controversial in higher education circle. Opponents of grade forgiveness policies believe that these policies diminish the importance of grades, which could make students less inclined to put forth the level of effort needed to succeed in the course the first time. They also worry that it might be a bad use of academic resources for students to retake courses to improve grades. Proponents argue that the policy could help the unprepared. Many early-stage college students, underrepresented minorities, and first-generation students struggle to thrive in college, both academically and socially, and this policy improves their chances of success.[1]

The key question of the controversies is that whether college should provide more forgiving or more strict academic standards to promote students' success. Intuitively, both pros and cons are sensible and it is unclear to what extent this policy

[1] For example, see discussion in news articles from Atlantic: https://www.theatlantic.com/education/archive/2018/06/college-grades-gpa/564095/ and https://www.miamistudent.net/article/2012/04/mu-implements-grade-forgiveness.

benefits or hurts students. In this paper, we comprehensively evaluate the effects of grade forgiveness on a number of key outcomes of students, including the decision to repeat a course, curriculum choice, performance, and graduation. We directly address the controversies of grade forgiveness in the higher education circle and go beyond that. We are interested in both the policy's direct effect on helping students persist in college via boosting their GPA and its indirect effect on encouraging students to take risks and challenge themselves via providing a safety-net. In short, we find that the opponents and proponents are both right and wrong: they point to different set of facts but are lack of precision on the magnitudes of the impacts.

We first present a simple conceptual framework to illustrate the changes in students' choice on course difficulty, time spent, and repeat decision in response to the policy change from grade averaging to grade forgiveness. We then provide empirical evidence using rich data with student-level admission information and student-course-level transcript information from Boise State University. A marked advantage of investigating grade forgiveness by using Boise State's data is that the University experienced a grade forgiveness policy implementation and reversal within recent years. The earliest cohort in our data (first enrolled in 1990) entered Boise State with a grade forgiveness policy in place. The university abolished grade forgiveness (changed to grade averaging) in 1995 and in fall 2001 reinstated grade forgiveness, which has been in place since then. Hypothetically, these two policy changes should back each other up by showing mirrored effects. Furthermore, any potentially different magnitudes of policy on-off effects and policy off-on effects could provide insights about how intervention would affect students' curriculum choices at different academic stages.[2]

Facilitating by the on-off-on variations of the policy, we use a model including students' individual fixed effects and academic progress fixed effects to estimate the policy's effects on probability of repeating a course, probability of taking a STEM course, first-attempt performance, and within-term grades variation. We begin by

[2]The policy is not retroactive and the calculation of GPA for each repeated course is according to the policy of the semester when repeat happens.

showing that students are significantly more likely to repeat courses when the policy is in place. Specifically, the policy increases the likelihood of repeating by up to 2.3 percentage points, which translates into a 65% increase from the policy-absent period repeating rate of 3.5%.

Changes in probability of repeating courses provide evidence on the mechanism of changes in students' curriculum choices. We find that the policy nudges students to enroll in more STEM courses courses that have more stringent grading policies. For example, students are 10.7% more likely to take a STEM course (including repeats) and 8.5% more likely to initially enroll in a STEM course. The nudge is even stronger for students who enter college without a declared major (10.9%). We also explore the policy's heterogeneous effects on STEM-underrepresented groups. We show that grade forgiveness's positive effect on STEM take-ups is weaker for women relative to men and weaker for students from low-income households than students from high-income households.

By estimating the policy's effect on student's first-attempt performance (with course fixed effects), we find that the concerns about students' slacking-off when their bad grades can be forgiven is trivial. On average, students' grades decrease by about 0.037 under the policy, explaining only less than 1.4% decline in grades. And this decline is driven by academically better prepared students who are, on average, taking more challenging courses in the meantime. By investigating students' within-semester grades variation (including repeats), we find that grades within a term are *not* more dispersed when the policy is in place, suggesting that students are *not* allocating their time or effort in terms of favoring some courses over the other by planning to repeat for better grades later on. This finding backs up the non-existence of slaking-off effect, together implying that students are not giving up easily under grade forgiveness.

For long-term outcome, we look at graduation rate, time to graduate, and STEM degree completion rate (at student-level) by comparing those who were more intensely treated by the policy in the early stage of their college career versus their counterparts who were less intensely treated. We find that grade forgiveness has no effect on the probability of ever graduating from the University for first-time students. However, time to graduation lengthens when the policy is in place. Among the sub-sample of students who did graduated from the University, we see a large increase in STEM degrees under the policy. In specific, the OPT-defined STEM degree increase by 25% and the conservative-defined STEM degree increase by 17%. This implies that, overall, the policy nudges students to obtain challenging degrees without a significant "price," i.e., without having them take longer to graduate.

To the best of our knowledge, we are the first to estimate the effects of grade forgiveness policies on students' outcomes. Indeed, more forgiving or inclusive academic policies are under-studied, relative to more strict policies, such as remedial education (e.g., Bettinger and Long (2009)) and grade retention policies (e.g., Tafreschi and Thiemann (2016)). Unlike the mandatory remedial education and retention policies, which raise standards for the least prepared students, grade forgiveness policies act as an insurance by providing a chance for all students to voluntarily retake courses to improve grades.

In a similar vein, there is a relatively larger, although still small, literature on retaking high-stakes tests, including the SAT, ACT, and college entrance exams. Retaking high-stakes tests substantially improves scores and increases four-year college enrollment rates, particularly for low-income and underrepresented minority students (Goodman, Gurantz, and Smith, 2020; Vigdor and Clotfelter, 2003). Retaking high-stakes tests improves grades both through increased familiarity with the test and through actual learning (Frisancho, Krishna, Lychagin, and Yavas, 2016). While both are inclusive policies in higher education, the essential difference between retaking college entrance exams and retaking college courses is that the former impacts students' success through college admission while the later impacts students' success through college curriculum choice, performance and persistence. Our paper, on the one hand, echoes the studies cited above on the positive consequences of retaking a course (or exam); on the other hand, it provides the first evidence on a more forgiving academic policy's effects on incentivize students to challenge themselves. We contribute to the literature by highlighting a key effect of this widely adopted but overlooked policy—nudging students to challenge themselves and promoting them to pursue and persist in difficult majors.

The findings in our paper are timely and policy-relevant as more and more universities are adopting this type of GPA policy as well as seeking new ways to mitigate the academic struggles of college students. Understanding whether bad grades should be forgiven is critical to promote students' achievement, including their progress at every stage of college and final job outcomes. Grades are important and powerful signals for students' later success, especially for major choices(Bandiera, Larcinese, and Rasul, 2015; Goodman, 2016; Marx and Meeler, 2013; Reshwan, 2016; Tafreschi and Thiemann, 2016). Most existing studies have attempted to address how to promote students' choices on STEM majors by suggesting applying more lenient *grading policies* on STEM courses (Minaya, 2020; Ahn, Arcidiacono, Hopson, and Thomas, 2019). However, modifying course grading policies at the institutional level is extremely difficult to justify, and even more difficult to implement. Professors and instructors of those courses may find it hard to accommodate a lenient grading policy without fundamentally changing the course material and the study goals for students from different backgrounds. The grade forgiveness policy we study, however, does not affect any particular course grading policy. It simply provides an option for students to fix their "failure" without requesting any changes to individual courses.

The existing literature has also documented that self-learning about ability through grades plays a very prominent role in college dropout decisions. Students that drop out between their first and second years would be largely reduced if no self-learning occurred about grade performance or academic ability (Stinebrickner and Stinebrickner, 2012, 2014). Students may revise their beliefs about their own abilities in response to (low) grades, leading some students to leave difficult majors (Ahn et al., 2019; Astorne-Figari and Speer, 2019; Stinebrickner and Stinebrickner, 2013).

However, the option to repeat courses, which most of the existing studies do not consider, would likely affect college dropout and major-switching decisions. In our working paper (Chen and Jiang, 2021), where we focus on lower-performing students' choice and outcomes, we find that retaking a course have a favorable impact on the number of same-subject credits that a repeating student registers for and complete in subsequent years, the difficulty level of the next-course she attempts and passes, and her performance in the next-course enrolled in the same field.

The remainder of the paper is organized as follows. Sections 2 introduces grade forgiveness and the institutional background of Boise State. Section 3 lays out a simple conceptual framework, predicting students' response to the policy change. Section 4 and 5 describe the data, key variables used in the analyses, and outline our main research methodology. Sections 6 presents our empirical results. Section 7 concludes.

2 The Policy Background

Grade forgiveness/replacement policies allow students to retake any courses in which they received low grades and utilize the most recent grade to calculate their overall GPAs. To date, 88% land grant universities have adopted grade forgiveness policy. We collect the implementation status and time for all 4-year universities with more than 10,000 enrollment (IPEDS) and present the fractions of schools with grade forgiveness over time, starting 1961, in Figure 1. We can see that the adoption of grade forgiveness has accelerated in recent years.[3] The increased popularity of grade forgiveness in college is often criticized by the media saying that colleges are trying to ensure their "customers" are satisfied. Although there is no survey on the institutional purpose and impact of this policy so far, the popular press has captured different opinions on this practice from students, both pros and cons.[4]

[3]Our growing list of grade forgiveness policy implementation by U.S. 4-year colleges can be found here. We exclude 133 schools missing a certain implementation time in the figure when calculating the time-specific fraction of implementation.

[4]*"I feel like it might be a bad thing. It would give kids an opportunity to feel like they don't have to work as hard the first time around"*—The Lantern. https://www.thelantern.com/2015/08/freshman-

The practice of grade forgiveness vary across institutions in terms of limits to the number of courses that are allowed to be repeated, with or without cut-off grades allowed to be repeated (e.g., letter grades below D), group of students allowed to utilize the policy,[5] and whether the registration of the repeated course requires approval from advisors. For example, the Ohio State University adopted a similar grade forgiveness policy in fall 2015, which allows students to petition to complete a second-course attempt and to replace the original grade.[6]

2.1 Boise State University's Policy

Boise State University (BSU) is a four-year public university located in the northwest United States, with an undergraduate population of approximately 22,000. During the observation period, BSU had the largest undergraduate enrollment in the State of Idaho and offers nearly 80 bachelor's degrees across seven academic colleges: Arts & Sciences, Business & Economics, Education, Engineering, Health Sciences, Innovation & Design, and the School of Public Service.

BSU has a long history with grade forgiveness policies, also referred to as grade replacement policies by the University. A grade replacement policy was implemented prior to 1970, which allows students who receive grades below "D" to retake a course to improve a grade and use the most recent grades for GPA calculation. Starting Fall 1988, students can retake a course to improve *any grades* and use the most recent grades for GPA calculation. This grade replacement policy was replaced with a grade averaging policy in Fall 1995 and then reinstated in Fall 2001.[7,8] The only difference between grade averaging policy and grade forgiveness policy

forgiveness-now-undergrad-amnesty/. *"It will allow students who are new to the rigor of a college education a period of adjustment. This policy will serve as a safety net for students to recover from a course that they might not have been prepared for"*—The Miami Student. https://www.miamistudent.net/article/2012/04/mu-implements-grade-forgiveness.

[5]Some institutions have applied grade forgiveness policies to the entire student body, like Boise State, some to only first-year students or students of certain programs.

[6]https://math.osu.edu/undergrad/non-majors/scheduling/repeating/forgiveness

[7]See Boise State University's 1995-1996 undergraduate catalog describing the repeat policy on pages 23-24 on this webpage: https://scholarworks.boisestate.edu/catalogs/8/.

[8]See Boise State University's 2001-2002 undergraduate catalog describing the repeat policy on pages 23-24 on this webpage: https://scholarworks.boisestate.edu/catalogs/67/.

is that the repeated grade and the first-attempted grade were averaged when calculating the student's GPA under grade averaging, which is a typical practice at all universities without a grade forgiveness. These policies have been applied to all students in the University regardless of programs, academic standing or grades. The university's official statements on the policies are listed below:[9]

"Courses repeated prior to Fall 1995 use a grade replacement policy. Only the most recent grade was used in calculating the cumulative GPA."

"Courses repeated Fall 1995 through Summer 2001 used a grade averaging policy. Courses repeated will be averaged, using both grades in the calculation of the GPA."

"Beginning Fall 2001 and on, courses repeated will use a grade replacement policy. Only the most recent grade will be used in the calculation of the cumulative GPA."

Why the Changes?

