Retaking to Succeed: Causal Evidence on Optional Course Retaking in College

Kelly Chen, Yingying Dong, and Xuan Jiang*

May 31, 2026

Abstract

Many colleges permit course retakes, yet their effects on academic success remain largely unknown. We study a four-year university that canceled a grade-replacement policy in 1995 and reinstated it in 2001. Under this policy, the retake grade replaces the original grade in GPA calculations. The policy substantially increased retaking among students who initially earned D/F grades. Exploiting these policy shifts and the disproportionate retaking response among D/F students relative to C students, we estimate the causal effects of retaking using an instrumental-variables design. Retaking improves same-subject persistence and subsequent same-subject performance, suggesting that students recover academically in the courses and subjects in which they initially struggled. The estimates for university-level retention and four-year graduation are less robust, ranging from positive effects to no detectable effects. The findings suggest that grade replacement can encourage course retaking among struggling students without harming their longer-run academic progress.

JEL Classification: I23, I24, J24, C36

Keywords: Optional course retaking, Same-subject persistence, Subsequent performance, Year-to-year retention, Four-year graduation, Low-achieving students, Instrumental-variables design.

*Chen: Department of Economics, Boise State University, USA KellyChen@boisestate.edu. Dong: Department of Economics, University of California-Irvine, USA yyd@uci.edu. Jiang: Department of Economics, Jinan University, China jiangxuan@jnu.edu.cn. This research received support from a Summer Faculty Research Grant from the College of Business and Economics at Boise State University. All opinions expressed, as well as any errors and omissions, are the sole responsibility of the authors. The authors declare no conflict of interest.

1 Introduction

Poor academic performance is a major driver of dropout, interrupted enrollment, delayed major declaration, and switches into less demanding majors (Stinebrickner and Stinebrickner 2014; Arcidiacono et al. 2012; Arcidiacono 2004; DesJardins et al. 1999). These challenges depress degree completion, lengthen time to graduation, and ultimately hinder labor-market outcomes. Recent national data show a decline in undergraduate degree earners, reversing nearly a decade of growth (National Student Clearinghouse Research Center 2024). Timely graduation also remains elusive for many U.S. students (Denning et al., 2022).

In response, many institutions permit underperforming students to retake courses as a form of remediation, though policies vary widely. A 2015 survey by Advancing Global Higher Education (AGHE) reports that while roughly 80% of colleges impose no limit on the number of repeats, eligibility thresholds differ—about 21% allow repeats for grades of D or below and 60% for C or below—and GPA accounting rules diverge as well, with some campuses replacing the original grade and others averaging initial and retake grades. These design choices—eligibility thresholds, caps, and GPA accounting—change the incentives to retake and may shape who benefits from the opportunity to retake.

Despite widespread adoption, the effectiveness of optional course retaking remains an open question. Conceptually, the effects are ambiguous. On the one hand, a retake can provide a second pass at foundational material, especially valuable for early-career students developing study strategies or confronting gateway coursework; improved grades may bolster confidence and reduce dropout risk. On the other hand, retaking consumes time and credits, can delay progress in other requirements, and may—if viewed as an easy fallback—dampen initial effort. The net effect on persistence and graduation is therefore an empirical matter.

Existing evidence offers mixed guidance, largely from settings with limited student agency. Studies of mandatory remediation and grade-retention policies report

gains in some contexts (e.g., improved persistence and completion; Scott-Clayton et al. 2014; De Paola and Scoppa 2015; Bettinger and Long 2009) but null or negative effects in others, including labor outcomes (Sanabria et al. 2020; Martorell and McFarlin 2011). Even when academic performance improves, dropout risks can rise (Tafreschi and Thiemann 2016). Related K–12 evidence suggests retention is more beneficial for younger students and more harmful for older ones (see discussions in Hill 2014). Taken together, these findings underscore that impacts hinge on institutional design, student characteristics, and autonomy—features that differ materially from the voluntary, course-specific retaking we study.

We establish causal effects of optional course retaking in college, where students choose whether to retake, and incentives are shaped by policy rules. Unlike mandatory remediation or grade retention—which often involve stigma, rigid thresholds, and delays in progression—optional course retaking allows students to choose what and when to retake. This autonomy can soften negative psychological effects and support recovery without derailing momentum. First, student choice can mitigate the loss of motivation associated with involuntary programs. Second, optional retaking can occur earlier—sometimes before outright failure. Third, because retakes target specific courses, students can address precise skill gaps while continuing progress elsewhere. Finally, retaking a single course is less time-intensive than repeating a full year or sequence, potentially making it a more efficient tool for academic recovery. These distinctions imply that the consequences and mechanisms of optional retaking are not readily inferred from studies of mandatory programs and underscore the need for direct causal evidence.

To our knowledge, this is the first study to provide causal evidence on how optional, course-specific retaking affects college achievement at scale. Prior work documents correlations for particular courses or settings (e.g., Biktimirov and Armstrong 2015), but such associations are vulnerable to selection and offer limited insight into population-wide impacts. We estimate causal effects for the broader undergraduate population.

A central challenge in evaluating the effects of retaking is selection: students who elect to retake differ systematically from those who do not. We leverage plausibly exogenous shifts in retake incentives induced by the cancellation (1995) and reinstatement (2001) of a grade-replacement policy at a four-year institution. Under grade replacement, which was in effect through academic year 1994 and again from academic year 2001 onward, the retake grade replaced the original grade in cumulative GPA calculations. During academic years 1995–2000, retaken courses were instead averaged with the original grade. The option to fully replace an initial grade strengthens incentives to retake, especially for students with initial D/F grades because a minimum of C commonly satisfies performance standards. Our instrumental-variables (IV) strategy exploits the disproportionate increase in retaking among students who initially earned D/F (relative to C) when grade replacement was in effect;[1] consistent with this first stage, 92% of retakers initially received a D or F.

We examine the effects of early-college retaking—particularly whether a student retakes by the end of the second year after college enrollment—on later academic outcomes at both the subject and university levels. Our preferred retaking window, Years 1–2 after college enrollment, captures a broad and natural early-college retaking opportunity set: many low grades earned in the first year, especially those earned in the second semester, can only realistically be retaken in the second year. For transparency and more importantly, to assess sensitivity to the length of the retaking observation window, we also report results using a narrower Year 1 window and a broader Years 1–3 window. As discussed in Section 3.1, tracking retaking cumulatively from college enrollment is important for mitigating selection

[1] As detailed in Section 3, our empirical analyses consist of (1) a student-course-level analysis for subject-level outcomes, and (2) a student-level analysis for university retention and graduation outcomes. At the student–course level, the comparison is between students who received an initial D/F in a focal course and those who received an initial C. At the student level, the comparison is between students who received any D/F during the relevant retaking window and those whose lowest grade during that window was C or higher. For brevity, we refer to these comparisons as D/F vs. C students throughout.

concerns, since conditioning only on students enrolled in later years would mechanically drop early leavers who may already have responded to the policy.

Our findings are threefold. First, grade replacement substantially increases retaking among students who initially earn D/F grades relative to students who earn C grades. Second, we find robust evidence that retaking improves both same-subject persistence and subsequent same-subject performance. Third, the evidence for university-level retention and four-year graduation is more suggestive: the estimates range from positive effects to no detectable effects. Importantly, we find no evidence that access to grade replacement makes struggling students worse off in the longer run. This is not trivial, since encouraging retakes could delay other coursework, slow credit accumulation toward new requirements, or lengthen time to degree. Instead, our results suggest that grade replacement encourages course retaking without detectable adverse effects on later retention or on-time graduation. Finally, the mechanism evidence points to course-specific recovery as an important channel: once we account for students' final performance in the potentially retaken course—the retake grade for retakers and the initial-attempt grade for non-retakers—the estimated effects on subsequent outcomes attenuate substantially or disappear.

As sensitivity checks, the Year 1 estimates support improvements in same-subject persistence and performance but show no detectable effects on university-level retention or graduation. The Years 1–3 estimates are broadly similar to the preferred Years 1–2 estimates. Because these windows are cumulative and overlapping, we do not interpret differences across them as clean evidence on the timing of retaking effects.

The remainder of the paper proceeds as follows. Section 2 describes the institutional setting and data. Section 3 outlines the empirical strategy. Section 4 reports the main results. Section 5 presents robustness checks. Section 6 examines mechanisms. Section 7 concludes.

2 Background and Data

2.1 Institutional Background

Boise State University (BSU) is a four-year public university located in the northwestern United States, with an undergraduate population of approximately 22,000. It has the largest undergraduate enrollment in the state of Idaho and offers nearly 80 bachelor's degrees across seven colleges: Arts Sciences, Business Economics, Education, Engineering, Graduate Studies, Health Sciences, and the School of Public Service.

During the study period, BSU required 120 credit hours and a minimum cumulative GPA of 2.0 (equivalent to a C average) for graduation, regardless of major.[2] The same 2.0 threshold was generally used to determine academic probation. Although a $D-$ was sufficient to earn credit for an individual course, many programs imposed higher standards: a $C-$ or even $B-$ was often required for the course to count toward the major. University-wide, unless a department specified otherwise, a $C-$ or better was needed to satisfy prerequisites and advance in course sequences. Consequently, students earning below $C-$ typically had to retake the course to remain on track academically.