In Fall 1995, the University replaced grade forgiveness with grade averaging to restrict the possibility of improving grades by course repeating thus raise academic standards. Starting Fall 2001, the University reverted the course repeat policy back to grade replacement policy. It is worth noting that the University was served by different sets of Academic Standards Committees during the years of policy changes. According to the 2001 Committee, they reinstated grade replacement policy for two primary reasons. First, it caall me to the faculty's attention that the grade averaging policy treated the first and the second attempts as two independent grades for GPA calculations, rather than averaging the two grade points, which appeared to "penalize students to a greater extent than was first proposed" and "it [was] more difficult for students to raise their GPA".[10] Second, most other colleges and universities in the state of Idaho implemented a grade replacement policy at that time, and "this has proved to be unfair to incoming transfer students" because those students took courses at their original institutions in good faith under

[9] See current official webpage describing policy changes here: https://www.boisestate.edu/policy/academic-affairs-student/policy-name-course-repetition-gpa-relationship/

[10] For example, a student has three entries on their transcript: ECON 101 (grade B); ECON 102 (first time, grade D); ECON 102 (second time, grade C). Both courses offered 3 credits. Based on the grade averaging policy in force, the GPA will be $\frac{3\times3+3\times1+3\times2}{3+3+3} = 2$, rather than $\frac{3\times3+3\times\frac{1+2}{2}}{3+3} = 2.25$.

the grade replacement rules.[11]

Students were well-informed of the grade replacement policy changes by academic advising, and were made aware of the policy changes in a timely manner. In addition, grade policies were published in each year's undergraduate catalog, and in the student newspaper, the "Arbiter". In the Arbiter, January 18, 1995, an article titled "New grade rules will greet students next fall" informed students of the academic policy changes that would take place in Fall 1995. In the Arbiter, August 30, 2001, an article titled "Grade replacement policy takes effect this semester" mentioned the re-introduction of grade forgiveness as "Students have a new tool this semester to improve their all-important grade point averages."[12]

Tuition Cost for Repeating A Course

For full-time students, who register for between 8 and 19 credits during Fall 1990–Summer 2008 or register for between 12 and 17 credits during Fall 2008–Summer 2016, the tuition is a flat rate. Thus for this group of student, repeating a course is free. For part-time students or "overload" students (who register for more than the upper limit of the full-time credit hours), the sticker price for each credit varies from $61.75 to $297 during 1990–2016 (same for in-state and out-of-state students). As a public institution, BSU also receives state funds to help subsidize the price of a degree for each student, based on eligibility. The sticker price should be considered as a upper bound of the tuition cost for repeating a credit. About 61% of BSU students are enrolled as full-time students, among which about 4.5% full-time students are "overloaded".[13] Full-time students are more likely to repeat a course than a part time students. Among the students who have repeated at least one course, about 80% repeated one course, which typically accounts 3 credits. Thus, for a typical part-time student, the sticker price for repeating a typical course is around $180–$900 during 1990–2016. On our rough calculation, the University's

[11] The decisions of the policy changes were mentioned in the meeting files of the Academic Standards Committee.

[12] See https://scholarworks.boisestate.edu/cgi/viewcontent.cgi?article=1985&context=student_newspapers and https://scholarworks.boisestate.edu/cgi/viewcontent.cgi?article=2194&context=student_newspapers.

[13] "Overloaded" students are often high-performing and attempting multiple majors/minors.

revenue from course repeating is about only $30,000–$40,000 per academic year.

3 Conceptual Framework

Conceptually, when students are making decisions to repeat a class after grades being revealed, the expected final grade of a certain course (or GPA) is higher under “Grade Forgiveness" policy than under grade averaging policy, while the expected cost of repeating remains the same under the two policies. Thus the policy will have a direct, or rather an “intended" effect from the view of policy makers, on students who are on the margin to retake a class. This set of effects, including the subsequent outcomes of the students who repeated courses, indicates whether the policy benefit its “targeted" group. When students are making decisions to enroll in any course, “Grade Forgiveness" policy offers students a costly insurance against a low grade, relative to grade averaging policy. If the insurance value of using the “Grade Forgiveness" policy outweighs its cost, we would see it incentivizes students to take risks. In this sense, the policy has an indirect effect—functioning through the intended possibility of repeating a course—on all students, who may or may not choose to retake a course.

In this section, we lay out a simple theoretical framework to illustrate the indirect effects of “Grade Forgiveness" policy, relative to the grade averaging policy. We are especially interested in students' decision making in course choice, time choice, and time allocation. This framework will help us interpret our results and provide insights for the mechanism of the empirical evidence on course choice, performance, and grades variation.

We consider the environment where students make two choices before they enroll in a course to maximize their single period utility: the type of courses (difficulty) to take and study time. We assume the students are myopic and only optimize the current period. The type of courses a student chooses for a semester depends on both the difficulty of each course (quality) and the total number of courses/credits enrolled (quantity). Here, we do not distinguish the quality and quantity and use a one dimensional difficulty definition, d.

We assume students value both the difficulty of courses, d, and the expected grade, g. The intuition is that students should value grades as well as the type of courses they study, which implicitly associated with the difficulty level. And the type of courses students choose constitute their major choices, although we are not directly measuring major choices. Studies have shown that subjects/majors gearing towards high-paying jobs (i.e., STEM) are often associated with tough grading policies. Thus, it is critical to consider d in the utility function.

A course grade received by a student (or grading policy), $g(d,t)$, is determined by the difficulty of the course, d, and study time, t. The cost of taking a course is also a function of course difficulty and study time, $c(d,t)$, which could be considered as a combination of a time cost and mental cost. Assume g is twice continuously differentiable and follows the law of diminishing marginal returns: $g'(d) < 0$, $g''(d) < 0$, $g'(t) > 0$, $g''(t) < 0$. The more difficult of the course, the lower grade will be received and the more time spent on the course , the higher grade will be received. Similarly, $c'(d) > 0$, $c''(d) < 0$, $c'(t) > 0$, $c''(t) < 0$.

The probability of repeating a course, is a function of the course grade, $f(g)$, and is decreasing in g: $f'(g) < 0$. The implicit threshold grade (reservation grade) for a student to repeat is g^*. In other words, a student will choose to repeat when the grade ranges between the lowest grade to the threshold grade, [F, g^*], and will choose to not repeat when the grade ranges between the threshold grade to the highest grade, [g^*, A]. We can write the utility function as the summation of the utility function of repeat and the utility function of not repeat integrated over a function of the first-attempt grade, $f(g)$.

$$U(g,c,d,t) = \int_F^{g^*} U^{(\text{repeat})} f(g)dg + \int_{g^*}^{A} U^{(\text{not repeat})} f(g)dg$$

When the student does not repeat a course, the utility is simply determined by the difficulty level, the grading policy, and cost function:

$$U^{(\text{not repeat})} = d * g(d,t) - c(d,t)$$

When the student makes the decision to repeat a course after revealing the first-attempt grade, the utility under each policy can be written as below:

$$U^{(\text{repeat})} = \begin{cases} d * \frac{E[G]+g(d,t)}{2} - c(d,t) - E[C], & \text{if Forgiveness = 0} \\ d * E[G] - c(d,t) - E[C], & \text{if Forgiveness = 1} \end{cases}$$

where $E[G]$ is the expected grades of repeating (the second-attempt grade) and $E[C]$ is the expected cost of the repeating. We assume that, under "Grade Forgiveness" policy, when students can replace the repeated grade with the original grade, students do not care about the original grade because the first-attempt grade will not be counted in the GPA. It is arguable that the first-attempt grade should not have a zero weight as both grades will show up on a transcript. However, as long as the weight on the expected repeating grade is lower than the first-attempt grade, our theoretical implications will remain the same and we choose to keep this way for simplicity.

We derive the following propositions regarding the choice on difficulty of course, study time, and time allocation, for students at the stage of making course enrollment decisions and before revealing first-attempt grade. We also derive propositions regarding the decision on repeat for students who have revealed the first-attempt grades. Proofs are in Appendix A.

3.1 Difficulty and Time Choice

Proposition 1. Students will choose higher level of difficulty under the "Grade Forgiveness" policy (i.e., Forgiveness = 1), relative to under the averaging policy (i.e., Forgiveness = 0).

Proposition 2. Students will choose less study time under the "Grade Forgiveness" policy, relative to under the averaging policy.

Revealed by these propositions, the major criticism of "Grade Forgiveness" that students will not work as hard the first time around is *right* in the sense that students will spend less time to study when having the option to improve their grades later by repeating; but *wrong* in the sense that students will also choose more difficult course to enroll in when having a safety-net to fall back on. Our empirical evidence in section 6 corresponds to these theoretical implications and discusses further how the policy benefits/harms students economically.

3.2 Time Allocation

Proposition 3. The time allocated among different courses is more dispersed under the "Grade Forgiveness" policy, relative to under the grade averaging policy.

When having the option to repeat any course to improve grades later, students will allocate time in favor of the course(s) that are more likely to obtain higher grades and allocate time away from the course(s) that are more likely to obtain lower grades in the current semester. We should expect to see a larger variation in a student's grades within a term under the Grade Forgiveness policy relative to the grade averaging policy.

3.3 Probability of Repeating and Threshold Grade

For students who have completed a course and revealed the first-attempt grade, g, we derive the probability of repeating and the threshold grades for repeating under the two different GPA policies.

Proposition 4. The average probability of repeating under Forgiveness = 1 will be higher than the probability of repeating under Forgiveness = 0.

Proposition 5. The threshold (highest) grade to repeat under Forgiveness = 1 is higher than the threshold (highest) grade to repeat under Forgiveness = 0: $g_1 > g_0$, and the difference between the two threshold grades is restricted as $g_1 - g_0 \leq E[C]$.

4 Data

We use administrative transcript data from undergraduate students at the University, who entered the university between the spring 1990 and spring 2017 semesters. The raw data include approximately 170,812 students and provides admissions information, including residency, SAT/ACT scores, and demographic characteristics. The student-section-level transcript data provide all courses enrolled in, credits attempted, credits earned, grades obtained, and information on the courses that the student has repeated.

Table 1 provides summary statistics on characteristics of interest. Of the 170,812 students in the full sample, 53.6% are female. Based on 2020–2021 figures, BSU has 73% White, 13% Hispanic or Latino, 5% Two or More Races, 3% Asian, 2% Black or African American, 0.397% Native Hawaiian or Other Pacific Islanders, and 0.389% American Indian or Alaska Native. 66% of students come from the state of Idaho, and 1% are international students.[14] About 45% students at BSU are transfer students. We exclude students that transferred-in coursework in some of our analysis because we do not observe their full transcripts, and courses taken from other institutions may not be comparable with those offered by BSU. The average course repeat rate is 0.035 during the policy-off period and 0.046 during the policy-on period.

It is worth mentioning that we achieve our final data from two major sources. One is the fully digitized transcript achieve for all transcript records from Summer 1998 and onward. BSU transitioned from paper recording to a centralized digital archiving system in the summer semester of 1998. Since then, a set of additional information also becomes available, including the specific course section the student attended, the instructor of the specific course section, and a more extensive set of demographic characteristics including the student's ethnicity, age at college entry, and in-state status. The other source of the transcript data was in PDF forms and we

[14] https://www.boisestate.edu/about/facts/

were able to parse all information based on PDF transcripts for students who first entered BSU from 1990–1998. The PDF transcripts provides all key variables for our main analysis, although is lack of the additional information in course section number, instructor, race and ethnicity.

5 Empirical Strategy

We aim to estimate the policy's total effect on all students who may or may not choose to retake a course. This set of outcome variables of interest includes the student's curriculum choice—measured by the type of courses and the number of credits students choose to enroll in, choice on time spent—revealed by first-attempt performance, and time allocation among courses—proxied by grade variation within a semester. Given the richness of our individual-course-level transcript data, we explore a series of outcomes affected by the policy using a fixed effect model including individual fixed effects and individual's academic progress fixed effects (academic term t) with two variations of the treatment over time. The estimation can be written as follows:

$$Y_{ijat} = \beta_0 + \beta Policy_t + \gamma_i + \delta_a + \theta_j + \Lambda S'_{jt} + \varepsilon_{ijat} \tag{1}$$

where i denotes students, j denotes courses, and t denotes the calendar semester. a is the academic progress/semester for each student, representing the number of semesters elapsed since a student's initial enrollment at the time of observation (i.e., 1st, 2nd, 3rd,..., semester). γ_i is a vector of individual fixed effects, and δ_a is a vector of academic progress fixed effects. The key independent variable, $Policy_t$, is an indicator of the policy's presence, which equals 1 for any course taken by any student during the Spring 1990 (start of our data)–Summer 1995 or Fall 2001–Fall 2019 (end of our data) periods, and 0 for any course taken by any student during the Fall 1995–Summer 2001 period. It is essential to distinguish the *academic semester for each student*, a, and the *calendar semester*, t. The policy varies by the calendar semester while the treatment groups considered varies for each student at their academic progress, a. The individual fixed effects, γ_i, and academic progress fixed effects, δ_a, could be considered facilitating a conventional difference-in-differences approach. This approach identifies, among all partially-treated individuals, the difference between students who entered the university at different times in the differences between outcomes from any two academic progress s, a' and a. Assuming parallel trends in a typical difference-in-differences design means that when we compare two students, A, who has not been treated in academic semesters a and a' while B has been treated in a' but not a, the difference between the outcome in a and a' of student B would have been identical to the difference between the outcome in a and a' of student A, if student B had not been treated.[15]

We also include course level controls when applicable. θ_j is a vector of course fixed effects, capturing any course-specific characteristics. We include θ_j in all of our specifications when the dependent variable varies at individual-section level. S_{jt} is a matrix of course-section characteristics, including mean GPAs and enrollments, further capturing within-course between-sections variations. β is the coefficient of interest here.

The primary advantage of taking this approach is to control for the unobserved individual heterogeneity, such as innate ability, academic preparation, family background, risk aversion, etc. Without individual FE, we are making a strong assumption that students' composition across time is alike; however, students' composition could be changed due to other contemporaneous institutional changes, thus confounding the grade forgiveness's effect.

To provide more insights into the policy's dynamic effect across time, we also apply a standard difference-in-differences (DD) event-study to show grade forgiveness's effects across cohorts. This is not our preferred specification due to omitted

[15]This approach is not an analogue to the staggered difference-in-differences. The criticism that recent work in econometric theory has on the staggered DiD designs, where groups adopt the policy or treatment of interest at a particular point in time and the key to causal interpretation of the estimates requires both a parallel trends assumption and treatment effects that are constant over time does not apply here.

variable concerns when comparing across cohorts; yet, it well serves as supportive evidence on how students respond to the two policy changes over time. Specifically, we estimate the differences in the outcome variable between a given year and the event year—the reintroduction of grade forgiveness in fall 2001 policy. We expect the abolition (fall 1995) and the reintroduction (fall 2001) of grade forgiveness policy potentially alter repetition behavior and curriculum choice in opposite directions. The specification is below, where the variable of interest, $\mathbb{1}(Year_t - 2001 = k)$, is an indicator for k years from the reintroduction of grade forgiveness (2001) which is equal to one when the years of observations is -11 (year 1990), -10, -9, ..., 5, 6 (year 2007) years from 2001.

$$Y_{ijat} = \beta_0 + \sum_{k=-11}^{6} \beta_k \mathbb{1}(Year_t - 2001 = k) + \delta_a + \theta_j + \Lambda P_i + \varepsilon_{ijat} \tag{2}$$

We include academic term-t fixed effects, δ_a, and course fixed effects, θ_j, in the regression of probability of repeating, as we do in our preferred specification (1). Although the beauty of the event-study is being able to compare policy's effects across time, the challenge it is facing is that the student composition is different across time. Essentially, different student composition could result in different repetition behavior or curriculum choice under different academic year, confounding with the policy effects. In order to minimize the omitted variable concern, we control for a set of variables varying at the individual-level, P_i, including a linear term of the first entry year, fixed effects of the season at the first entry, gender, average SAT, and home zip-code median income.

6 Results

6.1 First-order Effect: Do Students Repeat More?

We begin by analyzing students' responses to grade forgiveness in terms of course repeating behaviors. An increase in the probability of repeating should be the first-order effect of the policy, if students are responsive to it. The outcome variable of interest here is a binary indicator which equals to 0 if a course is shown in the transcript for the first time; and equals to 1 otherwise. Table 2 shows the policy's effect on the probability of repeating, where the outcome variable is an indicator of a course being a repeat (second attempt): $Y_{ijat} = 1$ indicates the course j taken by student i in academic progress a and in calendar semester t is a repeat, as opposed to a first attempt. The regression sample include all course level observations excluding incomplete grades, audits, or courses or sessions offering Pass/No Pass grades. Each column is a separate regression specified in Equation 1.

We present the fact that the policy significantly increases the probability of repeating by showing regression results from simple to rich specifications. Column (1)'s regression includes academic term-t fixed effects and course fixed effects, which is equivalent to the average of the point estimates from the event-study using the same sample. It indicates that grade forgiveness increases the probability of repeating by 1.6 p.p., counting for 48% increase from the baseline sample mean.[16] Column (2) includes individual fixed effects and academic term-t fixed effects; column (3) further adds in course fixed effects. The magnitude of the estimates in columns (2) and (3) are both larger than the one in column (1), suggesting that unobservables at individual level bias the estimate downward. As we are exploiting within-individual variations, we narrow the regression sample down to a more relevant time window, 1990-2008, in column (4) to minimize the time-variant factors influencing the estimates through academic progress fixed effects or course fixed effects.

The estimates are highly robust across different specifications. Specifically, under the policy period, the probability of a course being a repeated course is 2.27 percentage points higher than that in the no-policy period. On average, the repeat rate for policy-off periods is 3.3% in the analysis sample. Thus, our analysis shows the policy increases a student's probability of repeating by up to 68%.

[16]The mean probability of repeating a course during policy-off period, 1995-2001, is 0.033.

Figure 2 shows the event-study estimates of the grade forgiveness effects on the probability of course repetition using specification (2). Focusing on estimates around the events, we see a clear drop in the probability of repeating in the fall of 1995 when the University abolished grade forgiveness, and a sharp surge in the probability of repeating in the fall of 2001 when the University reintroduced the policy. We do not expect symmetric policy effects before the abolishment and after the reintroduction because the course-taking decisions and repeating decisions are not the same for cohorts affected by the two events. We do notice that the policy's effect is trending up to a larger extent after the reintroduction, relative to the period before the abolishment. We cannot rule out the possibility that differential policy effects are associated with unobservables by different student cohorts, for instance, student quality.

We thus further provide point estimates of the policy's effect on the probability of repeating during different periods—in place, abolished, and reintroduced—using the specification in (1), where any potential differences in student quality varying over time have been captured by the individual fixed effects. The results are shown in Table B1. We show that the policy's effect on the probability of repeating is robust and consistent for both the cohort who experienced a policy turn-off and the cohorts who experienced a policy turn-on. This provides convincing evidence for our identification strategy that the policy's effects are not subject to specific time periods nor are cohort-sensitive.

In addition to the changes in the probability of repeating, we observe the changes in grades that got repeated. As shown in the bottom panel of Table 1, among all first-attempt courses that got repeated, the share of A/B/C grades is higher than the share of D/F/W grades in policy-on period than in policy-off period. This implies that students are more likely to repeat "unsatisfying" grades (A/B/C), relative to "failing" grades (D/F/W) when grade forgiveness is in force. This finding is consistent with the theoretical implication that the threshold (highest) grade to repeat under grade forgiveness is higher than the threshold grade to repeat under grade averaging. Grade forgiveness may be particularly attractive to high-achieving students or students that need certain a GPA to be admitted into or progress in their program.

6.2 Choosing More Challenging Courses

6.2.1 STEM Courses

Having shown the policy's effect on the probability of repeating, we have justified both the mechanism and validity of exploring the policy's second-order (indirect) effect on all students, including those who did and did not actually repeat a course. One of the most important hypotheses we would like to empirically test is whether students choose to challenge themselves, given that grade forgiveness policy serves as insurance for potential bad outcomes. As suggested in proposition 1, when the option value of this insurance outweighs the expected cost, we should observe that more students are incentivized to take more challenging courses and (or) more credits per semester. We test this hypothesis by estimating the difficulty of courses taken as well as the number of credits attempted in each semester, between the treated and untreated semesters.

We start by estimating to what extent the policy nudges a student to take a course in Science, Technology, Engineering, and Mathematics (STEM). Literature has shown that the harsher grading practices observed in STEM subjects often deter students' participation, even though STEM jobs are among the highest-paying jobs. Thus, we use an indicator of whether a course in the transcript is a STEM subject course as the outcome variable. Table 3 shows the corresponding results: the policy's effect on the likelihood of taking a STEM course. We first use a conservative definition of STEM subjects for defining the outcome variable in Panel.[17] We carry this analysis in the full first-time students sample and two sub-samples. Column (1) shows that, on average, grade forgiveness policy nudges an average student to take

[17]Our conservative definition of STEM defines all natural sciences, engineering, and most medical sciences as STEM subjects.

a STEM course by 2.39 percentage points. Column (2) excludes non-first-attempt courses and shows that, for first-time students, the policy increases their probability of initially taking a STEM course by 1.91 p.p., accounting for a 9% increase from the sample mean. More importantly, for students who are not in a STEM major or have not declared a major at entry year, the policy increases their likelihood of enrolling in a STEM course by 2.05 p.p., accounting for a 11.3% increase from the sample mean. This is not trivial, considering the small body of STEM majored students and the possibility that this group of students might end up majoring in STEM. We use another definition of STEM subjects, based on DHS STEM Designated Degree Program List[18], which could be consider as a broader definition of STEM, in Panel 2 to further explore the curriculum choice as well as to check the robustness of Panel 1. Estimates in Panel 2 shows that the policy has a similar effect on nudging student to take STEM subject courses under this broader definition. Not surprisingly, the policy increases course take-ups more for broader STEM subjects.

Figures 3a and 3b show the event-study estimates of the probability of taking a STEM course based on the conservative definition and the OPT definition of STEM. We control for gender, average SAT, home zip-code median income, a linear time trend of entry cohort, term-season fixed effects, academic term t fixed effects, and shares of STEM courses offered by BSU in each semester. Besides the clear positive effects in the left and right policy-on windows, relative to the middle policy-off window, we see a mild trend-down effect in the first few years since the policy abolishment, instead of a cliff as we see in the probability of repeating. This is not surprising and is driven by students who had already been taking the STEM track and would continue taking STEM courses.[19]

One could argue that using STEM to measure difficulty is not comprehensive and non-STEM courses could be found difficult. To broaden the definition of dif-

[18]The U.S. Department of Homeland Security (DHS) STEM Designated Degree Program List is a complete list of fields of study that DHS considers to be science, technology, engineering or mathematics (STEM) fields of study for purposes of the 24-month STEM optional practical training extension described at 8 CFR 214.2(f). See: https://www.ice.gov/sites/default/files/documents/stem-list.pdf

[19]The decline in the year 2002 ($k = 1$) is still mysterious to us.

ficulty, we construct alternative and continuous measures based on grading harshness. The idea is, any course that offers relatively low grades are considered difficult (to succeed) by students. Basically, we collapse grades to course-level, averaging all same course grades offered from different years and different sections. We use either the numerical grades or letter grades to construct this measure, we describe the construction of which in detail in Section B and present the results in Table B3. In brief, we find that grade forgiveness leads to more enrollment in courses that offer relatively lower grades. Furthermore, we see that the number of courses and credits attempted in each semester increases when the policy is in force, suggesting that students are also incentivized to challenge themselves in terms of quantity in addition to quality.

6.2.2 Differential Effects on STEM-underrepresented Groups

Gender

Policymakers and researchers have been concerned about the under-representation of women in STEM fields, given the expected shortage of STEM workers and the likely effects of the gender gap in college major choice on the pre-existing gender wage gap. A consensus in the recent literature is that women value grades significantly more than men, and the nature of STEM subjects' often-harsher grading practices deters women's participation in STEM (Rask and Tiefenthaler, 2008; Ost, 2010; Owen, 2010; Butcher, McEwan, and Weerapana, 2014; Ahn et al., 2019; Minaya, 2017). A related and relevant question that our study can answer is that, if given another chance, whether women would be more likely than their male counterparts to retake a course for better grades, and thus to pursue a STEM major. Unlike altering STEM courses' grading policies thus "inflate" grades to encourage more women to enroll in STEM courses, as suggested in the previous literature, grade forgiveness does not promise higher grades. It merely offers students costly insurance against low grades. Students face the time, effort, mental cost, and opportunity cost of taking other courses when retaking a course. If the net benefit—

insurance value of grade forgiveness less all the costs—for women excesses the net benefit for men, we would see grade forgiveness nudging more women to retake courses than men.

We estimate the policy's gender-specific effect on course repeating and course enrollment in STEM subjects and show the estimates in Panel 1, table 4. Column (1) takes the regression specification in (1) and adds an interaction term of gender and policy. The coefficient of the interaction term indicates that, on average, the policy has a significantly weaker effect on nudging women to repeat a course than nudging men.

One concern of this estimation is due to the known gender differences in grades. Since the dependent variable in this specification is the indicator of whether a course is a repeat course (being taken for the second time), we do not observe nor control for the repeated grades. If women are in general performing better than men, we will overestimate the gender differences in grade forgiveness's effect. We thus present an alternative specification with the dependent variable being an indicator of whether a course has been repeated and control for the course grades of the first attempt. The regression can be written as follows.[20]

$$Repeated_{ijat} = \beta_0 + \beta Policy_{t'} + Grade_{ijat} + \gamma_i + \delta_a + \theta_j + \varepsilon_{ijat} \quad t' \geq t \tag{3}$$

Similar to column (1), column (2) indicates that, on average, the policy has a significantly weaker effect on course repeating for women than for men, even after controlling for the course grade of the first attempt.[21]

[20]Note that the dependent variable here is an indicator of whether a course is being repeated in any subsequent semesters. Only in this way, we can observe and control for the grades *being repeated*, the first-attempt grades. However, we would not know the exact timing of the repeat. The repetition decision could been made in response to the policy in any subsequent period after receiving $Grade_{ijt}$. Although we only show the estimate of ($Policy_t$) in column (2), Table 4, we run robustness checks using different timing of policy, one-semester-lagged ($Policy_{t+1}$) and two-semester-lagged policy ($Policy_{t+2}$), and we see robust results from these specifications. As supportive evidence, we show that most repeats happen in the next two semesters in Figure 4.

[21]In order to make sure that columns (1) and (2) are comparable, we drop all course pairs that the first-attempt and the second-attempt (repeat) are across policies. For example, a student's first-attempt of MATH 101 is taken between 1990-1995 while the second-attempt of MATH 101 is taken between 1995-2001. Dropping those cases would probably underestimating the policy's effect on repeating; thus we should consider columns (1) and (2) are conservative estimates. This sample

Therefore, it is not surprising that the policy incentivizes women significantly less, relative to men, to enroll in STEM courses. Indeed, in columns (3) and (4), we see that women are 1.26 percentage points less likely to enroll in an OPT STEM course and 1.43 percentage points less likely to enroll in an conservative STEM course.

The findings above suggest 1) the total cost of repeating a course is much higher for women than for men, off-setting women's higher preferences on higher grades, making them less likely to repeat a course, conditional on the same grades. As the financial cost of repeating is the same for women and men, the gender difference in repeating cost must be mental cost, which could be related to the gender differences in negative signal taking and overconfidence (Stinebrickner and Stinebrickner, 2012): low grades may be perceived as low ability or major mismatch for women; thus, women are less likely than men to persist in the subject. 2) the option value provided by grade forgiveness is larger for men, relative to women, thus they are more likely to take difficult courses. Besides the perceived cost of repeating is lower for men than for women, another possible explanation resides in the literature on that women are more risk-averse than men (Borghans, Heckman, Golsteyn, and Meijers, 2009).

Low-income Students

Besides women, students from the lower social-economic background are also known as underrepresented in college STEM majors. We next show the policy's effects on students from financially disadvantaged backgrounds. We proximate a measure of whether a student comes from a low-income background by linking their home address zipcode to the median household income by zipcode in 1999 from the Census. Basically, we define a student as a low-income student if his/her home is located in a zipcode where the median household income is below the US median household

restriction the reason why the number of observations in Table 4 is different from the number of observations in Table 3.

income. We then take this binary variable into the regression.[22]

Panel 2 in Table 4 shows the policy's differential effects on lower-income students, relative to higher-income students. First, we find that, although low-income students are performing worse on average than their high-income counterparts, they are not different in terms of repeating a course, conditional on grades received on the first-attempt. However, low-income students are less likely than high-income students to be nudged by the policy to take a STEM subject class. Although both groups are more likely to take a STEM course under grade forgiveness, the positive effect is 1.45–1.66 percentage points smaller on lower-income students than higher-income students.