BSU's course-retaking policies were relatively permissive during the study period. Students could generally retake any course regardless of the original grade, with few restrictions by course type. Although the university nominally capped total retakes at six and limited attempts per course to two or three, these limits were applied flexibly, and petitions for additional attempts were often approved.[3] Re-

[2]BSU adopted a standard 4.0 GPA scale to convert letter grades to numerical values: A+ = 4, A = 4, A- = 3.7, B+ = 3.3, B = 3, B- = 2.7, C+ = 2.3, C = 2, C- = 1.7, D+ = 1.3, D = 1, D- = 0.7, and F = 0. For simplicity, we refer to plus and minus grades (e.g., $C-$ and $C+$) as straight letter grades (e.g., C) in this text. However, the different variations of straight letter grades are distinguished when included as part of the initial-grade fixed effects in our regression models (see Section 3.2).

[3]As noted in BSU's Faculty Senate Meeting Minutes, December 6, 2001, "appeal(s) by students regarding the [repetition] policy was usually successful."

taking was not cost-free: standard per-credit tuition and fees applied (ranging from $61 to $297 during 1990–2016), unless the student was already paying a flat full-time rate and the added credits kept them within that full-time tuition band—in which case no additional per-credit charges would typically apply. During the observation period, about one in four students at BSU retook at least one course; among retakers, most (68%) retook one or two courses during the entire college time, while the rest retook three or more.

When a student retook a letter-graded course,[4] both the original and the retake appeared on the transcript, with the original marked "repeated." The effect on cumulative GPA depended on timing. Before 1995 and after 2001, BSU used grade replacement: only the most recent grade counted toward GPA. From 1995 to 2001, the university applied grade averaging, combining the original and retake grades. The averaging policy was introduced to raise academic standards but was later rescinded amid fairness concerns, especially because most other Idaho institutions used grade replacement. Neither policy was retroactive; each applied only to retakes occurring after its announcement. Students were informed of these policy changes through the annual University Catalog and subsequent routine advising. The BSU Undergraduate Catalog was typically issued in late spring or early summer, around June or July, before registration for the upcoming academic year. Therefore, there is limited scope for students to have adjusted behavior well in advance of either change.

For Pass/Fail courses, the institutional policy remained entirely unchanged throughout our sample period: a failed course could be retaken and the retake grade could be used to replace the original Fail. Accordingly, the reforms had minimal implications for Pass/Fail courses. We exploit this invariance in a robustness check.

[4] A "letter-graded course" is the one that assigns letter grades (A, B, C, etc.).

2.2 Data

We analyze admission records and transcripts for first-time undergraduate students who enrolled at BSU between academic years 1990 and 2016.[5] For each student, we observe enrollment histories, course grades, graduation date (if any), gender, and transfer status. Following the university's transition to a centralized digital archive in 1998, we also observe course-section-level details, including section attended and instructor, as well as additional demographic information for students observed after 1998, including race, age at college admission, and in-state residency.

We exclude transfer students (80,351 students, or 52% of the initial sample) so that we can consistently define each student's year since college enrollment. During the observation period, BSU offered certificate-only programs for teaching certification in Idaho, as well as continuing-education programs for credentialed teachers; neither required degree completion. We identify and further exclude 1,361 students, less than 1% of the initial sample, who were affiliated exclusively with these programs, as indicated by their enrollment solely in program-specific courses. Together, these exclusions allow us to focus on degree-seeking first-time undergraduates. Finally, we exclude students who first enrolled before 1990 or after 2012 to ensure that each student in the analysis sample can be tracked through the fourth year after college enrollment.

Unless otherwise noted, the focal courses used to define initial grades and retaking are restricted to letter-graded courses. Students who took only Pass/Fail courses during the relevant retaking window are excluded from the corresponding analysis sample, since they have no letter-graded focal course from which to define initial grade or retaking status. However, Pass/Fail courses—and students who took them—are retained when we construct subsequent persistence outcomes, both within the retaken subject and at the university level. To maintain consistency and measurement clarity, we also exclude nontraditional courses, as well as stu-

[5]Unless otherwise stated, all years discussed in this study refer to academic years rather than calendar years.

dents who took only such courses. These include tutorials, labs, study-abroad and satellite-campus courses, concurrent-enrollment courses, and zero-credit courses.

We construct two types of analysis samples: student–course-level samples and student-level samples. The student–course-level samples are used for subject-level outcomes. These samples are restricted to student–course observations in which the student received an initial grade in a letter-graded course during the relevant retaking observation window: Year 1, Years 1–2, or Years 1–3. Overall, our student–course-level data span 3,336 courses in 171 academic subjects. The student-level samples are used for university-level outcomes and are restricted to students who completed at least one letter-graded course during the relevant retaking observation window.

The retaking observation windows are defined cumulatively from college enrollment. At the student–course level, we track whether a student retakes a given course by the end of the first year, the second year, or the third year after college enrollment, and then link this retake status to subsequent subject-level outcomes. At the student level, we track whether a student ever retakes any letter-graded course by the end of the first, second, or third year after college enrollment, and then link this ever-retaking status to subsequent university-level outcomes.

Tracking retaking cumulatively from college enrollment is important for mitigating selection concerns. In particular, we do not condition on students being enrolled in later years, since doing so would mechanically drop early leavers who may already have responded to the policy. For example, the Years 1–2 sample is not restricted to students who return for year 2. Students who leave after year 1 remain in the data whenever their first-year records place them in the relevant sample, and their subsequent retention and graduation outcomes are coded accordingly. Specifically, if a student is not retained into year 2, later retention indicators are set to zero.

Our preferred analysis window is Years 1–2. This window captures the main early-college retaking opportunity while avoiding an overly narrow definition of

retaking. For example, a student who receives a D/F in a second-semester first-year course and retakes that course in the second year is naturally included in the Years 1–2 window but not in the Year 1 window. The Years 1–2 window therefore better reflects the practical opportunity students have to retake early low grades. For transparency, we report the main estimates for all three cumulative windows: Year 1, Years 1–2, and Years 1–3.

The subject-level outcomes comprise two sets of measures. First, same-subject persistence is measured by the number of subsequent same-subject credits completed and an indicator for completing the next credit in the same subject. Second, same-subject performance is measured, conditional on enrolling in a subsequent same-subject course, by the course grade on a 4.0 scale and an indicator for passing that course. All same-subject persistence and performance outcomes exclude the retaken course itself. For students who enroll in multiple subsequent same-subject courses, we select a single "next" course—the first same-subject course taken in the semester immediately after the initial attempt—so that we can include next-course-by-semester-year fixed effects in the performance specifications (Section 3.2). If a student takes more than one same-subject course in that semester, we randomly select one attempt.

The university-level outcomes comprise year-to-year retention indicators and a four-year graduation indicator. These outcomes capture overall persistence and progress at the university. Because the cancellation and reinstatement of grade replacement are six years apart, tracking beyond the fourth year is largely infeasible. First-year retention equals one if a first-year student enrolls at any point during the second year; for example, a student who enrolls in 1991, skips 1992, and returns in 1993 is not retained for the first-year measure.[6] Second-year retention equals one if the student enrolls during the third year. By construction, if first-year retention

[6]The National Center for Education Statistics defines the retention rate as the share of undergraduates enrolled in the previous fall who are again enrolled in the subsequent fall (https://nces.ed.gov/programs/coe/glossary#retention; retrieved on September 25, 2025). We adopt a broader retention measure because BSU allows first-time freshmen to matriculate in the spring as well as the fall.

equals zero, second-year retention is also set to zero. Third-year retention is defined analogously for enrollment during the fourth year. Four-year graduation is an indicator for earning a bachelor's degree within four years of college enrollment.

The relevant university-level outcomes depend on the retaking observation window. For the Year 1 window, we define first-, second-, and third-year retention. For the Years 1–2 window, we define second- and third-year retention. For the Years 1–3 window, we define third-year retention only. Together with the subject-level outcomes, these measures summarize students' subsequent enrollment behavior and academic progress.

Finally, for the mechanism analysis, we construct a course-level "final performance" measure for courses that may be retaken. Final performance is measured either as the numerical grade on the 4.0 scale or as an indicator for earning at least a C. For retakers, final performance is the grade earned on the retake attempt; for non-retakers, it is the grade earned on the initial attempt.

2.3 Descriptive Statistics

Table 1 presents the grade distribution and student characteristics for the entire initial-attempt and retake sample, using the student–course level data from 1991 to 2016.

Among 1,109,598 first-time attempts, 18% ended with a C while 14% with a D/F, indicating that D/F grades were relatively uncommon. By contrast, retaking was overwhelmingly concentrated among D/F earners: 92% of the 48,282 retake attempts followed an initial D (40%) or F (52%). This pattern aligns with institutional incentives—a minimum cumulative GPA of 2.0 for good standing and graduation, and a $C-$ or better to satisfy prerequisites and progress in the course sequences (Section 2.1).

In addition, retakers were more likely to be male and to have weaker prior academic performance, as measured by both most-recent-semester GPA and most-recent-cumulative GPA.

Table 2 (student–course level data) reports descriptive statistics for our main sample that tracks Years 1-2 course-retaking status after college enrollment and links it to subsequent same-subject outcomes. Table 3 (student level data) reports descriptive statistics for the corresponding student level sample, which tracks ever-retaking in Years 1-2 and links it to subsequent university retention and on-time graduation.

Across both samples, D/F students are much more likely than C students to retake courses when grade replacement is in effect. Descriptively, their same-subject and university-level outcomes also improve more during grade-replacement periods. These patterns are correlations rather than causal estimates; we turn to causal identification in the next section.