It is worth highlighting that the policy's between-genders differential effects and between-SES differential effects are through different channels. Conditional on first-attempt grades, women are less likely than men to repeat while lower-income students are not repeating at a different rate than higher-income students. This evidence suggests that the negative signaling effect of low grades is different between genders but not different between social-economic status. Thus, while both women and lower-income students are less likely to take challenging courses than their counterparts, we can infer that the gender difference is more driven by gender difference in perceived mental cost while the SES difference is more driven by the perceived financial cost.

Overall, this section shows that grade forgiveness incentivized all students to take more STEM courses; however, this positive effect is relatively smaller on STEM-underrepresented groups. Note that we are not yet able to conclude that this policy hurt the STEM-underrepresented groups, but our finding confirms that grade forgiveness is not efficient in mitigating the gap between demographic groups in college STEM majors. We complement previous literature by showing a need for effective policies in reducing the corresponding costs for STEM-underrepresented

[22]Since we only have race and ethnicity variables for the post-1998 sample, we are not able to utilize the policy's on-off-on variations to rigorously identify the policy's effect on students of different races or ethnicity. Instead, we show policy's differential effects on students from different financial backgrounds.

groups to enroll in STEM college majors. We continue to compare the policy's differential effects on STEM degree between demographic groups in Section 6.4, where we further provide evidence on that grade forgiveness enlarges the gap between STEM-represented and underrepresented groups.

We also discuss the differential choices of taking challenging courses among students with different academic preparation in Appendix C. We find that, overall, the policy's encouragement of taking challenging courses is more pronounced among students who have stronger academic preparation than students with weaker academic preparation.

6.3 Performance and Time Allocation

As argued by the opponents of grade forgiveness, students would not work as hard around the first time when they have the option to retake the class and replace the low grade. In the theoretical framework, we assume students are myopic and do not count for future cost to derive an implication (Proposition 2) consistent with this argument. That is, students would choose less study time under grade forgiveness. In this section, we test this hypothesis empirically by estimating the policy's effect on students' first-attempt performance.

Ideally, we would like to test the policy's effect on time spent on each course. Unfortunately, we don't directly observe time spent; instead, we use performance to proximate students time spent on each course. We would like to regress the (numerical) grades on the policy indicator using the sub-sample of all non-retake classes. The estimate of interest is the changes in first-attempt grades in response to the policy change. One challenge of this estimation is that grades are potentially influenced by the general grade inflation over time (Bar, Kadiyali, and Zussman, 2009; Chen, Hansen, and Lowe, 2021; Denning, Eide, Mumford, Patterson, and Warnick, 2022), which is rooted in the supply side of grades (instructors) instead of the demand side of grades (students). Since our policy is captured by time variation, grade inflation over time might mask students' performance changes in response to grade forgiveness. In order to control for general grade inflation trend, we construct a de-trend numerical grades variable to replace the original numerical grades. Specifically, we first predict numerical grades using course-specific linear time trend and take the difference between the original numerical grades and the predicted numerical grades. It is worth highlighting that we use the policy-on time periods (1990-1995 and 2001-2008) only to estimate the course-specific linear time trend, in order to further minimize any potential influence of grade forgiveness on grade offering.

We then regress the de-trend numerical grades on the policy indicator, following our main specification in (1). The results are shown in Table 5. Column (1) shows that first-time class takers receive 0.0382 points lower when grade forgiveness is in place, which translate to a 1.46% decrease from the mean. To explore the heterogeneity across students with different academic preparedness, we add in a set of academic preparedness indicators and the interaction terms of the policy indicator and academic preparedness indicators. In specific, we use each student's first term GPA as a measure of the student' academic preparedness; then we categorize the first term GPA to quartiles 1 to 4 to represent students with lowest to highest academic preparedness. The estimates of the interaction terms are shown in column (2). We can see that the least-prepared students, corresponding to the baseline estimate in row 1, perform neither better or worse when the policy is in place, relative to the policy is not in place. Students who are better prepared, quartiles 2-4, perform relatively worse when grade forgiveness is in place. For example, students in quartile 3 receive 0.0671 points less in response to the policy, which translates to a 2.56% decrease from the mean.

Overall, we don't see economically significant decrease in performance when grade forgiveness is in place; especially, we don't see any of that among the least prepared students. This finding put away the concern that students might "slack off" when they have the opportunity to re-do it. The discrepancy between empirical evidence and the theoretical implication—consistent with opponent's opinion—is that students are not as myopic as what has been assumed in the theory. In other words, bad performance is costly for the future periods and students are not fully discounting the cost when making decision on how much time to spend on a class.

Similarly, we should not expect any changes in time allocation among courses taken in one semester; that is, we should not expect students favoring (allocating more time towards) some course over the others, the grades of which could be improved by repeating in the following semesters. Again, since we do not directly observe the time spent on each course, we proximate time allocation among courses by using the performance among courses. Specifically, we create two individual-semester-level variables measuring the grade dispersion within a semester: the standard deviation of grades and the difference between the max and min grades in a semester.[23] We then regress each of the two variables on the policy indicator, controlling for individual fixed effects and academic term t fixed effects, and present the results in Panel 2, Table 5. We find that the grade dispersion does not vary by policy, which echoes the findings in Panel 1 and further demonstrates that students do not "giving up" on classes even when having the grade forgiveness option.

6.4 College Completion

6.4.1 Probability of Graduation and Time-to-Graduation

We have shown above that grade forgiveness policy incentivizes students to repeat more and take more challenging courses. One may wonder about the policy's ultimate effect on students' final outcome, graduation, which is the primary interest of most institutional research as well as a commonly used measure of students' success. In this section, we examine grade forgiveness's effect on students' probability of graduation, time to graduation, and probability of graduating with a STEM degree.

On average, the graduation rate during the observation period at BSU is 23.7% for the entry cohorts who first enrolled between 1990 and 2008, among which the

[23] We use the de-trended numerical grades to construct the two measures.

first-time (non-transfer) students' graduation rate is 15.3%. It is both common to transfer in and out of BSU and we are unable to observe students who transferred out from BSU to another institution. Thus, we cannot distinguish between dropouts and transfer-outs and are forced to treat all the unobserved outflow of students as non-graduates. Among the observed graduation of first-time students, the average time to graduate is 12 semesters (same as the median), and the 75th percentile and 95th percentile of the time to graduate are 14 and 19 semesters, respectively. Again, we use a sample including cohorts who entered the University from 1990 to 2008 to ensure that we observe the majority of the graduation events.

We first estimate the policy's impact on the probability of graduation by comparing the outcome between different entry cohorts, who were exposed to different levels of treatment of the policy. We cannot adopt the same estimation strategy in the previous sections, where the key independent variable is an indicator of the policy–varying at the individual-time level. Our outcome variables in this section are at the individual levels and we have to construct a new independent variable measuring the policy treatment. Considering that the first several semesters in college are the key semesters to explore the curriculum that will later determine college major, repeat courses, and in general, to "survive" college, we construct a series of treatment (dummy) variables, $Treated_X$, based on whether a student experience grade forgiveness in his/her first X semesters. Specifically, we define treatment variable, $Treated_X = 1$ for individuals (the treatment group) who were *continuously* exposed to the policy in their first X semester(s) and $Treated_X = 0$ for individuals (the control group) who were not exposed to the policy in any of their first X semester(s).

Panel 1 in Table 6 shows the policy's effect on the probability of graduation, where the dependent variable is an indicator of graduating from BSU and the independent variable of interest is the treatment dummy variable defined above. We control for a linear and a quadratic time trend based on the first entry semester, entry season fixed effects (i.e., fall, summer, spring), gender, and imputed composite SAT score. Standard errors are clustered at entry cohort level. Each column is a separate regression with a different $Treated_X$. In specific, column (1) shows the difference of the outcomes between cohorts (the treatment group) who had grade forgiveness in their first semester and cohorts (the control group) who did not have grade forgiveness in their first semester. Column (2) is the difference of the outcomes between cohorts who had grade forgiveness in their first 2 semester and cohorts who have no grade forgiveness in either of their first 2 semesters. The regression sample size gradually reduces from column (1) to (6) due to dropouts. Overall, we find a zero effect of grade forgiveness on the probability of graduating from BSU as first-time students.

Next, we estimate grade forgiveness's effect on time-to-graduation among the sub-sample of BSU graduates. To do so, we regress the number of years from first enrollment to graduation on $Treated_X$, controlling for the same set of variables as in panel 1. The estimates are shown in panel 2, Table 6. Overall, we see an increase in time-to-graduation as treatment time lengthens. Cohorts who experience grade forgiveness in their first 6 semesters take 0.53 more years to graduate, compared to cohorts who experience no grade forgiveness in their first 6 semesters, counting for $\sim 8\%$ increase in time-to-graduation from the 1995 entry cohort mean.

6.4.2 Probability of Graduating with A STEM Degree

Now that we have shown grade forgiveness's positive effect on enrollment in STEM courses, we expect the policy to have a positive effect on graduation in STEM. Among the sub-sample of students who *graduated* from BSU, we do find a significant increase in obtaining a STEM degree for those who were exposed to the policy in the early stages of their study relative to their counterparts. We run a similar set of regressions on the sub-sample of graduates by replacing the outcome variable with an indicator of obtaining a STEM degree and present the estimates in Table 7.[24] Panel 1 and 2 in Table 7 shows the policy's effect on the probability of obtain-

[24]The numbers of observations in Table 7 are different from the ones in Panel 2 of Table 6 because that a considerable portion of our data have missing values for major or fail to update the declared major due to unknown reasons, leaving graduating major as "Course of interest", "Generic Undergraduate Plan for Conversion", or "General-Undeclared". We exclude those cases in this analysis

ing a conservative STEM degree and a OPT-defined STEM degree. Each column is presented in the same fashion as the ones in Table 6.

Unsurprisingly, we see that the policy has a positive effect on the likelihood of obtaining a STEM degree among graduates, under both the broader definition from OPT and the conservative definition of STEM majors. We start observing the positive effect between cohorts who were exposed to the policy in their first 2 semesters versus cohorts who were not exposed to the policy in their first 2 semesters. When we compare groups who have a larger difference in treatment intensity, we see larger differences in probability of obtaining a STEM degree. For example, we see that cohorts who were continuously exposed to the policy in their first 4 semesters are more likely to obtain a conservative-STEM (OPT-STEM) degree by 5.49 (9.14) percentage points, relative to cohorts who were not exposed to the policy in any of their first 4 semesters. These effects translate to a 22% (27.6%) increase in OPT-STEM (conservative-STEM) degrees from the sample mean. The policy's effect gets as large as 33% when we compare cohorts who were 6 semesters apart in the policy treatment.

Taken all together, grade forgiveness incentivizes students to enroll, persist and succeed in STEM majors; however, it comes with a unneglectable price–taking longer to graduate. We do not intend to conduct a rigorous cost-benefit analysis but here are some numbers to consider. According to NSF, in 2019 (the most recent pre-pandemic data), STEM workers had higher median earnings ($55,000) than non-STEM workers ($33,000) with a bachelor's degree or higher.[25] BSU's in-state tuition and out-of-state tuition was $8,068 and $24,988 in 2019. Conservatively, assuming that the increased time-to-graduation is all resulted from the increased STEM graduates, the marginal cost of producing a STEM graduate is the additional tuition and the forgone earning of a non-STEM job of the increased portion of time-to-graduation. Our back-of-an-envelope calculation tells that, for example, those who experienced at least 4 semesters of grade forgiveness, he/she would pay off the

since they are not informative in defining STEM or non-STEM major/degree.

[25]Data is from NSF: The STEM Labor Force of Today.

additional cost of obtaining a STEM degree by about 4 years after graduation.[26]

6.4.3 STEM Degrees among Underrepresented Groups

We further examine the heterogeneity of the effects among demographic groups, women versus men and low-income background students versus high-income background students, by running the same specification in 7 on demographic sub-samples. Results are shown in Table D1 and D2.[27] Consistent with the policy's heterogeneous effects on STEM course enrollment among demographic groups, we find that the policy's positive effect on the probability of obtaining a conservatively defined STEM degree is stronger for men than for women and stronger for students from high-income background than for students from low-income background. This implies that, while grade forgiveness promotes STEM education, it might as well enlarge the existing gaps between STEM-underrepresented group and their counterparts. Moreover, since STEM degrees lead towards higher earning jobs, we can imagine a widening earning gap between genders and between students from two income backgrounds, adding to the known existing earning gaps among each group.

7 Conclusion

Whether college students should be given a second chance in coursework is a fundamental topic in institutional research. We study the so-called "Grade Forgiveness" policies in-depth for the first time by investigating changes in course repetition, curriculum choice, effort allocation, and graduation. We use a unique administrative dataset from Boise State University, which includes an observational period that covers two recent changes in grade forgiveness. We use a fixed effects model,

[26] Students who were continuously treated in their first 4 semesters are 9.14 percentage points more likely to obtain a STEM degree and taking 0.1895 years longer to graduate, relative to their counterparts. The marginal cost divided by the wage differences is $\frac{(33000+8068)*0.1895}{(55000-33000)*0.0914} = 3.87$.