3 Research Design and Empirical Specifications

3.1 Research Design

We exploit BSU's cancellation (1995) and reinstatement (2001) of grade replacement to obtain plausibly exogenous variation in retaking. Under grade replacement, the retake grade replaces the original in the cumulative GPA, substantially enhancing incentives for lower-performing students to retake a course.

These incentives are strongest for students with initial D/F grades. Although retake entails financial, time, and effort costs, the potential GPA gains can be pivotal for this group: BSU requires a minimum cumulative GPA of 2.0 (C average) for graduation, good standing, and many prerequisites. A successful retake that lifts a D/F to a C or higher can unlock prerequisites, avert probation risk, and restore an on-time degree path. By contrast, the academic payoff to retake is much weaker for students whose initial grade is a C. A further psychological motive is stigma: a D or F on the transcript carries a salient mark, and the opportunity to correct it can itself encourage retaking.

Consistent with this logic, descriptive statistics (Section 2.3) show that D/F stu-

dents retake at much higher rates than C students. Our research design leverages this differential sensitivity together with abrupt policy shifts over time. Specifically, we estimate IV models in which retaking is instrumented with the interaction between an indicator for receiving a D/F (versus C) on the initial attempt and the grade-replacement policy indicator.

Figure 1 illustrates the identification strategy by plotting retake rates over time for students who received their initial grades in the same courses. Retake rates for D/F students fall sharply in 1995 and rise sharply in 2001—coinciding with the cancellation and reinstatement of grade replacement—whereas rates for C students remain low and comparatively flat throughout the period.

Our empirical analysis has three parts. First, we estimate how grade replacement affects early-college retaking. Second, we examine the consequences of early-college retaking for subject-level outcomes—persistence in the same subject and performance in subsequent same-subject courses. Third, we examine the impacts of early-college retaking on university-level outcomes, including year-to-year retention and four-year graduation. In addition, in Section 6, we investigate mechanisms by assessing whether improved final performance in the potentially retaken course helps account for subsequent outcomes. Together, these analyses document the grade replacement policy's effect on retaking, trace its downstream consequences, and provide evidence on the channels through which retaking may matter.

3.2 Main IV Specifications

3.2.1 Same-subject Persistence and Performance

We analyze subject-level outcomes using student–course level data that follow each student–course over time. The unit of observation is a student–course–semester–year. Thus, if a student retakes multiple courses, each attempt enters the sample as a separate observation. Standard errors are clustered at the student level to account for

repeated observations of the same student. The first stage is:

$$Retake_{jct} = \beta\left(GR_t \times DF_{jc,t_{-1}}\right) + X'_{jt}\eta + \Theta_l + \Gamma_{c,t_{-1}} + \Psi_s + \varepsilon_{jct}, \tag{1}$$

where $Retake_{jct} = 1$ if student j retakes course c in semester–year t (after an initial attempt in t_{-1}, a semester-year prior to t), and 0 otherwise; GR_t indicates whether grade replacement is in effect at t; $DF_{jc,t_{-1}} = 1$ if the student's initial grade in c at t_{-1}, was D/F, and 0 if it was C. X_{jt} includes student characteristics (e.g., gender, most-recent-semester GPA and most-recent-cumulative GPA).[7] Θ_l are initial-letter-grade fixed effects, absorbing permanent differences across students who received the same initial letter grade (e.g., systematic gaps between initial C vs. D/F earners). $\Gamma_{c,_{-1}}$ are course-by–semester–year fixed effects, capturing all factors specific to a particular offering of course c in semester-year t in which student j received a D/F grade—including peers, instructor, content, and format—that may affect both the later retake decisions and subsequent same-subject outcomes. Ψ_s are entry-cohort (semester–year of first enrollment at BSU) fixed effects, absorbing cohort-wide differences such as exposure time in college when retakes and outcomes are observed, as well as cohort-specific shocks. The interaction $GR_t \times DF_{jc,t_{-1}}$ is the instrument.

The structural same-subject outcome equation is:

$$SameSubjectOutcome_{jct} = \pi Retake_{jc,t_{-1}} + X'_{jt}\lambda + \Theta_l + \Gamma_{c,t_{-1}} + \Psi_s + \epsilon_{jct}, \tag{2}$$

where $SameSubjectOutcome_{jct}$ is either (i) persistence in the same subject (same-subject credits completed; indicator of completing a same-subject follow-on course) or (ii) performance in the same subject (conditional on enrolling in a follow-on course): 4.0-scale grade and pass indicator. The coefficient π captures the causal effect of retaking, identified by the first-stage instrument above.

In our analysis of same-subject performance (instead of persistence), we ad-

[7] The most-recent-semester and -cumulative GPA enter the model linearly. For first-semester students, they are not available, so we code them as zero and include a dummy to indicate such cases.

ditionally control for next-course-by–semester–year fixed effects, $\Delta_{q,t}$, where q indexes the specific follow-on course taken in semester–year t. This allows us to compare the performance of D/F with C students who attended the same follow-on course within the field.

3.2.2 University-level Retention and Graduation

We analyze year-to-year retention and four-year graduation using student level data. A student may retake multiple courses within a relevant observation window—year 1, years 1–2, or years 1–3 since first enrollment at BSU (Section 2.2). However, because retention and graduation are modeled at the student (not student–course) level, we study the impacts of ever-retaking in a given relevant observation window on subsequent retention and graduation outcomes. For the year 1 group, we study first-, second-, and third-year retention; for the years 1–2 group, we study second- and third-year retention; and for the years 1–3 group, we study third-year retention.

We follow students by entry cohort year (i.e., the academic year of the student's first enrollment at BSU), and define $EverRetake^k_{jy} = 1$ if student j in entry cohort year y has ever retaken any graded course within that group's window k, for $k \in \{1, 2, 3\}$ denoting year 1, years 1–2, or years 1–3, and 0 otherwise; and similarly, $EverDF^k_{jy} = 1$ if the student has earned at least one D or F within that window, and 0 if their minimum grade is at least C.

We estimate the following first-stage equation:

$$EverRetake^k_{jy} = \sigma(GR_y \times EverDF^k_{jy}) + \delta\, EverDF^k_{jy} + Z_j'\kappa + \Lambda_y + \omega^k_{jy}, \qquad (3)$$

where $GR_y = 1$ if the entry cohort y was exposed to grade replacement;[8] Z_j con-

[8]Our main analysis sets $GR_y = 1$ for partially exposed cohorts—e.g., students whose entry cohort year is 1994 might experience grade replacement in their first year and grade averaging in the second; similarly, students whose entry cohort year is 2000 might experience grade averaging in their first year and grade replacement in their second. Alternatively, one may exclude them from estimation. We imple-

tains student characteristics measured at first enrollment at BSU, such as gender, as well as an indicator for students whose degree-seeking status is potentially ambiguous because they took both teacher-certification or teacher continuing-education courses and regular letter-graded courses during the relevant retaking window;[9] Λ_y are entry-cohort-year (academic year of first enrollment) fixed effects.

We define the university retention and graduation outcomes l years after k as $UniversityOutcome_{jy}^{k+l}$ for $l \geq 0$ and $k + l \leq 4$. For example, if $k = 2$, then $k + l = 2, 3, 4$, and $UniversityOutcome_{jy}^{k+l}$ in this case represents the binary indicators for second-year retention, third-year retention, and four-year graduation. For the structural university-outcome equation, we estimate

$$UniversityOutcome_{jy}^{k+l} = \varphi\, EverRetake_{jy}^{k} + \tau\, EverDF_{jy}^{k} + Z_j'\vartheta + \Lambda_y + \varpi_{jy}^{k+l}. \quad (4)$$

3.3 Event Study

One useful feature of our design is that grade replacement was in effect both before the 1995 cancellation and after the 2001 reinstatement, with grade averaging in between. If, absent the policy, the D/F–C outcome gap were drifting approximately linearly over time for reasons unrelated to grade replacement, then comparing the pre-1995 grade-replacement years with the grade-averaging years would tend to be affected by this drift in one direction, while comparing the post-2001 grade-replacement years with the grade-averaging years would tend to be affected in the opposite direction. Pooling the two grade-replacement periods therefore re-

ment both. The results are quantitatively similar, so we report the former in the main text and relegate the latter to Table A1.

[9] We exclude students who took only courses in those programs and hence can be identified as affiliated exclusively with those programs. Additionally, some students took both teacher-certification and teacher continuing-education courses and regular letter-graded courses, in which case we cannot determine whether they were degree-seeking students simultaneously pursuing certification or non-degree-seeking students enrolled in teacher-certification or teacher continuing-education programs who also took regular letter-graded courses. We therefore include an indicator for those ambiguous cases. This affects 7%, 8%, and 9% of Year 1, Years 1-2, and Years 1-2 sample, respectively.

duces the concern that our estimates are driven solely by a simple linear differential trend. This logic is only a useful benchmark and does not replace the parallel-trends assumption that is needed for identification in our setup; we therefore examine the event-study evidence.

3.3.1 Same Subject Persistence and Performance

We assess no anticipatory effects and parallel trends using the event-study versions of the first-stage and reduced-form outcome equations specified in (1) and (2). We replace the binary policy indicator GR_t with a set of event-time indicators. Let $E_t^{(m)} = 1\{t - 2000 = m\}, m = \{-9, \ldots, -1, 0, 1, \ldots, 9\}$, where m indexes years relative to 2000, the last pre-reinstatement year. We omit $m = 0$ as the reference period and bin tails at $|m| > 9$.

The event-study first-stage is:

$$Retake_{jct} = \sum_{m \neq 0} \beta_m \left(E_t^{(m)} \times DF_{jc,t_{-1}}\right) + X'_{jt}\alpha + \Theta_l + \Gamma_{c,t_{-1}} + \Psi_s + \mu_{jct}. \tag{5}$$

The event-study reduced-form same-subject outcome is:

$$SameSubjectOutcome_{jct} = \sum_{m \neq 0} \pi_m \left(E_t^{(m)} \times DF_{jc,t_{-1}}\right) + X'_{jt}\gamma + \Theta_l + \Gamma_{c,t_{-1}} + \Psi_s + \nu_{jct}. \tag{6}$$

Here $DF_{jc,t_{-1}}$ indicates an initial D/F (vs. C) in course c at t_{-1}; X_{jt} is the control vector, and $\Theta_l, \Gamma_{c,t_{-1}}, \Psi_s$ are the fixed effects defined earlier (initial-letter-grade; course-by–semester-year; entry-cohort). The coefficients β_m and π_m trace the dynamic differential in retaking and outcomes for D/F relative to C students around the policy change.

Because grade replacement was in effect before 1995 and after 2001, the pre-1995 coefficients are treated-period coefficients rather than placebo coefficients. Under parallel trends and no anticipation, the non-omitted coefficients during the grade-averaging period should be zero relative to the omitted grade-averaging year, i.e., parallel trends and no anticipation imply $\beta_m = \pi_m = 0$ for $-5 \leq m \leq -1$, and

changes in the D/F–C differential should occur at the 1995 cancellation and 2001 reinstatement rather than before those policy changes.

3.3.2 University Retention and Graduation

For university retention and graduation outcomes, we conduct event studies. The event-study design replaces GR_y in the main specification with event-time indicators relative to the last entry cohort that is not exposed to grade replacement reinstated in 2001. More specifically, for $k = 1$, the last unexposed entry cohort year (or the reference entry cohort year) is 2000; for $k = 2$, it is 1999 (since those who enrolled in 2000 were subject to the grade replacement policy in their second year); and similarly for $k = 3$, it is 1998. Let $E_t^{(m)}$ = 1{y - (2001 - k) = m}, m = {-9, . . . , -1, 0, 1, . . . , 9}, The event-study first stage equation is:

$$EverRetake_{jy}^{k} = \sigma_m(E_y^{(m)} \times EverDF_{jy}^{k}) + \rho\, EverDF_{jy}^{k} + Z_j'\theta + \Lambda_y + \nu_{jy}^{k}, \tag{7}$$

and the event-study reduced-form outcome equation is:

$$UniversityOutcome_{jy}^{k+l} = \sum_{m\neq 0} \varphi_m(E_y^{(m)} \times EverDF_{jy}^{k}) + \mu\, EverDF_{jy}^{k} + Z_j'\xi + \Lambda_y + \upsilon_{jy}^{k+l}. \tag{8}$$

The coefficients σ_m and φ_m trace the dynamic gap in retention for D/F versus C students relative to the last unexposed cohort, which has $m = 0$. Parallel trends and no anticipatory effects imply coefficients for $-6 + k \leq m \leq -1$ are zero. For example, for the Years 1-2 sample, $k = 2$, so the last unexposed cohort is entry cohort year 1999, and zero coefficients for $-4 \leq m \leq -1$, corresponding to entry cohort years 1995-1998 would support parallel trends and no anticipation. The event study for four-year graduation is analogously specified.

4 Main Results

This section presents our main results and event study evidence. We focus on estimates using the Years 1–2 window, which is our preferred specification. As mentioned, this window captures the main early-college retaking margin, including courses first attempted late in the first year and retaken in the second year. In contrast, the Year 1 window captures only retakes completed within the first semester after enrollment at BSU and therefore misses many first-year low grades that can only realistically be retaken in the second year. Estimates using the narrower Year 1 window and the broader Years 1–3 window are reported for transparency; because these windows are cumulative and overlapping, we do not emphasize formal comparisons across them.

4.1 Impacts of Course Retaking on Same-Subject Outcomes

4.1.1 IV Estimates

Table 4 reports the first-stage, reduced-form, and IV estimates for same-subject persistence, measured by (i) an indicator for completing the next credit in the same subject and (ii) the total number of subsequent credits completed in that subject.

The first-stage estimate is strong for our preferred Years 1–2 window. The first-stage F-statistic is 779, and the estimated coefficient on the instrument is 0.087, statistically significant at the 1% level. Thus, grade replacement increases the D/F–C gap in the probability of retaking by 8.7 percentage points.

Panel C of Table 4 reports the IV estimates. D/F students induced to retake within the Years 1–2 window are 13.4 percentage points more likely to complete the next credit in the same subject and complete, on average, 2.4 additional same-subject credits thereafter.

Table 5 reports the corresponding estimates for same-subject performance, measured by (i) an indicator for passing the next course in the same subject and (ii) the

numeric grade in that course on the 4.0 scale. Among students who continue to a subsequent same-subject course, D/F students induced to retake within the Years 1–2 window are 27.2 percentage points more likely to pass the next same-subject course, and their grade in that course increases by 0.49 points on average.

These performance outcomes are observed only for students who take a subsequent course in the same subject, and therefore are conditional on same-subject persistence. Given the IV structure and fixed-effect requirements, formally separating the extensive margin of persistence from the intensive margin of performance is infeasible in this setting. Nevertheless, the likely direction of selection is informative. Because retaking increases the probability that D/F students continue to the next same-subject course, and because the marginal students induced to persist are plausibly weaker than students who would have persisted regardless, the conditional performance estimates may understate the effect of retaking on same-subject performance.

The discussion above focuses on the preferred Years 1–2 window. For transparency, Tables 4 and 5 also report estimates using the narrower Year 1 window and the broader Years 1–3 window. These results similarly indicate that grade replacement increases retaking and that retaking improves same-subject persistence and performance, although the Year 1 estimates are less precise and the Years 1–3 estimates are broadly comparable to those from the Years 1–2 window. Because these windows are cumulative and overlapping, we do not emphasize formal comparisons across them.

4.1.2 Event-Study Evidence

Figure 2 presents event-study plot for the D/F–C gap in retaking during Years 1-2. The event time is academic year, which is in line with the event-study specification in Equation (5). The omitted reference year is academic year 2000, the final grade-averaging year before reinstatement. Figure 3 presents the corresponding reduced-form event-study estimates for same-subject persistence and performance.

To summarize the grade-averaging period more transparently, we supplement the year-by-year event-study coefficients with the average coefficient over the non-omitted grade-averaging years, relative to the omitted grade-averaging year, and report the associated point estimate and p-value in each event-study plot (Callaway and Sant'Anna, 2021).

The pattern in Figure 2 aligns closely with the policy changes: the estimated D/F–C retaking gap falls sharply when grade replacement is canceled in 1995 and rises sharply when grade replacement is reinstated in 2001. The estimates during the non-omitted grade-averaging years are close to zero and statistically insignificant, while the estimates in the grade-replacement years are positive and statistically significant.

The outcome event studies in Figure 3 are noisier than the first-stage event study, and some coefficients during the grade-averaging period between the 1995 cancellation and the 2001 reinstatement are statistically significant. Specifically, significant coefficients appear for academic year 1996 in the number of subsequent same-subject credits completed (Figure 3a) and for academic year 1997 in completing the next credit in the same subject (Figure 3b).

These significant coefficients appear early in the grade-averaging period, which may partly reflect that these years are farthest from the omitted reference year, academic year 2000. The omitted year only normalizes the event-study coefficients and does not affect the main IV estimates reported in the tables.

The statistically significant grade-averaging-period coefficients in Figure 3 are positive, the same direction as the estimated grade-replacement effects. Thus, including these years in the grade-averaging comparison raises the comparison-period benchmark and, if anything, makes the estimated positive effects of grade replacement more conservative. Excluding these years or treating them separately would therefore not weaken the main same-subject findings.

Further, none of these average coefficients is statistically significant. This does not prove parallel trends, especially given the well-known limitations of pre-trend

tests (Roth, 2022), but it suggests that the statistically significant individual coefficients do not reflect a large systematic differential trend over the grade-averaging period.

4.2 Impacts of Ever-retaking on Retention and Graduation

4.2.1 IV Estimates

Table 6 reports the first-stage, reduced-form, and IV estimates of the impacts of ever-retaking within each observation window—Year 1, Years 1–2, and Years 1–3—on year-to-year retention and four-year graduation. These estimates use student-level data, with students tracked by their college-enrollment cohorts.

For our preferred Years 1–2 window, the first-stage F-statistic is 637, and the estimated coefficient on the instrument is 0.139, statistically significant at the 1% level. Thus, grade replacement increases the D/F–C gap in the probability of ever-retaking within the first two years after college enrollment by 13.9 percentage points.

Panel C of Table 6 reports the IV estimates. D/F students induced to ever-retake within the Years 1–2 window are 20.0 percentage points more likely to enroll in year 3 and 16.2 percentage points more likely to enroll in year 4. The estimated effect on four-year graduation is 9.4 percentage points. All three estimates are statistically significant at the 1% level.

For transparency, Table 6 also reports estimates using the narrower Year 1 window and the broader Years 1–3 window. In the Year 1 window, the estimated effects on later retention and four-year graduation are smaller and generally not statistically significant. This pattern is consistent with many first-year D/F students being far from the margin of on-time graduation: a first-year retake may improve same-subject persistence and short-run academic progress without necessarily shifting later retention or graduation. The Years 1–2 window captures an additional second-year retaking margin, when GPA recovery, prerequisite completion, and academic-standing concerns may be more directly linked to continued enrollment and four-year graduation.