[27] We run separate regressions, one on each demographic subgroup, instead of adding a group indicator and an interaction term of the group indicator and the treatment variable, because of concerns about the differential dropout timing among demographic subgroups. Differential timing of dropping out or transferring out between women and men and between low-income and high-income students will result in differential sorting effect in each treatment and control groups.

strengthened by the on-off-on policy variations, to estimate the policy's effects on college students' decision-making and outcomes throughout different stages of college.

We directly address the controversies of grade forgiveness in the higher education circle and go beyond that. We find that students are slacking-off but only to a minor extent under grade forgiveness, indicating the opponents' concern is trivial. What's more, the policy incentives students to take more challenging courses and STEM courses, which gear toward high-paying jobs. However, the less-prepared and STEM-underrepresented groups are less incentivized by the policy to initially take a STEM course, despite proponents' wishes. The policy has no impact on the overall graduation rate, but it increases the time to graduation by 5%. In the meantime, the probability of getting a STEM degree increases by up to 25% among the graduates. The price for getting higher-paying (STEM) degrees and higher GPAs is longer time to graduate, leaving it an empirical question that whether the improvements in ultimate success and increases associated with more STEM graduates may be worth the cost.

Our findings have generalizable policy implications on similar practices of providing a safety net for students' academic performance, such as allowing students to take the pass/fail grading options under certain scenarios. In 2018, the Massachusetts Institute of Technology initiated an experimental freshman grading policy, allowing up to three science core General Institute Requirements (GIRs) to be graded on a Pass/No Record basis as part of their Committee on the Undergraduate Program Experimental Grading Policy.[28] These type of policies could effectively help students to survive in college as well as nudge students to challenge themselves and explore more difficult courses and even majors. As the first study estimating the impacts of grade forgiveness policy, we call for more attention to be directed towards institutional policies on academic requirements and their impact on students.

[28]https://registrar.mit.edu/classes-grades-evaluations/grades/grading-policies/experimentalgrading-policy/entering-fall-2018

References

Ahn, T., P. Arcidiacono, A. Hopson, and J. R. Thomas (2019). Equilibrium grade inflation with implications for female interest in stem majors. Technical report, National Bureau of Economic Research.

Astorne-Figari, C. and J. D. Speer (2019). Are changes of major major changes? the roles of grades, gender, and preferences in college major switching. *Economics of Education Review 70*, 75–93.

Bandiera, O., V. Larcinese, and I. Rasul (2015). Blissful ignorance? a natural experiment on the effect of feedback on students' performance. *Labour Economics 34*, 13–25.

Bar, T., V. Kadiyali, and A. Zussman (2009). Grade information and grade inflation: The cornell experiment. *Journal of Economic Perspectives 23*(3), 93–108.

Bettinger, E. P. and B. T. Long (2009). Addressing the needs of underprepared students in higher education does college remediation work? *Journal of Human resources 44*(3), 736–771.

Borghans, L., J. J. Heckman, B. H. Golsteyn, and H. Meijers (2009). Gender differences in risk aversion and ambiguity aversion. *Journal of the European Economic Association 7*(2-3), 649–658.

Butcher, K. F., P. J. McEwan, and A. Weerapana (2014). The effects of an anti-grade-inflation policy at wellesley college. *Journal of Economic Perspectives 28*(3), 189–204.

Chen, K., Z. Hansen, and S. Lowe (2021). Why do we inflate grades? the effect of adjunct faculty employment on instructor grading standards. *Journal of Human Resources 56*(3), 878–921.

Denning, J. T., E. R. Eide, K. Mumford, R. W. Patterson, and M. Warnick (2022).

Why have college completion rates increased? *American economic journal: applied economics 14*(3), 1–29.

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.

Goodman, J., O. Gurantz, and J. Smith (2020). Take two! sat retaking and college enrollment gaps. *American Economic Journal: Economic Policy 12*(2), 115–58.

Goodman, S. (2016). Learning from the test: Raising selective college enrollment by providing information. *Review of Economics and Statistics 98*(4), 671–684.

Marx, J. and D. Meeler (2013). Strike four! do-over policies institutionalize gpa distortion. *Quality Assurance in Education* ,(,), ,.

Minaya, V. (2017). Do differential grading norms across fields matter for major choice? evidence from a policy change in florida. Technical report, Working Paper.

Minaya, V. (2020). Do differential grading standards across fields matter for major choice? evidence from a policy change in florida. *Research in Higher Education 61*(8), 943–965.

Ost, B. (2010). The role of peers and grades in determining major persistence in the sciences. *Economics of Education Review 29*(6), 923–934.

Owen, A. L. (2010). Grades, gender, and encouragement: A regression discontinuity analysis. *The Journal of Economic Education 41*(3), 217–234.

Rask, K. and J. Tiefenthaler (2008). The role of grade sensitivity in explaining the gender imbalance in undergraduate economics. *Economics of Education Review 27*(6), 676–687.

Reshwan, R. (2016). Does gpa matter when applying for a job?

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*(3), 601–644.

Stinebrickner, R. and T. R. Stinebrickner (2013). A major in science? initial beliefs and final outcomes for college major and dropout. *Review of Economic Studies 81*(1), 426–472.

Stinebrickner, T. and R. Stinebrickner (2012). Learning about academic ability and the college dropout decision. *Journal of Labor Economics 30*(4), 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.

Vigdor, J. L. and C. T. Clotfelter (2003). Retaking the sat. *Journal of Human resources 38*(1), 1–33.

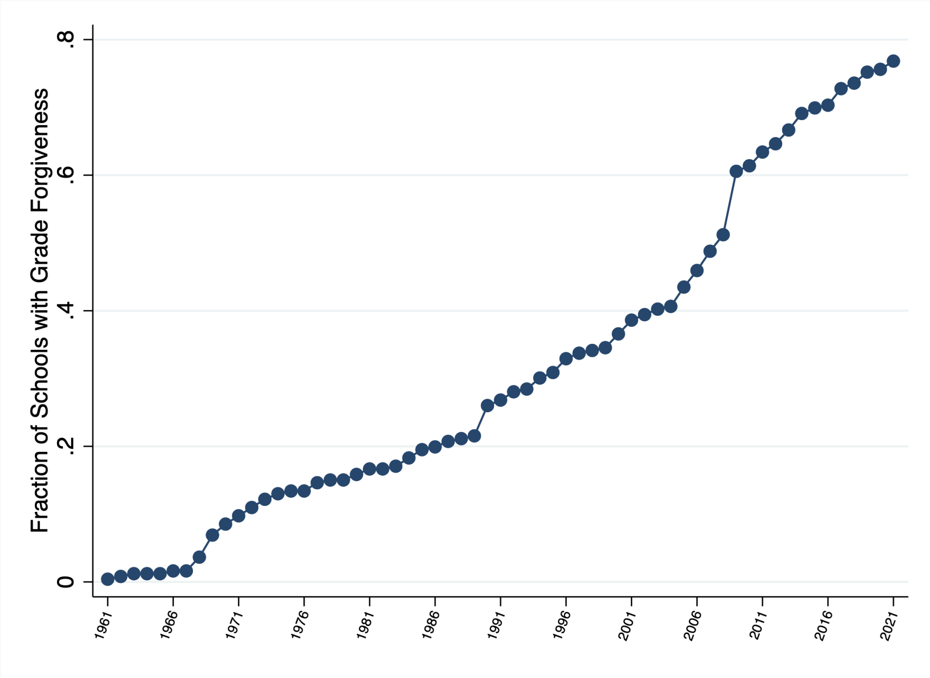


Figure 1: Fraction of Schools with Grade Forgiveness

Note: We collect year of implementation of grade forgiveness for the 380 IPEDS universities with above 10,000 enrollment. We exclude 133 schools missing a certain implementation time in this figure.

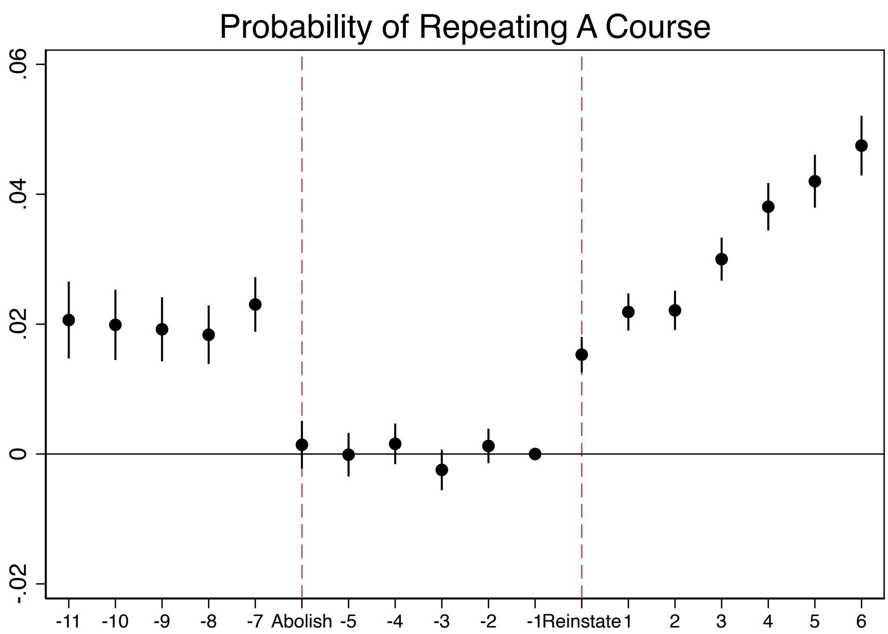
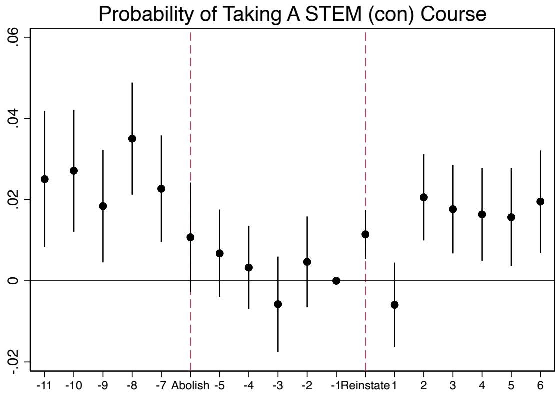
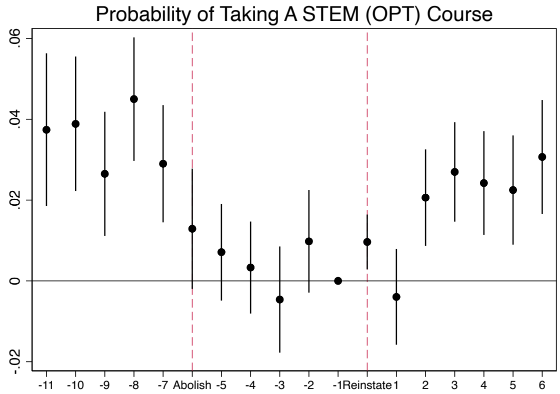


Figure 2: Grade Forgiveness's Effect on Probability of Repeating

Note: This figure plots the estimates of the standard difference-in-differences (DD) event study on grade forgiveness policy's effect on the probability of repeating a course. The variable of interest is an indicator which is equal to one when the years of observations is -11 (Fall 1990–Summer 1991), -10, -9, . . . , 5, 6 years relative to the year Fall 2001–Summer 2001, when grade forgiveness policy is reinstated. We control for gender, an indicator of having SAT scores, (imputed) SAT composite score, home zip-code median income, first enrollment year fixed effects, term-season fixed effects, academic term-t fixed effects, and course fixed effects. The horizontal dashed lines indicate the two time points when grade policy changed: Fall 1995 (abolish) and the Fall 2001 (reinstate).



(a) STEM (Conservative)



(b) STEM (OPT)

Figure 3: Grade Forgiveness's Effects on the Probability of Taking A STEM Course

Note: This figure plots the estimates of the standard difference-in-differences (DD) event study on grade forgiveness's effect on the probability of taking a STEM course. The dependent variable of is an indicator of a course being a STEM course based on the conservative definition of STEM (subfigure (a)) and based on the OPT definition of STEM (subfigure (b)). The variable of interest is an indicator which is equal to one when the years of observations is -11 (Fall 1990–Summer 1991), -10, -9, . . . , 5, 6 years relative to the year Fall 2001–Summer 2001, when grade forgiveness policy is reinstated. We control for gender, average SAT, home zip-code median income, a linear time trend of entry cohort, term-season fixed effects, academic term t fixed effects, and shares of STEM courses offered by BSU in each semester. The horizontal dashed lines indicate the two time points when grade policy changed: Fall 1995 (abolish) and Fall 2001 (reinstate).