Extending the observation window to Years 1–3 yields estimates that are broadly similar to those from the Years 1–2 window. Ever-retaking within the first three years increases third-year retention by 27.6 percentage points and four-year graduation by 14.0 percentage points, both statistically significant at the 1% level. Because the Year 1, Years 1–2, and Years 1–3 windows are cumulative and overlapping, we do not emphasize formal comparisons across them. Instead, we focus on the Years 1–2 window as our preferred early-college retaking window and use the other windows as sensitivity checks.

Appendix Table A1 reports estimates that exclude partially exposed cohorts, i.e., cohorts whose retaking windows include both grade-replacement and grade-averaging years. This restriction is relevant only for the Years 1–2 and Years 1–3 windows, since these windows can span a policy transition. Excluding partially exposed cohorts leads to slightly larger first-stage and reduced-form estimates. The IV estimates for retention and four-year graduation, however, remain broadly comparable to the main estimates in Table 6, which include these cohorts.

4.2.2 Event-Study Evidence

Figure 4 presents the event-study estimates for the D/F–C gap in ever-retaking within the Years 1–2 window. Event time is indexed by students' college-enrollment cohort, following Eq. (7). For this window, the 2000 college-enrollment cohort is partially exposed: students in this cohort have a first-two-year window that spans 2000, when grade averaging was still in place, and 2001, when grade replacement had been reinstated. Therefore, the last clean grade-averaging cohort is the 1999 cohort, which is used as the omitted reference cohort. Similarly, the 1994 cohort is partially exposed because its first-two-year window spans the 1995 cancellation.

Figure 4 shows that the D/F–C ever-retaking gap is close to zero and statistically insignificant for cohorts whose first two years were entirely in the grade-averaging period. By contrast, the gap is positive and statistically significant for cohorts exposed to grade replacement. The post-reinstatement coefficients also increase grad-

ually over time, indicating that the D/F–C gap in ever-retaking became larger in later post-reinstatement cohorts.

Figure 5 presents the corresponding reduced-form event-study estimates for second-year retention, third-year retention, and four-year graduation. These outcome event studies are less clean than the first-stage event study. In particular, some coefficients in the grade-averaging period are positive and statistically significant, especially for the 1996 college-enrollment cohort in the second- and third-year retention outcomes. However, the average coefficients over the clean grade-averaging years are not statistically significant. Similar to before, we interpret this result cautiously: it does not prove parallel trends, but it suggests that the statistically significant individual coefficients do not reflect a large systematic differential trend over the grade-averaging period.

4.3 Post-2008 Changes

The event-study plots for the same-subject outcomes in Figure 3 and the university-level outcomes in Figure 5 show some visible upward movement in the later post-reinstatement years, particularly after 2008. These increases could partly reflect the gradual strengthening of the corresponding first stages for retaking and ever-retaking, shown in Figures 2 and 4. At the same time, they coincide with university-wide initiatives implemented around 2009 that were intended to support struggling students and improve degree completion. These initiatives included: (i) a comprehensive restructuring of the university's lower-level mathematics courses and (ii) the creation of the Office of the Vice Provost for Undergraduate Studies, which led to university-wide standards for first-year writing, student success coaching, and more consistent academic advising.

If these initiatives differentially affected D/F students, they could confound the estimated effects of grade replacement in the later post-reinstatement years. To assess this concern, we re-estimate the IV models for all outcomes after excluding observations from 2008 onward. These estimates are reported in Tables A2–A4.

The results show a clear distinction between same-subject and university-level outcomes. The same-subject effects remain positive and robust when observations from 2008 onward are excluded. By contrast, the positive estimates for university-level retention and four-year graduation are less pronounced in the shorter sample and appear to rely more heavily on the later post-reinstatement years. We therefore interpret the university-level results cautiously, as they may partly reflect the stronger later first stage or complementary institutional changes in later years.

The estimates nevertheless support a cautious but policy-relevant conclusion. We find no evidence that grade replacement worsens longer-run academic progress for struggling students. This is not a mechanical result, because retaking could, in principle, crowd out alternative coursework, slow credit accumulation, or delay graduation. Instead, the evidence suggests that grade replacement increases retaking without detectable adverse effects on subsequent retention or on-time graduation.

5 Robustness and Sensitivity Analyses

Our research design relies on two key identifying assumptions: the exclusion restriction for the IV and parallel trends between the treatment and control groups. The exclusion restriction requires that grade replacement affects outcomes only through its differential effect on retaking among D/F students relative to C students; this rules out concurrent changes that coincide with the cancellation or reinstatement of grade replacement and differentially affect these groups. The parallel trends assumption requires that, absent grade replacement, outcomes for D/F and C students would have evolved similarly.

This section complements the event-study diagnostics with additional robustness checks and sensitivity analyses that probe these identifying assumptions. These include a Pass/Fail placebo analysis, composition and grade-inflation checks, and alternative specifications that use a narrower treatment-group definition and add

additional covariates. For brevity, the corresponding tables are collected in the Appendix.

5.1 Concurrent Policies

During the observation period, BSU expanded support for struggling students through early-intervention resources (e.g., Writing Center in 1994 and Math Learning Center in 2002) and gradually enhanced academic counseling to guide course selection, study skills, and other challenges.

To gauge the influence of contemporaneous changes, we examine outcomes for students in Pass/Fail courses over the same period. These students faced the same academic supports but were largely insulated from GPA mechanics: under both grade-replacement and grade-averaging regimes, an F could be retaken with the retake grade replacing the original Fail, and a Pass does not enter the GPA. Pass/Fail credits still count toward degree progress and prerequisites. Hence, the outcome gap between students receiving F versus P in Pass/Fail courses should be invariant to the policy regime. Systematic shifts would therefore point to other forces (e.g., concurrent supports or unobserved shocks), whereas limited or no change supports the exclusion restriction.

We estimate a reduced-form analogue of Equation (2) on the sample of students who took Pass/Fail courses, replacing $Retake_{jc,t_{-1}}$ with the interaction $(GR_t \times F_{jc,t_{-1}})$, where $F_{jc,t_{-1}}$ indicates that student j received an F in Pass/Fail course c in semester-year t_{-1}.

Table A5 reports the student—course-level reduced-form estimates for the Pass/-Fail placebo analysis. We do not find the positive grade-replacement effects observed for letter-graded courses. The estimates are generally small and statistically insignificant; when statistically significant, they are negative, suggesting that students who initially failed a Pass/Fail course were, if anything, less likely to complete subsequent same-subject courses under grade replacement than under grade averaging. This placebo analysis cannot fully rule out alternative explanations,

especially if students' incentives to take courses Pass/Fail changed across policy regimes or if Pass/Fail courses differ systematically from major-required letter-graded courses. Nevertheless, the pattern provides suggestive evidence that the positive same-subject findings for letter-graded courses are unlikely to be driven solely by contemporaneous policies.

5.2 Initial-Grade Composition and Grade Inflation

Another potential concern is that grade replacement may change the composition of students in the D/F and C groups. For example, under grade replacement, a student who anticipates earning a C may reduce effort, accept a lower initial grade, and plan to retake the course because the retake grade will fully replace the original grade in GPA calculations. Under grade averaging, by contrast, the original grade still affects the final GPA, giving students stronger incentives to maintain effort on the initial attempt. This type of strategic-effort response would imply a policy-aligned shift in the initial-grade distribution: the shares of D/F grades should be higher during grade-replacement periods and lower during the grade-averaging period.

A related concern is secular grade inflation. If grades became systematically higher over time, this could change the composition of D/F and C students or make it easier for D/F students to persist and graduate in later years. To examine both concerns, Figure 6 plots the yearly fractions of C, D, and F grades in letter-graded courses, using the student–course-level data for the Years 1–2 sample.

Figure 6 does not provide clear support for either concern. The D share remains relatively stable over time. The F share increases modestly over the sample period, but this increase does not correspond closely to the grade-replacement periods: it begins during the grade-averaging period and does not show a sharp change at either the 1995 cancellation or the 2001 reinstatement. The C share declines modestly over time. Thus, the figure does not suggest that grade replacement induced a systematic shift from C grades into D/F grades. Nor does it show a systematic decline

in low grades during the post-reinstatement period, which would be expected if the later results were driven by broad grade inflation along the margins most relevant to our design.

We complement this graphical evidence with checks of observed student characteristics. Tables A1 and A2 report means and standard deviations for variables available in the full sample: gender, most-recent-semester GPA, and most-recent-cumulative GPA. For students observed from 1998 onward, Table A6 additionally reports race, age at college enrollment, and in-state residency. We then formally test whether grade replacement is associated with significant changes in these characteristics by regressing each characteristic on the grade-replacement indicator, separately for D/F and C students, using the same fixed effects as in our main specifications, particularly Equations (1) and (3). The results, reported in Table A7, show no statistically significant changes in the observed characteristics of D/F or C students across policy regimes.

Overall, Figure 6 and the composition checks provide little evidence that our estimates are driven by policy-induced changes in the composition of D/F versus C students or by secular grade inflation. Consistent with this conclusion, our main findings are also robust to adding observed student characteristics as controls and to using D students, rather than D/F students, as the treatment group in Table A8.

5.3 Alternative Treatment Group and Additional Covariates

Because event-study evidence cannot prove parallel trends, we further probe the parallel-trends assumption with three complementary specifications that tighten treatment-control comparability: (i) using D students instead of D/F students as the treatment group; (ii) adding controls for White and Hispanic indicators, age at first college enrollment, and in-state residency; and (iii) for the same-subject outcomes only, adding the same controls as in (ii) and replacing course-by-semester-year fixed effects with course-section fixed effects. The last specification narrows comparisons to students in the same course section and therefore holds constant the instructor

and section-specific grading environment.

Table A8 reports the results from these robustness checks, with the baseline Years 1–2 estimates reported in Column (1) for reference. Overall, the results support the main findings from Section 4. Adding additional covariates produces estimates very similar to the baseline estimates across same-subject and university-level outcomes, indicating that the main results are not driven by these observed student characteristics. The same-subject results are also robust to the more demanding course-section fixed-effect specification: all estimates remain positive and statistically significant.

The alternative specification using D students as the treatment group is somewhat more mixed for same-subject persistence but remains broadly supportive. The estimate for subsequent same-subject credits completed remains positive, while the estimate for completing the next same-subject credit is close to zero and statistically insignificant. By contrast, the same-subject performance estimates remain positive and statistically significant. The university-level estimates are also positive and statistically significant under this specification. We interpret these latter results cautiously, however, for two reasons. First, redefining the treatment group from D/F to D changes the complier population. Second, as in the main analysis, the university-level estimates are sensitive to excluding post-2008 observations.[10]

Taken together, Table A8 reinforces the main conclusion that retaking improves same-subject performance and provides additional evidence that the estimates are robust to added covariates and more restrictive course-section comparisons. At the same time, the same-subject persistence estimates are less robust under the D-only treatment definition, and the university-level retention and graduation estimates remain subject to the caution discussed above and in the post-2008 sensitivity analysis.

[10]When post-2008 observations are excluded, the estimate for four-year graduation is 0.148 (std. err. 0.083), while the estimates for second- and third-year retention are both small and statistically insignificant.

6 Potential Mechanism

Section 4 provides robust evidence that course retaking within the first two years after college enrollment improves same-subject persistence and subsequent same-subject performance. The estimates for university-level retention and four-year graduation are non-negative, and we interpret them more cautiously. A natural question is whether these effects operate through course-specific learning or recovery. Intuitively, retaking may deepen mastery of the material and better prepare students for subsequent courses in the same subject. Improved performance may also affect students' confidence or beliefs about their ability in that subject, with potential downstream effects on persistence and graduation. For example, Owen (2010) finds that earning an A in an introductory economics course significantly increases the likelihood of majoring in economics, even after conditioning on the numeric grade received.

Although we cannot directly observe changes in beliefs or confidence, we can examine whether retaking improves final performance in the course in which the student initially struggled, and whether this course-specific improvement helps account for the subsequent positive results documented above.

We first estimate the effect of course retaking on final course performance using the IV specifications in Eqs. (1)–(2), replacing the outcome in Eq. (2) with two measures of final performance: (i) the numeric grade on the 4.0 scale and (ii) an indicator for earning at least a C. For retakers, final performance is measured by the retake grade; for non-retakers, it is measured by the initial grade. We estimate these models using three samples: (1) our primary Years 1–2 sample; (2) a same-instructor sample, in which the initial and retake attempts were taught by the same instructor; and (3) a solo-section sample, limited to courses that offered only one section throughout the sample period. The latter two samples help align grading standards across attempts and reduce concerns that students retake into more lenient sections. The results are reported in Table 7.

Table 7 shows that course retaking consistently and significantly improves final course performance. Numeric grades increase by 1.4–1.7 points on the 4.0 scale, and the probability of earning at least a C increases by 68–85 percentage points. Although grade inflation or grading leniency is a potential concern, the same-instructor and solo-section estimates suggest that the improvement is not driven solely by changes in grading standards across sections or instructors.

Figure 7 presents event-study estimates for final course performance using the Years 1–2 sample. The figure shows that grade replacement is associated with higher final performance for D/F students relative to C students during grade-replacement years.

We next investigate whether course-specific improvements help account for the subsequent positive findings for same-subject and university outcomes. In particular, we re-estimate the main IV models for same-subject and university-level outcomes, augmenting them with each final-performance measure in turn: either (i) the numeric grade on the 4.0 scale or (ii) an indicator for earning at least a C. For the student-level university outcomes, we use the corresponding student-level measures: average final course performance and the share of potentially retaken courses in which the student earned at least a C. The results are reported in Table 8.

Final performance in the potentially retaken course is measured after the retake decision but before the subsequent same-subject and university-level outcomes. We therefore treat it as a mechanism variable rather than a baseline control.[11] Table 8 shows that, once we account for final performance in the potentially retaken course, the estimated effects on subsequent outcomes attenuate substantially. A couple of statistically significant estimates point in the opposite direction. In particular, the estimates for second- and third-year retention are negative when we control for the indicator for earning at least a C. This pattern suggests that course-specific improve-

[11] This exercise is not a formal mediation analysis. Although our specifications include the same fixed effects and covariates as the main models, a causal direct-effect interpretation would require that no residual unobserved factors jointly affect final course performance and later outcomes.

ment is an important proximate channel: students improve their final performance in the course in which they initially struggled, which may help them satisfy prerequisites, meet C-or-better requirements, improve GPA, or avoid academic-standing constraints.

Improved performance in the retaken course may also affect students' confidence or beliefs about their ability in that subject, thereby encouraging continued same-subject persistence and performance. Because we do not directly observe confidence, beliefs, or broader skill acquisition, we interpret these as plausible channels rather than direct evidence. The results do not rule out broader academic spillovers, but they show most directly that retaking improves performance in the specific course retaken and that this course-specific improvement helps account for the downstream effects.

7 Conclusion

We study the effects of optional course retaking on persistence, performance, and graduation using an instrumental-variables design. Our strategy exploits two plausibly exogenous policy shifts—the 1995 cancellation and 2001 reinstatement of BSU's grade-replacement policy—together with the differential sensitivity to the policy among students who initially earned D/F grades relative to their C counterparts. This design identifies the causal effects of retaking for the relevant subgroup of compliers: D/F students whose retaking decisions were affected by the grade-replacement policy. More broadly, the paper illustrates how institutional grading rules can be used to study student decision margins.

Taken together, the evidence shows that grade replacement meaningfully changes student behavior by increasing retaking among students with initial D/F grades. Retaking, in turn, improves same-subject persistence and subsequent same-subject performance. The university-level estimates are less definitive: effects on retention and four-year graduation range from positive effects to no detectable effects and are

sensitive to the inclusion of later post-reinstatement years; they therefore require cautious interpretation. Nevertheless, we find no evidence that access to grade replacement worsens longer-run academic progress. This is an important finding because retaking could, in principle, delay new coursework or slow progress toward degree completion. Our results instead suggest that grade replacement promotes course-specific recovery without detectable harm to later retention or on-time graduation.

The mechanism evidence points most directly to course-specific learning or recovery. Retaking improves final performance in the course in which the student initially struggled, and accounting for this final course performance substantially attenuates the estimated downstream effects. We do not directly test whether retaking also generates broader academic-skill spillovers. The evidence therefore supports a narrower but policy-relevant conclusion: grade replacement encourages struggling students to retake courses, and retaking improves their performance in the retaken courses and in subsequent same-subject courses, without evidence of compromising longer-run retention or graduation outcomes.

Several limitations remain. First, we do not estimate peer or spillover effects. Prior work suggests that large retake cohorts can affect first-time takers in high school settings (e.g., Hill, 2014). In our setting, retakers comprise about 4% of students, so such externalities may be limited, but documenting spillovers remains an important avenue for future research.

Second, external validity warrants caution. Our estimates apply to compliers in this IV framework and to an institutional context with relatively low baseline retention and graduation rates, leaving room for improvement. Whether similar effects would arise at institutions with different student composition, grading norms, retaking policies, or baseline completion rates is an empirical question.

Finally, optional retaking may have longer-run effects beyond the academic outcomes observed here. If improved performance in retaken courses translates into greater skill acquisition, confidence, or persistence in chosen fields, retaking may

affect later labor-market outcomes for initially low-performing students. Quantifying these longer-run effects, and mapping any peer externalities, offers a natural agenda for future research.

References

Advancing Global Higher Education (2015). Course repeat practices. Accessed on June 9, 2023.

Arcidiacono, P. (2004). Ability sorting and the returns to college major. Journal of Econometrics, 121(1-2), 343-375.

Arcidiacono, P., Aucejo, E. M., & Spenner, K. (2012). What happens after enrollment? An analysis of the time path of racial differences in GPA and major choice. IZA Journal of Labor Economics, 1, 1-24.

Bettinger, E. P., & Long, B. T. (2009). Addressing the needs of underprepared students in higher education: Does college remediation work?. Journal of Human Resources, 44(3), 736-771.

Biktimirov, E. N., & Armstrong, M. J. (2015). Is the second time the charm for students repeating introductory finance?. Journal of Financial Education, 32-49.

Callaway, B., Sant'Anna, P. H. C. (2021). Difference-in-Differences with Multiple Time Periods. Journal of Econometrics, 225(2), 200-230.

De Paola, M., & Scoppa, V. (2015). Procrastination, academic success and the effectiveness of a remedial program. Journal of Economic Behavior & Organization, 115, 217-236.

Denning, J. T., Eide, E. R., Mumford, K. J., Patterson, R. W., & Warnick, M. (2022). Why have college completion rates increased?. American Economic Journal: Applied Economics, 14(3), 1-29.