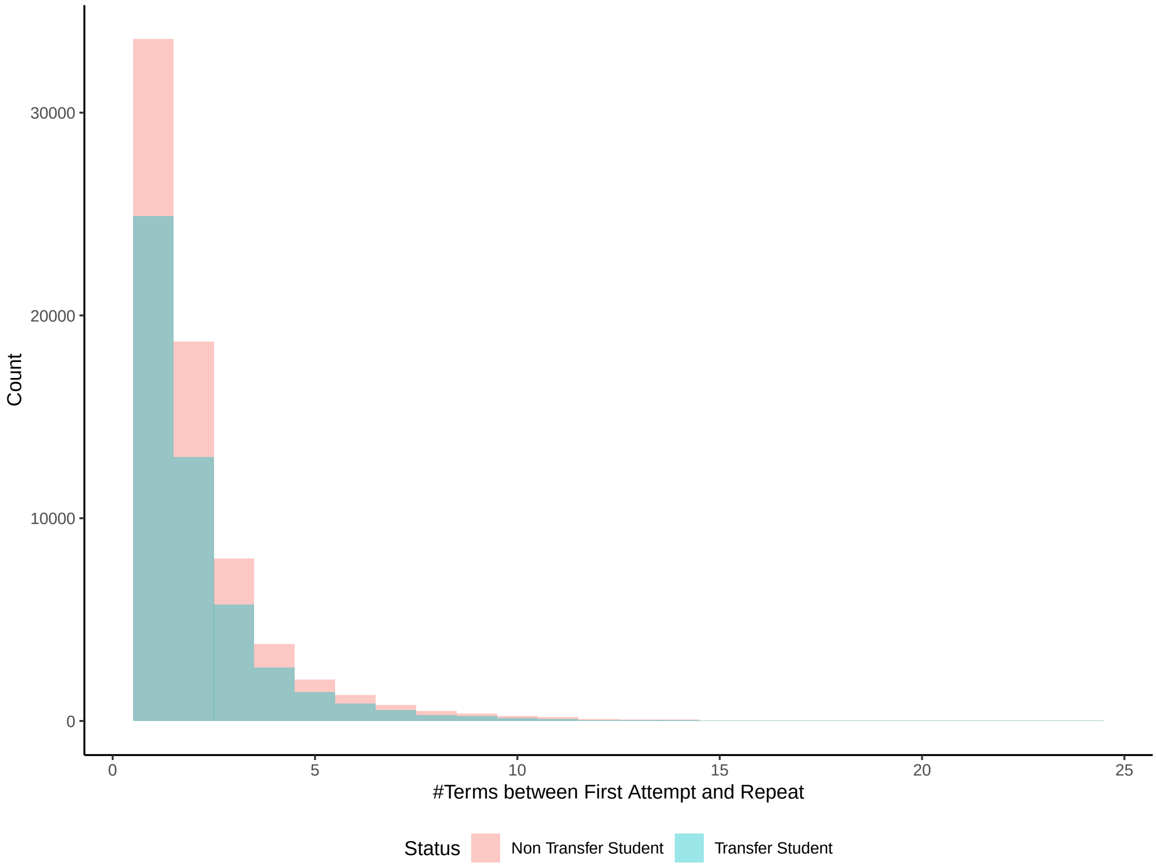


Figure 4: Density of #Terms between First Attempt and Repeat
Note: This figure shows the distribution of number of terms between a first attempt and the second attempt (repetition) by transfer students (blue) and non-transfer students (pink) separately. It indicates most of repeats happened in the next few semesters after the first-attempt.

Table 1: Summary Statistics on Key Variables

Variable	Mean	Std. Dev.	N
Person Level			
Female	.5415	.4983	169,722
SAT composite	1478.821	253.0717	86,628
Transfer student	.4504	.4975	170,812
Total #Repeat	.7919	1.9024	170,812
Exit GPA	2.62288	1.1136	170,812
Graduation Rate (Entry Cohort 1990-2008)	.2777	.4478	112,874
Graduate GPA at BSU	3.1750	.4548	47,614
# Years to Graduate	5.1804	3.4529	47,614
%STEM Degrees (conservative definition)	.2435	.4292	45,674
%STEM Degrees (OPT definition)	.3848	.4865	45,674
Term Level			
#Courses Attempted	3.729965	1.873759	806,577
#Credits Attempted	10.56368	4.868073	806,577
#Courses Earned	3.60792	1.842826	806,577
#Credits Earned	9.115406	5.124696	806,577
#Courses Withdrew	.1220454	.4083085	806,577
Course Level			
Repeat	.0480351	.2138405	2,769,204
Grade Points	2.760448	1.301253	2,769,204
STEM (conservative)	.256222	.4365459	2,769,204
STEM (OPT)	.3121514	.4633713	2,769,204
Repeated Grade			
By Semesters	1990-1995	1995-2001	2001-2008
A/B/C	5.08%	4.67%	5.79%
D/F/W	94.92%	95.33%	94.21%

Note: This table describes the data we used in the analysis, which covers students who enter BSU between 1990–2017. The upper, middle, and bottom panels shows person-level, term-level, and course-level statistics, respectively. ACT scores are converted to SAT composite scores when only ACT scores are available.

Table 2: Policy's Effect on the Probability of Repeating

	(1)	(2)	(3)	(4)
Policy	0.0160***	0.0240***	0.0207***	0.0227***
	(0.0006)	(0.0011)	(0.0011)	(0.0011)
Academic Progress F.E.	YES	YES	YES	YES
Course F.E.	YES		YES	YES
Individual F.E.		YES	YES	YES
Narrow Window: Fall 1990–Summer 2008				YES
Sample Mean (policy-off)	.0330675	.0330675	.0330675	.0330729
Observations	1279867	1279867	1279867	741122

Note: This table shows the policy's effect on the probability of retaking a course. The dependent variable is a binary indicator which equals to "1" if the course shown on a student's transcript is repeated (non-first time taking). The regression sample include all course level observations excluding incomplete grades, audits, or courses or sessions offering Pass/No Pass grades. Each column is a separate regression based on the specification in Equation 1. Column (1) includes academic term-t fixed effects and course fixed effects. Column (2) includes academic term-t and individual fixed effects. Column (3) includes all three fixed effects. Column (4) narrows the sample time window to Fall 1990–Summer 2008, focusing on a more relevant time window based on the two changes of grade forgiveness policy. The policy indicator equals 1 for calendar semesters Spring 1990–Summer 1995 and Fall 2001 and after; equals 0 elsewhere. Standard errors are clustered at the individual level.

Table 3: Policy's Effect on the Probability of Taking A STEM Course

	(1) Full Sample	(2) 1990–2008	(3) Excl. Retakes	(4) Entry Major non-STEM
Panel 1: Conservative STEM Definition				
Policy	0.0217***	0.0227***	0.0176***	0.0194***
	(0.0025)	(0.0024)	(0.0024)	(0.0025)
Sample Mean (policy-off)	.2128782	.2128941	.2084056	.1782125
Panel 2: OPT STEM Definition				
Policy	0.0241***	0.0239***	0.0182***	0.0218***
	(0.0026)	(0.0026)	(0.0026)	(0.0027)
Sample Mean (policy-off)	.2720815	.2720909	.2663202	.2305067
N	1279867	741203	706094	547307

Note: This table shows the policy's effect on the likelihood of taking a STEM course by four subsamples. The dependent variable of interest is an indicator of a course being a STEM course. Panel 1 and panel 2 uses two different definition of STEM: conservative definition and OPT-based definition. Column (1) covers the full sample; column (2) restrict the time window to 1990–2008; column (3) further excludes repeated courses and focuses on the initial takes of STEM courses, and column (4) includes only the initial takes of STEM courses for the first-time college students who did not declare a major when entering Boise State. Each regression includes individual fixed effects, academic term-t fixed effects, share of STEM courses offered in a certain semester, and a policy indicator. Standard errors are clustered at the individual level.

Table 4: Heterogeneity: STEM-underrepresented Groups

	(1) isRepeat	(2) Repeated	(3) STEM (OPT)	(4) STEM (conservative)
Panel 1: Female v.s. Male				
Policy	0.0128***	0.0184***	0.0320***	0.0269***
	(0.0015)	(0.0014)	(0.0042)	(0.0041)
Policy × Female	-0.0061***	-0.0044**	-0.0134**	-0.0088*
	(0.0019)	(0.0018)	(0.0057)	(0.0053)
Grade Points		-0.0644***		
		(0.0006)		
Sample Mean (policy-off)	.0268536	.0268536	.2748007	.2168394
Observations	721205	721205	721205	721205
Panel 2: Low-income v.s. High-income				
Policy	0.0103***	0.0150***	0.0302***	0.0273***
	(0.0013)	(0.0012)	(0.0039)	(0.0036)
Policy × Low Income	-0.0017	0.0021	-0.0114**	-0.0110**
	(0.0018)	(0.0018)	(0.0056)	(0.0052)
Grade Points		-0.0644***		
		(0.0005)		
Sample Mean (policy-off)	.0268588	.026975	.2747244	.2167862
Observations	722500	722500	722500	722500

Note: This table shows the policy's differential effects on the probability of repeating a course and on the probability of taking a STEM course between demographic groups. Panel 1 shows the policy's effect on female relative to male students and panel 2 shows the policy's effect on low-income students relative to high-income students. The independent variable in columns (1)–(4) are the probability of one course being a repeat/non-first-attempt course, the probability of one course being repeated, an indicator of a course being a STEM course by OPT definition and by conservative definition. Each regression includes individual fixed effects, academic term t fixed effects, course fixed effects (for columns 1 and 2), a policy indicator and an interaction term of the policy and gender (or low-income) indicator. The key difference between columns (1) and (2) is that column (1) reveals the more time-sensitive estimates on the policy's effect on repeating; while column (2) allows us to control for and condition on the first-attempt grades. We exclude the cross-policy repeats (the first-attempt and repeat-attempt are not under the same policy) to keep the sample consistent across the four columns. Columns (3) and (4) have the same regression specifications as regressions in Table 3, except adding the interaction terms. The regression sample includes transcript between 1990–2008. Standard errors are clustered at the individual level.

Table 5: Policy's Effect on Performance

	(1) All	(2) By Preparation Quartiles
Panel 1: First-Attempt Performance		
Policy	-0.0374***	0.0112
	(0.0077)	(0.0252)
Policy × Quartile 2		-0.0489*
		(0.0278)
Policy × Quartile 3		-0.0708**
		(0.0288)
Policy × Quartile 4 (Best-prepared)		-0.0536*
		(0.0291)
Sample Mean	2.615789	
Observations	706011	693928

Panel 2: Grades Variation within Semester

	SD of Grades	Max-Min Grade
Policy	-0.0048	0.0079
	(0.0045)	(0.0103)
Observations	195927	195927

Note: This table shows the policy's effect on students' performance. Panel 1 shows the policy's effect on students' first-attempt numerical grades (ranges 0-4), where the regression sample include only the first-attempt (non-repeated) course records. The dependent variable is the de-trended numerical grades using the method described in Section 6.3. Each regression includes individual fixed effects, academic term t fixed effects, course fixed effects, number of students enrolled the section, and the policy dummy. Column 1 shows the overall on-average estimate and column 2 shows the by preparedness estimates. The baseline group in column 2 is the least-prepared group, defined by the first quartile of the entry semester GPA. Panel 2 shows the policy's effect on students' grades variation within a semester, which measured by the standard deviation (column 1) and the max-min difference (column 2) of all grades within a semester. Both the standard deviation and the max-min gap are constructed using the de-trended numerical grades. Each regression includes individual fixed effects, academic term t fixed effects, and a policy indicator. All regressions use transcript between 1990–2008. Standard errors are clustered at the individual level.

Table 6: Probability of Graduating from BSU

	(1) First 1 Term	(2) First 2 Terms	(3) First 3 Terms	(4) First 4 Terms	(5) First 5 Terms	(6) First 6 Terms
Panel 1: Probability of Graduation						
$Treated_X$	0.0171	0.0134	0.0135	0.0035	0.0085	0.0203
	(0.0126)	(0.0153)	(0.0234)	(0.0323)	(0.0394)	(0.0405)
Mean of Entry Cohort 1995	.1692607					
Control Group Mean	.1945891	.2530396	.3466471	.4336536	.4962465	.5481639
Observations	45862	34443	22416	17361	13480	11333
Panel 2: Years to Graduation (Graduated Sub-sample)						
$Treated_X$	-0.2577**	-0.1035	0.0722	0.1895*	0.3764***	0.5312***
	(0.1016)	(0.1130)	(0.1028)	(0.1057)	(0.1118)	(0.1323)
Mean of Entry Cohort 1995	6.158301					
Control Group Mean	6.626732	6.442745	6.339001	6.260529	6.1352	6.038986
Observations	9870	9427	8477	8051	7331	6896

Note: This table shows the policy's effect on students' probability of graduation. This sample include first-time college students who first entered BSU between 1990 and 2008. The outcome variable in Panel 1 is an indicator of graduating from BSU and the outcome variable in Panel 2 is a continuous variable of years to graduation (for the graduated subsample). The regression sample is individual-level and the treatment group in columns (1)-(4) are: continuously being treated by the policy for the first 1, 2, 3, and 4 semester(s); while the corresponding control group for each treatment group in columns (1)-(4) are *never* treated by the policy in the first 1, 2, 3, and 4 semester(s). Each regression controls for individual's average SAT score, gender, home zip-code median income, a linear trend of the entry year, a squared trend of the entry year, and an indicator of the entry term season (spring, summer. In order to control for SAT score and not reduce sample size, we impute SAT scores by assigning zero to the missing SAT scores and including a dummy for missing SAT. We have tried another specification that does not include SAT control, which delivers qualitatively similar estimates. Standard errors are clustered at the entry cohort level.

Table 7: Probability of Obtaining STEM Degrees from BSU, Graduates Subsample

	(1) First 1 Term	(2) First 2 Terms	(3) First 3 Terms	(4) First 4 Terms	(5) First 5 Terms	(6) First 6 Terms
Panel 1: Dependent Variable: Indicator of Conservative STEM Degrees						
$Treated_X$	0.0056	0.0107	0.0405**	0.0549***	0.0571***	0.0656***
	(0.0247)	(0.0254)	(0.0152)	(0.0130)	(0.0151)	(0.0142)
Mean of Entry Cohort 1995	.2372					
Control Group Mean	.2887	.2844	.2583	.2424	.2408	.2340
Observations	6421	6269	5763	5570	5107	4869
Panel 2: Dependent Variable: Indicator of OPT STEM Degrees						
$Treated_X$	0.0407	0.0443*	0.0774***	0.0914***	0.0907***	0.1071***
	(0.0251)	(0.0256)	(0.0168)	(0.0151)	(0.0207)	(0.0159)
Mean of Entry Cohort 1995	.2881					
Control Group Mean	.3698	.3659	.3451	.3288	.3357	.3198
Observations	6421	6269	5763	5570	5107	4869

Note: This table shows the policy's effect on the probability of obtaining a STEM degree (among graduates). The regression sample include first-time students who entered BSU between 1990 and 2008. The outcome variable in all columns is an indicator of obtaining a STEM degree, where the definition of a STEM degree is conservative (panel 1) and OPT (panel 2). The regression sample is individual-level and the treatment group in columns (1)-(6) are: continuously being treated by the policy for the first 1, 2, 3, 45 and 6 semester(s); while the corresponding control group for each treatment group in columns (1)-(6) are *never* treated by the policy in the first 1, 2, 3, 4, 5, and 6 semester(s). Each regression controls for individual's average SAT score, gender, home zip-code median income, a linear trend of the entry year, a squared trend of the entry year, and an indicator of the entry term season (spring, summer, or fall). Standard errors are clustered at the entry cohort level.