DesJardins, S. L., Ahlburg, D. A., & McCall, B. P. (1999). An event history model of student departure. Economics of Education Review, 18(3), 375-390.

Hill, A. J. (2014). The costs of failure: Negative externalities in high school course repetition. Economics of Education Review, 43, 91-105.

Jewell, R. T., McPherson, M. A., & Tieslau, M. A. (2013). Whose fault is it? Assigning blame for grade inflation in higher education. Applied Economics, 45(9), 1185-1200.

Martorell, P., & McFarlin Jr, I. (2011). Help or hindrance? The effects of college remediation on academic and labor market outcomes. The Review of Economics and Statistics, 93(2), 436-454.

Marx, J., & Meeler, D. (2013). Strike four! Do-over policies institutionalize GPA distortion. Quality Assurance in Education, 21(1), 39-53.

National Student Clearinghouse. (2024). Undergraduate Degree Earners Report. https://nscresearchcenter.org/undergraduate-degree-earners/ Assessed on June 14, 2024.

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

Roth, J. (2022). Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends. American Economic Review: Insights, 4(3), 305-322.

Sanabria, T., Penner, A., & Domina, T. (2020). Failing at remediation? College remedial coursetaking, failure and long-term student outcomes. Research in Higher Education, 61, 459-484.

Scott-Clayton, J., Crosta, P. M., & Belfield, C. R. (2014). Improving the targeting of treatment: Evidence from college remediation. Educational Evaluation and Policy Analysis, 36(3), 371-393.

Stinebrickner, R., & Stinebrickner, T. R. (2014). A major in science? Initial beliefs and final outcomes for college major and dropout. Review of Economic Studies, 81(1), 426-472.

Tafreschi, D., & Thiemann, P. (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.

Table 1: Grade distribution and characteristics, initial-attempt and retake samples

	Initial-attempt	Retake	Mean Diff.
Grade A	0.37 (0.48)	0.00 (0.06)	0.37***
Grade B	0.31 (0.46)	0.01 (0.10)	0.30***
Grade C	0.18 (0.39)	0.07 (0.25)	0.11***
Grade D	0.05 (0.22)	0.40 (0.49)	-0.35***
Grade F	0.09 (0.28)	0.52 (0.50)	-0.43***
Female (0/1)	0.53 (0.50)	0.47 (0.50)	0.06***
Most-recent-semester GPA	2.40 (1.35)	1.91 (1.09)	1.67***
Most-recent-cumulative GPA	2.42 (1.21)	2.19 (0.71)	0.23***
First-semester students (0/1)	0.15 (0.35)	0.00 (0.00)	0.15***
N	1,109,598	48,282	

Note: This table reports sample means, standard deviations in parentheses, and mean differences using student–course-level data. The "initial-attempt" sample includes all first attempts in letter-graded courses completed by first-time undergraduate students between 1991 and 2016. The "retake" sample includes all retake attempts in letter-graded courses during the same period. Letter grades refer to grades earned on the initial attempt. Mean differences compare the initial-attempt and retake samples. *** $p < 0.01$.

Table 2: Sample summary statistics, Years 1-2 sample (student–course-level data)

	C students		D/F students	
Grade Replacement (0/1)	No	Yes	No	Yes
Retake (0/1)	0.003	0.01	0.09	0.20
	(0.05)	(0.10)	(0.29)	(0.40)
Panel A: Student characteristics				
Female (0/1)	0.55	0.50	0.50	0.46
	(0.50)	(0.50)	(0.50)	(0.50)
Most-recent-semester GPA	1.64	1.80	1.13	1.28
	(1.32)	(1.35)	(1.18)	(1.21)
Most-recent-cumulative GPA	1.67	1.86	1.20	1.40
	(1.29)	(1.32)	(1.16)	(1.18)
First-semester students (0/1)	0.32	0.28	0.37	0.30
	(0.47)	(0.45)	(0.48)	(0.46)
Panel B: Same-subject persistence				
# of subsequent credits completed	3.55	3.49	1.08	1.09
	(9.41)	(8.88)	(4.42)	(4.50)
Complete next credit (0/1)	0.35	0.34	0.15	0.15
	(0.48)	(0.47)	(0.36)	(0.35)
N	27,355	81,299	21,922	67,147
Panel C: Same-subject performance *(conditional on persistence)*				
Grade of next course	1.93	1.91	1.43	1.48
	(1.23)	(1.30)	(1.35)	(1.36)
Pass next course (0/1)	0.80	0.78	0.61	0.63
	(0.40)	(0.42)	(0.49)	(0.48)
N	4,634	13,206	1,973	6,543

Notes: This table reports means, with standard deviations in parentheses, using the student–course-level sample from 1991 to 2016. The sample tracks course-retaking status through the end of the second year after first enrollment at BSU and links it to subsequent same-subject outcomes. The performance subsample is restricted to first-time students who enroll in a subsequent letter-graded course in the same subject; therefore, performance measures are conditional on same-subject persistence. "D/F students" refer to students who received a D/F grade and "C students" refer to those who received a C grade in a given letter-graded course. "Grade Replacement (0/1)" indicates whether the grade-replacement policy was in effect during the relevant retaking window; it does not indicate whether the student retook a course.

Table 3: Sample summary statistics, Years 1-2 sample (student-level data)

	C students		D/F students	
Grade Replacement (0/1)	No	Yes	No	Yes
Ever-retake (0/1)	0.004	0.01	0.19	0.35
	(0.061)	(0.10)	(0.39)	(0.48)
Panel A: Student characteristics				
Female (0/1)	0.58	0.57	0.51	0.48
	(0.49)	(0.50)	(0.50)	(0.50)
Panel B: Year-to-year retention				
Second-year retention (0/1)	0.27	0.28	0.27	0.32
	(0.44)	(0.45)	(0.44)	(0.46)
Third-year retention (0/1)	0.22	0.24	0.19	0.24
	(0.41)	(0.42)	(0.39)	(0.43)
Four-year graduation (0/1)	0.13	0.14	0.06	0.08
	(0.34)	(0.35)	(0.24)	(0.28)
N	7,375	30,925	7,506	28,822

Notes: This table reports means, with standard deviations in parentheses, using the student-level sample from 1991 to 2016. The sample tracks ever-retaking status through the end of the second year after first enrollment at BSU and links it to subsequent university-level retention and four-year graduation outcomes. The sample is restricted to first-time students who completed at least one letter-graded course during the observation window. "D/F students" refer to students who received any D/F and "C students" refer to those whose lowest grade was C or higher during the two-year retaking window. "Grade Replacement (0/1)" indicates whether the grade-replacement policy was in effect during the relevant retaking window; it does not indicate whether the student retook a course.

Table 4: Impacts of course retaking on same-subject persistence

	Year 1 (1)	Years 1–2 (2)	Years 1–3 (3)
Panel A: First stage			
Retake (0/1)	0.044***	0.087***	0.114***
	(0.002)	(0.003)	(0.003)
F-statistic	361	779	1086
Panel B: Reduced form			
# of subsequent credits completed	0.156*	0.210***	0.197***
	(0.084)	(0.077)	(0.074)
Complete next credit (0/1)	0.014***	0.012***	0.012***
	(0.005)	(0.004)	(0.004)
Panel C: IV			
# of subsequent credits completed	3.516*	2.416***	1.730***
	(1.911)	(0.886)	(0.650)
Complete next credit (0/1)	0.314***	0.134***	0.107***
	(0.114)	(0.051)	(0.037)
N	141,075	197,723	240,669

Notes: This table reports estimates based on Eqs. (1)–(2), using student–course-level data from 1991 to 2016. The sample is restricted to first-time students who completed a letter-graded course during the relevant retaking window. Subsequent same-subject credit measures exclude the retaken course itself. Columns (1)–(3) track course-retaking status through the end of Year 1, Years 1–2, and Years 1–3 after college enrollment, respectively. All specifications control for gender, linear terms in most-recent-semester GPA and most-recent-cumulative GPA, initial-grade fixed effects, course-by-semester-year fixed effects, and entry-semester-year fixed effects, where entry semester-year refers to the semester-year of first enrollment at BSU. For first-semester students, both GPA measures are set to zero, and an indicator flags these cases. Standard errors, reported in parentheses, are clustered at the student level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Impacts of course retaking on same-Subject performance, conditional on same-subject persistence

	Year 1 (1)	Years 1–2 (2)	Years 1–3 (3)
Panel A: First stage			
Retake (0/1)	0.162***	0.232***	0.247***
	(0.012)	(0.013)	(0.012)
F-statistic	196	338	399
Panel B: Reduced form			
Grade of next course	0.092	0.112**	0.118***
	(0.062)	(0.050)	(0.045)
Pass next course (0/1)	0.049**	0.063***	0.055***
	(0.024)	(0.019)	(0.017)
Panel C: IV			
Grade of next course	0.571	0.485**	0.479***
	(0.380)	(0.214)	(0.180)
Pass next course (0/1)	0.303**	0.272***	0.224***
	(0.147)	(0.082)	(0.069)
N	16,675	26,356	33,758