A Proof of the Theoretical Implication

A.1 Difficulty Choice

To solve the optimal choice under each policy, we first expand the utility functions and then take the first order condition with respect to d, for each utility function under each policy. With the envelope theorem, we will take t as given when solving d; vise versa. Again, we assume the expected grade of the second attempt (retake), $E[G]$, and the expected cost for retaking, $E[C]$, are constant. We denote $g(d,t)$ as g for brevity.

The expanded utility function for policy = 0 is:

$$U = f(g)\{d * \frac{E[G] + g(d,t)}{2} - c(d,t) - E[C]\} + (1 - f(g)\{d * g(d,t) - c(d,t)\} \quad (4)$$

Take the first order condition (FOC) with respect to (w.r.t.) d,

$$\frac{\partial U}{\partial d} = f(g) * \{\frac{E[G] + g}{2} - c'(d) + \frac{dE[G]}{2}\} + f'(g)g'(d)\{d * \frac{E[G] + g}{2} - c - E[C]\} \\ - f'(g)g'(d)[d * g - c] + (1 - f(g))[g + d * g'(d) - c'(d)] = 0 \quad (5)$$

Denote the optimal difficulty as $d = d_0^*$. Simplify the equation to get:

$$\frac{\partial U}{\partial d} = \frac{1}{2}\{f'(g)g'(d)dE[G] - f'(g)g'(d)dg + f(g)E[G] - f(g)g - f(g)dg'(d)\} \\ - \{f'(g)g'(d)E[C] - g - dg'(d) + c'(d)\} = 0 \quad (6)$$

The expanded utility function for policy = 1 is:

$$U = f(g)\{d * E[G] - c(d,t) - E[C]\} + (1 - f(g)\{d * g(d,t) - c(d,t)\} \quad (7)$$

FOC w.r.t. d, for policy = 1:

$$\frac{\partial U}{\partial d} = f(g) * \{E[G] - c'(d)\} + f'(g)g'(d)\{dE[G] - c - E[C]\}$$
$$- f'(g)g'(d)[dg - c] + (1 - f(g))\{g + dg'(d) - c'(d)\} = 0 \quad (8)$$

where the optimal $d = d_1^*$. Simplify the equation to get:

$$\frac{\partial U}{\partial d} = \{f'(g)g'(d)dE[G] - f'(g)g'(d)dg + f(g)E[G] - f(g)g - f(g)dg'(d)\}$$
$$- \{f'(g)g'(d)E[C] - g - dg'(d) + c'(d)\} = 0 \quad (9)$$

We can re-write the two FOCs above as follows:

$$\frac{1}{2}A(d_0^*) - B(d_0^*) = 0;$$

$$A(d_1^*) - B(d_1^*) = 0;$$

where $A() = \{f'(g)g'(d)dE[G] - f'(g)g'(d)dg + f(g)E[G] - f(g)g - f(g)dg'(d)\}$ and $B() = \{f'(g)g'(d)E[C] - g - dg'(d) + c'(d)\}$.

$$\frac{\partial A}{\partial d} = \{f'(g)g'(d)g'(d)d(E[G] - g) + f'(g)g''(d)d(E[G] - g) +$$
$$2f'(g)g'(d)(E[G] - g) - 2f'(g)g'(d)g'(d)d - 2f(g)g'(d) - f(g)dg''(d)\} = 0 \quad (10)$$

$$\frac{\partial B}{\partial d} = f'(g)g'(d)g'(d)E[C] + f'(g)g''(d)E[C] - 2g'(d) - dg''(d) + c''(d) = 0 \quad (11)$$

By assumption, $g'(d) < 0$, $g''(d) < 0$, $f'(g) < 0$, $c'(d) > 0$, $-g''(d) > g'(d)^2$. Again, we assume the expected grade is larger than the first attempted grade, i.e., $E[G] > g(d)$, for a student to be willing to repeat a course. We can derive that function $A()$ is strictly increasing in d and function $B()$ is strictly increasing in d. Finally, we obtain $d_1^* > d_0^*$. Therefore, student will be more likely to choose a difficult course under the Grade Forgiveness policy.

A.2 Time/Effort Choice

Since the weight of the first-attempt grade is zero under policy = 1, it is trivial to prove that the optimal study time under policy = 1 is $t_1^* = 0$. While the optimal study time under policy = 0 is $t_0^* > 0$, we can conclude that $t_1^* < t_0^*$.

A.3 Time/Effort Allocation

Students choose to allocate time among courses they enroll in a semester. Each student has a certain endowment of time to study, t, which does not change by the policy. To illustrate the time allocation among courses taken in the same semester, let's consider the simplest case that a student takes two courses in this period and allocates time on each course, t_1 and t_2, i.e., $t_1 + t_2 = t$. Assume that the two courses have different difficulty level, $d_1 > d_2$. The utility function can be written as the summation of the utility gained from the two courses:

For policy = 0:

$$\begin{aligned} U(g,c,d,t) = f(g_1) * \{d_1 \frac{E[G_1] + g(d_1,t_1)}{2} - c(d_1,t_1) - E[C]\} \\ +\{1 - f(g_1)\} * \{d_1 g(d_1,t_1) - c(d_1,t_1)\} \\ +f(g_2) * \{d_2 \frac{E[G_2] + g(d_2,t_2)}{2} - c(d_2,t_2) - E[C]\} \\ +\{1 - f(g_2)\} * \{d_2 g(d_2,t_2) - c(d_2,t_2)\} \end{aligned} \tag{12}$$

For policy = 1:

$$\begin{aligned} U(g,c,d,t) = f(g_1) * \{d_1 E[G_1] - c(d_1,t_1) - E[C]\} \\ +\{1 - f(g_1)\} * \{d_1 g(d_1,t_1) - c(d_1,t_1)\} \\ +f(g_2) * \{d_2 E[G_2] - c(d_2,t_2) - E[C]\} \\ +\{1 - f(g_2)\} * \{d_2 g(d_2,t_2) - c(d_2,t_2)\} \end{aligned} \tag{13}$$

For a student who chooses the optimal time spent on course 1, t_1, we take the first order condition of the utility function under each policy with respect to t_1. By assumption, $t_2 = t - t_1$. We obtain the FOC as:

For policy = 0, FOC w.r.t. t_1,

$$
\begin{aligned}
\frac{\partial U}{\partial t_1} &= f(g_1) * \{d_1 \frac{g'(t_1)}{2} - c'(t_1)\} + \{1 - f(g_1)\} * \{d_1 g'(t_1) - c'(t_1)\} \\
&+ f'(g)g'(t_1) * \{d_1 \frac{E[G_1] + g(d_1, t_1)}{2} - c(d_1, t_1) - E[C]\} \\
&- \{f'(g)g'(t_1)\} * \{d_1 g(d_1, t_1) - c(d_1, t_1)\} \\
&+ f(g_2) * \{d_2 \frac{-g'(t_1)}{2} + c'(t_1)\} + \{1 - f(g_2)\} * \{-d_2 g'(t_1) + c'(t_1)\} \\
&- f'(g)g'(t_1) * \{d_2 \frac{E[G_2] + g(d_2, t_1)}{2} - c(d_2, t - t_1) - E[C]\} \\
&+ \{f'(g)g'(t_1)\} * \{d_2 g(d_2, t - t_1) - c(d_2, t - t_1)\} \\
&= 0
\end{aligned}
\tag{14}
$$

Simplify to be:

$$
\begin{aligned}
\frac{\partial U}{\partial t_1} &= \{-d_1 \frac{f(g_1)g'(t_1)}{2} + d_1 g'(t_1)\} + \{d_2 \frac{f(g_2)g'(t_1)}{2} - d_2 g'(t_1)\} \\
&+ \{d_1/2 * f'(g)g'(t_1)E[G_1] - d_2/2 * f'(g)g'(t_1)g(d_1, t_1)\} \\
&+ \{-d_2/2 * f'(g)g'(t_1)E[G_2] + d_2/2 * f'(g)g'(t_1)g(d_2, t_1)\} \\
&= 0
\end{aligned}
\tag{15}
$$

For policy = 1, FOC w.r.t. t_1,

$$
\begin{aligned}
\frac{\partial U}{\partial t_1} &= -f(g_1) * \{c'(t_1)\} + \{1 - f(g_1)\} * \{d_1 g'(t_1) - c'(t_1)\} \\
&+ f'(g)g'(t_1) * \{d_1 E[G_1] - c(d_1, t_1) - E[C]\} \\
&+ \{-f'(g)g'(t_1)\} * \{d_1 g(d_1, t_1) - c(d_1, t_1)\} \\
&+ f(g_2) * \{c'(t_1)\} + \{1 - f(g_2)\} * \{-d_2 g'(t_1) + c'(t_1)\} \\
&- f'(g)g'(t_1) * \{d_2 E[G_2] - c(d_2, t - t_1) - E[C]\} \\
&+ \{f'(g)g'(t_1)\} * \{d_2 g(d_2, t - t_1) - c(d_2, t - t_1)\} \\
&= 0
\end{aligned}
\tag{16}
$$

Simplify to be:

$$
\begin{aligned}
\frac{\partial U}{\partial t_1} &= \{-d_1 f(g_1) g'(t_1) + d_1 g'(t_1)\} + \{+d_2 f(g_2) g'(t_1) - d_2 g'(t_1)\} \\
&+ \{d_1 * f'(g)g'(t_1)E[G_1] - d_2 * f'(g)g'(t_1)g(d_1, t_1)\} \\
&\{-d_2 * f'(g)g'(t_1)E[G_2] + d_2 * f'(g)g'(t_1)g(d_2, t_1)\} \\
&= 0
\end{aligned}
\tag{17}
$$

Re-arranging (16) and (18) we obtain:

$$
\begin{aligned}
&\frac{1}{2}\{-d_1 f(g_1) + d_1 f'(g)E[G_1] - d_1 f'(g)g(d_1, t_1) \\
&+ d_2 f(g_2) - d_2 f'(g)E[G_2] + d_2 f'(g)g(d_2, t_2)\} \\
&= \frac{1}{2} LHS(t_1^A) = [d_2 - d_1]
\end{aligned}
\tag{18}
$$

and

$$
\begin{aligned}
&\{-d_1 f(g_1) + d_1 f'(g)E[G_1] - d_1 f'(g)g(d_1, t_1) \\
&+ d_2 f(g_2) - d_2 f'(g)E[G_2] + d_2 f'(g)g(d_2, t_2)\} \\
&= LHS(t_1^F) = [d_2 - d_1]
\end{aligned}
\tag{19}
$$

Compare (19) and (20), we can find $\frac{1}{2}LHS(t_1^A) = LHS(t_1^F)$, where t_1^A and t_1^F are the optimal t_1 choice under the averaging policy (A) and the Forgiveness policy(F). By assumption, $d_1 > d_2$, $g'(t) > 0$, $g''(t) < 0$, $E[G_1] > g(d_1, t_1)$, and $E[G_2] >$

$g(d_2, t_2)$. We can find that the function $LHS(t_1)$ is decreasing in t_1. Finally, we can conjecture $t_1^A < t_1^F$. If we define the gap between the time allocated in the two courses as $t_1 - t_2 = t_1 - (t - t_1) = 2t_1 - t$, we can conclude that the gap increases as t_1 increases.

A.4 Probability of Repeating and the Threshold

We now assume that a student has already enrolled and completed a course and realized the grade, g, and the cost, c, at the first-attempt and has a constant expectation on the second-attempt grade, $E[G]$, and a constant expectation on the second-attempt cost, $E[C]$. The "Grade Forgiveness" policy basically changes the calculation of the GPA if the student chooses to repeat. That is, when repeating happens during the time that policy is off, i.e., Forgiveness = 0, the final grade is simply an average between the first-attempt grade and the second-attempt grade. When repeating happens during the time that policy is on, i.e., Forgiveness = 1, the final grade is the second-attempt grade. Here, we assume the difficulty and time allocation as given for brevity to derive the probability of repeating.

$$U(g, c) = \begin{cases} g - c, & \text{if Not Repeat} \\ \frac{E[G]+g}{2} - c - E[C] & \text{if Repeat under Forgiveness = 0} \\ E[G] - c - E[C], & \text{if Repeat under Forgiveness = 1} \end{cases} \tag{20}$$

The condition for the student chooses to repeat is the utility of repeating is larger than the utility of not repeating. Let's denote the utility of repeating as U^R and the utility of not repeating as U^{NR}.

$$U^R(g, c) > U^{NR}(g, c) \equiv \begin{cases} \frac{E[G]+g}{2} - c - E[C] > g - c, & \text{if Forgiveness = 0} \\ E[G] - c - E[C] > g - c, & \text{if Forgiveness = 1} \end{cases} \tag{21}$$

We further denote the realized first-attempt grade under each policy as g_0 and

g_1 and simplify the inequalities above as:

$$U^R(g,c) > U^{NR}(g,c) \equiv \begin{cases} E[G] - 2E[C] > g_0, & \text{if Forgiveness = 0} \\ E[G] - E[C] > g_1, & \text{if Forgiveness = 1} \end{cases} \tag{22}$$

Thus the probability of repeating a course can be written as:

$Pr(E[G] - 2E[C] > g_0)$ and $Pr(E[G] - 2E[C] = g_0) = 0$ if Forgiveness = 0;

$Pr(E[G] - E[C] > g_1)$ and $Pr(E[G] - E[C] = g_1) = 0$ if Forgiveness = 1.

By assumption, the belief on the expected grade, $E[G]$, and the expected cost, $E[C]$, is unchanged. The probability of repeating a course is determined by the inequality. We can easily obtain that:

(1) If the first-attempt grades are constant under different policies, $g_0 = g_1 = g$, the region of grades for one to prefer repeating over not is larger under the policy Forgiveness = 1 than under the policy Forgiveness = 0: $E[C] < E[G] - g < 2E[C]$. Thus, the average probability of repeating under Forgiveness = 1 will be higher than the probability of repeating under Forgiveness = 0.

(2) The threshold (highest) grade to repeat under Forgiveness = 1 is higher than the threshold (highest) grade to repeat under Forgiveness = 0: $g_1 > g_0$, and the difference between the two threshold grades is restricted as $g_1 - g_0 \leq E[C]$.

B Robustness Checks

B.1 Policy's Effect on Repeat: turn-on versus turn-off

We exploited two variations of the grade replacement policy in our main analysis. Here, we further investigate the variation-specific effects regarding the policy's turn-on and turn-off by showing the effects on two cohort windows: Fall 1990–Summer 2001 and Fall 1995–Summer 2006. We run the same regression specification in columns (3) and (4) of Table 2 on the two sub-samples. The results in Table B1 show that the probability of repeating is significantly higher when the policy is in force and it holds true for both the cohorts who experienced a policy turn-off and the cohorts who experienced a policy turn-on. This evidence provide convincing support for our identification strategy that the policy's effects are not subject to specific time or cohort-sensitive.

Table B1: Policy's Effect on Repeat: turn-on versus turn-off

	Policy on-off Fall 1990–Summer 2001	Policy off-on Fall 1995–Summer 2006
Policy	0.0190***	0.0146***
	(0.0017)	(0.0018)
N	367918	593297

Note: This table shows the policy's effect on the probability of repeating by two cohorts: cohorts under the first change of the policy (policy on to off period) and cohorts under the second change of the policy (policy off to on period). The regression sample in each column are sub-samples from regression sample in column (3) and (4) of Table 2. Specifically, column (1) and (2) covers Fall 1990 to Summer 2001 observations and columns (3) and (4) cover Fall 1995–Summer 2006 observations. The independent variable is an indicator of whether a course shown on a student's transcript is a repeat (second time taking). Each regression includes individual fixed effects, academic term t fixed effects, course fixed effects, number of credits attempted in the semester, and a policy indicator where policy = 1 for calendar semesters Spring 1990–Summer 1995 and Fall 2001 and after; policy = 0 elsewhere. Standard errors are clustered at the individual level.

Table B2: Policy's Effect on STEM Course Taking: turn-on versus turn-off

	Policy on-off Fall 1990–Summer 2001		Policy off-on Fall 1995–Summer 2006	
	(1)	(2)	(3)	(4)
Policy	0.0087**	0.0049	0.0200***	0.0171***
	(0.0039)	(0.0039)	(0.0037)	(0.0038)
Sample Mean	.2188818	.2138401	.2189143	.2138647
N	368021	355886	593401	563090
Policy	0.0131***	0.0083**	0.0142***	0.0097**
	(0.0042)	(0.0042)	(0.0040)	(0.0041)
Sample Mean	.2779074	.2716199	.2779779	.271698
N	368021	355886	593401	563090

Note: This table shows the policy's effect on the probability of taking a STEM course by two cohorts: cohorts under the first change of the policy (policy on to off period) and cohorts under the second change of the policy (policy off to on period). The regression sample in each column are sub-samples from regression sample in column (3) and (4) of Table 3. Specifically, column (1) and (2) covers Fall 1990 to Summer 2001 observations and columns (3) and (4) cover Fall 1995–Summer 2006 observations. The independent variable is an indicator of whether a course shown on a student's transcript is a STEM course. Each regression includes individual fixed effects, academic term t fixed effects, number of credits attempted in the semester, and a policy indicator where policy = 1 for calendar semesters Spring 1990–Summer 1995 and Fall 2001 and after; policy = 0 elsewhere. Standard errors are clustered at the individual level.

B.2 Alternative Measures of Difficulty Courses

We construct a series of alternative measures *difficulty* by collapsing all grades offered by the same course over different sections and different semesters to the mean. This measure is not subject to the policy change or any instructor-specific effects. We first use the de-trended numerical grades, described in Section 6.3 to construct the mean grades for each course; then use the product of the mean grade and number of credit of each course to construct a secondary course-level difficulty measure.

Columns (1) and (2) in Table B3 shows the estimation results. The results indicate that under grade forgiveness policy, students chose to enroll in courses that are more harshly graded. Taking the number of credits of each course in to account, the outcome variables in column (2) proximate total difficulty, showing the same results. Columns (3) and (4) shows that students are also taking more credits and more number of courses under grade forgiveness.

Table B3: Policy's Effect on Choice of Difficulty: Alternative Measures

	(1) Measure 1: Grades	(2) Measure 2: Grades $\times$ Credits	(3) Measure 3: #Credits/Term	(4) Measure 4: #Courses/Term
Policy	-0.0175***	-0.0332***	0.2504***	0.0984***
	(0.0008)	(0.0021)	(0.0384)	(0.0149)
Sample Mean	.0209905	.0352624	10.29701	3.649454
Observations	679225	679225	196770	196770

Note: This table shows the policy's effect on course-level (panel 1) and term-level (panel 2) measures of difficulty. Outcome variables in columns (1)–(4) of Panel 1 are course-level difficulty measured by fraction of letter grade D or F by course, fraction of letter grade D or F or W by course, the product of the fraction of letter grade D or F by course and the course credits, and the product of the fraction of letter grade D or F or W by course and the course credits. Outcome variables in columns (1)–(3) of Panel 2 are: number of credits attempted in each term, number of courses attempted in each term, sum of the product of the fraction of letter grade D or F (or W) by course and the course credits in each term (column 4). Each regression includes individual fixed effects, academic term t fixed effects, and a policy indicator where policy = 1 for calendar semesters Spring 1990–Summer 1995 and Fall 2001 and after; policy = 0 elsewhere. Standard errors are clustered at the individual level.

C Heterogeneity of the Effects among Students with Different Academic Preparedness

We have also explored how students with different levels of academic preparation respond to the policy in terms of choosing STEM courses. We conduct a similar analysis as in Table 3 by differentiating students by their first term GPA. Specifically, we regress the average course difficulty and number of credits attempted on the interaction of policy and the quartiles of the first term GPA. We use the first-term GPA as a measure of students' academic preparation instead of using SAT or ACT scores because there are about 35% of students do not have an SAT or ACT score on file.

Table C1 shows the results for this exercise. Quartiles 1–4 represent students with the lowest–highest academic preparation (first term GPA). Overall, there are no significant differences in the policy's effect on taking a STEM course across the 4 groups of students. However, by including only the first 6 semesters (similar results for only including the first 4 semesters), when students are still exploring the curriculum and when course repeats happen most frequently, we see that the policy's effect is driven by students who are better prepared. In specific, students in quartiles 2, 3, and 4 have a statistically significant effect on choosing a STEM course than students in quartile 1. The magnitudes increase slightly when GPA increases, but are not statistically significant. Additionally, students whose GPA is in quartile 4 are driving the increased number of credits per course. Overall, it is evident that the policy's nudge is especially pronounced among students who have stronger academic preparation.

Table C1: Heterogeneity: Policy's Nudge across Academic Preparation

	(1) STEM Conservative	(2) STEM OPT	(3) #Credits
Policy (Baseline: least prepared)	-0.0062	0.0012	0.0019
	(0.0046)	(0.0050)	(0.0093)
Policy $\times$ Quartile 2	0.0249***	0.0205***	0.0193
	(0.0068)	(0.0073)	(0.0130)
Policy $\times$ Quartile 3	0.0254***	0.0178**	0.0139
	(0.0085)	(0.0091)	(0.0158)
Policy $\times$ Quartile 4	0.0244***	0.0207**	0.0328*
	(0.0087)	(0.0093)	(0.0172)
N	873210	873210	873210

Note: This table shows the policy's different effects on course choice by academic preparation. The sample includes the first 6 semesters of all non-transfer students. The independent variable in column (1) is an indicator of STEM course (conservative definition), column (2) is an indicator of STEM course (OPT definition), and column (3) is the number of credits of each course. Students' academic preparation is defined by their first term GPA and separated into quartiles, where quartile 1 with the lowest GPA and quartile 4 with the highest GPA. Each regression includes individual fixed effects, academic term t fixed effects, number of courses offered in each semester, number of peers in each semester and each major, a policy indicator, and interaction of policy and the quartile. Policy = 1 for calendar semesters Spring 1990–Summer 1995 and Fall 2001 and after; policy = 0 elsewhere. Standard errors are clustered at the individual level.

D Additional Tables and Figures

Table D1: Policy's Effect on Probability of Obtaining STEM Degrees by Gender

	(1) First 1 Term	(2) First 2 Terms	(3) First 3 Terms	(4) First 4 Terms	(5) First 5 Terms	(6) First 6 Terms
Panel 1: Conservative STEM, Women						
$Treated_X$	0.0028	0.0072	0.0240	0.0353*	0.0324	0.0405*
	(0.0232)	(0.0240)	(0.0216)	(0.0189)	(0.0196)	(0.0208)
#Women	3585	3510	3228	3118	2851	2708
Panel 2: Conservative STEM, Men						
$Treated_X$	0.0068	0.0119	0.0568	0.0744**	0.0849**	0.0939**
	(0.0424)	(0.0431)	(0.0354)	(0.0338)	(0.0376)	(0.0395)
#Men	2836	2759	2535	2452	2256	2161
Panel 3: OPT STEM, Women						
$Treated_X$	0.0506*	0.0546*	0.0761***	0.0858***	0.0815**	0.0966***
	(0.0279)	(0.0285)	(0.0278)	(0.0273)	(0.0365)	(0.0315)
#Women	3585	3510	3228	3118	2851	2708
Panel 4: OPT STEM, Men						
$Treated_X$	0.0249	0.0267	0.0728*	0.0918**	0.0965**	0.1137***
	(0.0444)	(0.0453)	(0.0369)	(0.0351)	(0.0400)	(0.0416)
#Men	2836	2759	2535	2452	2256	2161

Note: This table shows the policy's heterogeneous effects on the probability of obtaining a STEM degree among genders. The regression sample include first-time students who entered BSU between 1990 and 2008. The outcome variable in all columns is an indicator of obtaining a STEM degree, where the definition of a STEM degree is conservative in panels 1 and 2 and OPT in panels 3 and 4. We compare the policy's effects across genders by running the same regression separately on the women's (panels 1 and 3) and men's sub-sample (panels 2 and 4). The regression sample is at individual-level and the treatment group in columns (1)-(6) are: continuously being treated by the policy for the first 1, 2, 3, 4, 5 or 6 semester(s); while the corresponding control group for each treatment group in columns (1)-(6) are *never* treated by the policy in the first 1, 2, 3, 4, 5, or 6 semester(s). Each regression controls for individual's average SAT score, gender, home zip-code median income, a linear trend of the entry year, a squared trend of the entry year, and an indicator of the entry term season (spring, summer, or fall). Standard errors are clustered at the entry cohort level.

Table D2: Policy's Effect on Probability of Obtaining STEM Degrees by Income

	(1) First 1 Term	(2) First 2 Terms	(3) First 3 Terms	(4) First 4 Terms	(5) First 5 Terms	(6) First 6 Terms
Panel 1: Conservative STEM, Low-income						
$Treated_X$	-0.0274	-0.0190	0.0130	0.0250	0.0194	0.0270
	(0.0323)	(0.0313)	(0.0229)	(0.0223)	(0.0255)	(0.0253)
#Low-income	2891	2815	2556	2471	2247	2125
Panel 2: Conservative STEM, High-income						
$Treated_X$	0.0335	0.0355	0.0625***	0.0790***	0.0866***	0.0936***
	(0.0234)	(0.0256)	(0.0186)	(0.0159)	(0.0186)	(0.0199)
#High-income	3530	3454	3207	3099	2860	2744
Panel 3: OPT STEM, Low-income						
$Treated_X$	0.0116	0.0182	0.0516**	0.0613***	0.0533*	0.0633**
	(0.0280)	(0.0270)	(0.0212)	(0.0210)	(0.0312)	(0.0255)
#Low-income	2891	2815	2556	2471	2247	2125
Panel 4: OPT STEM, High-income						
$Treated_X$	0.0666**	0.0670**	0.0985***	0.1173***	0.1195***	0.1400***
	(0.0283)	(0.0297)	(0.0232)	(0.0207)	(0.0209)	(0.0224)
#High-income	3530	3454	3207	3099	2860	2744

Note: This table shows the policy's heterogeneous effects on the probability of obtaining a STEM degree among students from low and high income households. We compare the policy's effects across family income background by running the same regression separately on the lower-income students' (panels 1 and 3) and higher-income students' sub-sample (panels 2 and 4). All other specifications are the same as in Table D1.