Notes: This table reports estimates based on Eqs. (1)–(2), using student–course-level data from 1991 to 2016. The sample is restricted to first-time students who completed a letter-graded focal course and subsequently enrolled in a letter-graded course in the same subject; therefore, performance measures are conditional on same-subject persistence. The retaken course itself is excluded from the subsequent same-subject performance measures. Columns (1)–(3) track course-retaking status through the end of Year 1, Years 1–2, and Years 1–3 after college enrollment, respectively. All specifications control for gender, linear terms in most-recent-semester GPA and most-recent-cumulative GPA, initial-grade fixed effects, course-by-semester-year fixed effects, entry-semester-year fixed effects, and next-course-by-semester-year fixed effects, where entry semester-year refers to the semester-year of first enrollment at BSU. For first-semester students, both GPA measures are set to zero, and an indicator flags these cases. Standard errors, reported in parentheses, are clustered at the student level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Impacts of ever-retaking on year-to-year retention and four-year graduation

	Year 1 (1)	Years 1–2 (2)	Years 1–3 (3)
Panel A: First stage			
Ever-retake (0/1)	0.114***	0.139***	0.145***
	(0.005)	(0.005)	(0.006)
F-statistic	621	637	606
Panel B: Reduced form			
First-year retention (0/1)	0.001	–	–
	(0.008)	–	–
Second-year retention (0/1)	−0.006	0.027***	–
	(0.007)	(0.007)	–
Third-year retention (0/1)	−0.002	0.023***	0.040***
	(0.006)	(0.007)	(0.006)
Four-year graduation (0/1)	−0.003	0.013***	0.020***
	(0.005)	(0.005)	(0.005)
Panel C: IV			
First-year retention (0/1)	0.009	–	–
	(0.072)	–	–
Second-year retention (0/1)	−0.052	0.200***	–
	(0.066)	(0.049)	–
Third-year retention (0/1)	−0.017	0.162***	0.276***
	(0.060)	(0.046)	(0.042)
Four-year graduation (0/1)	−0.029	0.094***	0.140***
	(0.043)	(0.034)	(0.032)
N	73,721	74,628	75,056

Notes: This table reports estimates based on Eqs. (3)–(4), using student-level data from 1991 to 2016. The sample is restricted to first-time students who completed at least one letter-graded course during the relevant retaking window. Columns (1)–(3) track ever-retaking status through the end of Year 1, Years 1–2, and Years 1–3 after college enrollment, respectively. All specifications control for gender, an indicator for ambiguous degree-seeking status, and first-enrollment-year fixed effects, where first-enrollment year refers to the academic year of first enrollment at BSU. Robust standard errors are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Impacts of course retaking on final course performance, Years 1-2 sample

	(1) All courses	(2) Same-instructor	(3) Solo-section
Grade (4.0 scale)	1.547***	1.352***	1.674***
	(0.049)	(0.086)	(0.219)
C or above (0/1)	0.752***	0.682***	0.848***
	(0.019)	(0.035)	(0.062)
F-statistic	779	245	34
N	197,723	122,896	8,556

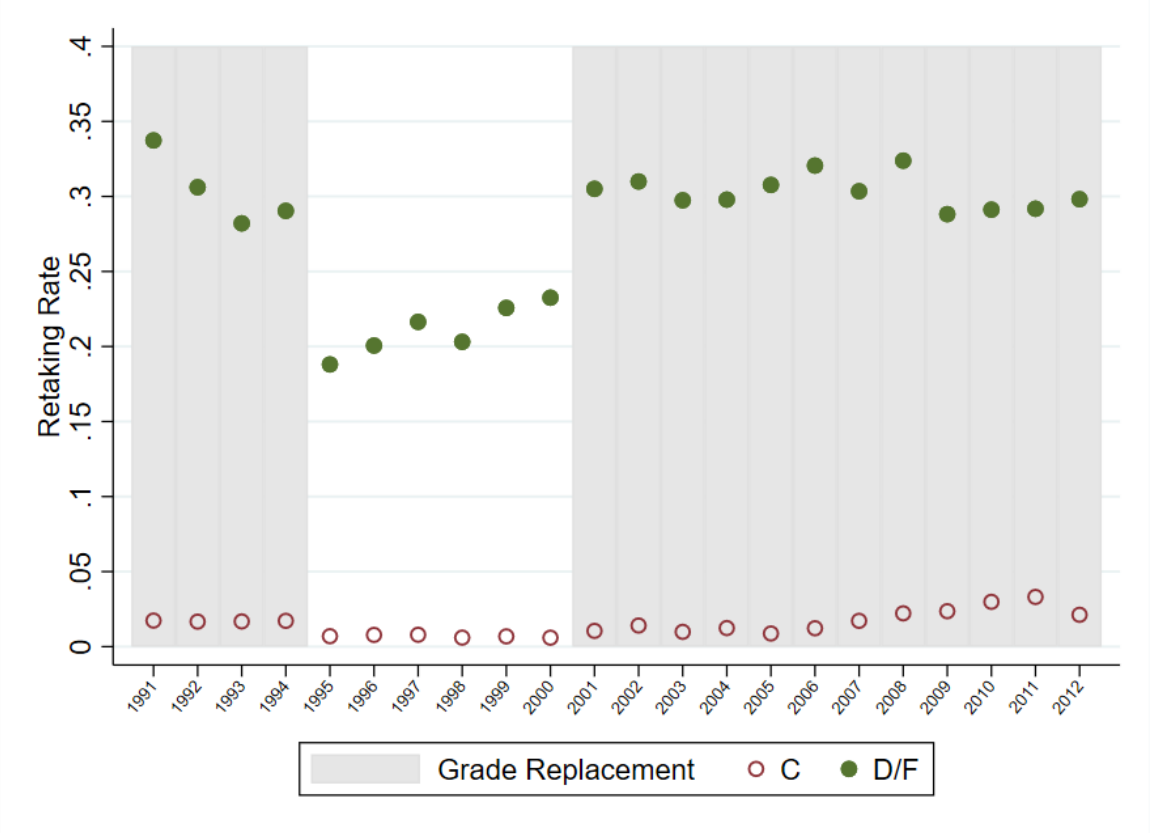
Notes: This table reports IV estimates for the effects of course retaking on final course performance, using student–course-level data from 1991 to 2016 and the specifications in Eqs. (1)–(2). Final course performance is measured either as the numeric grade on the 4.0 scale or as an indicator for earning at least a C. For retakers, final performance is measured using the retake grade; for non-retakers, it is measured using the grade on the initial attempt. Column (1) uses the full Years 1–2 sample, corresponding to Column (2) of Table 4. Column (2) restricts the sample to courses in which the initial and retake attempts were taught by the same instructor. Column (3) restricts the sample to courses that offered only one section throughout the sample period. All specifications control for gender, linear terms in most-recent-semester GPA and most-recent-cumulative GPA, initial-grade fixed effects, course-by-semester-year fixed effects, and entry-semester-year fixed effects, where entry semester-year refers to the semester-year of first enrollment at BSU. For first-semester students, both GPA measures are set to zero, and an indicator flags these cases. Standard errors, reported in parentheses, are clustered at the student level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: Potential mechanism: accounting for final performance in the potentially retaken course, Years 1–2 Sample

	Grade (4.0 scale) (1)	C or above (0/1) (2)
Panel A: Same-subject persistence		
# of subsequent credits completed	1.546	1.457
	(1.747)	(2.293)
Complete next credit (0/1)	−0.025	−0.083
	(0.100)	(0.132)
N	197,723	197,723
Panel B: Same-subject performance *(conditional on persistence)*		
Grade of next course	−0.123	−0.014
	(0.429)	(0.902)
Pass next course (0/1)	0.248	0.502
	(0.188)	(0.372)
N	26,356	26,356
Panel C: Retention and graduation		
Second-year retention (0/1)	0.043	−0.115*
	(0.052)	(0.059)
Third-year retention (0/1)	0.031	−0.094*
	(0.050)	(0.056)
Four-year graduation (0/1)	0.039	0.011
	(0.038)	(0.042)
N	74,628	74,628

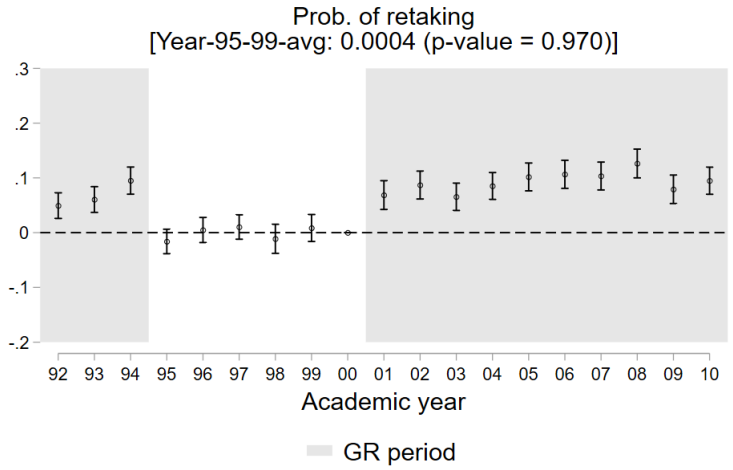
Notes: Panels A–B use student–course-level data from the Years 1–2 sample, 1991–2016. These panels report IV estimates for same-subject outcomes based on Eqs. (1)–(2), additionally controlling for final course performance in the potentially retaken course: either the numeric grade on the 4.0 scale or an indicator for earning at least a C. Standard errors, reported in parentheses, are clustered at the student level. Panel C uses student-level data from the Years 1–2 sample, 1991–2016. This panel reports IV estimates for university-level outcomes based on Eqs. (3)–(4), additionally controlling for the corresponding student-level final-performance measure: either average final course performance on the 4.0 scale or the share of potentially retaken courses in which the student earned at least a C. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 1: Differences in retaking rates between students with D/F vs. C grades over time



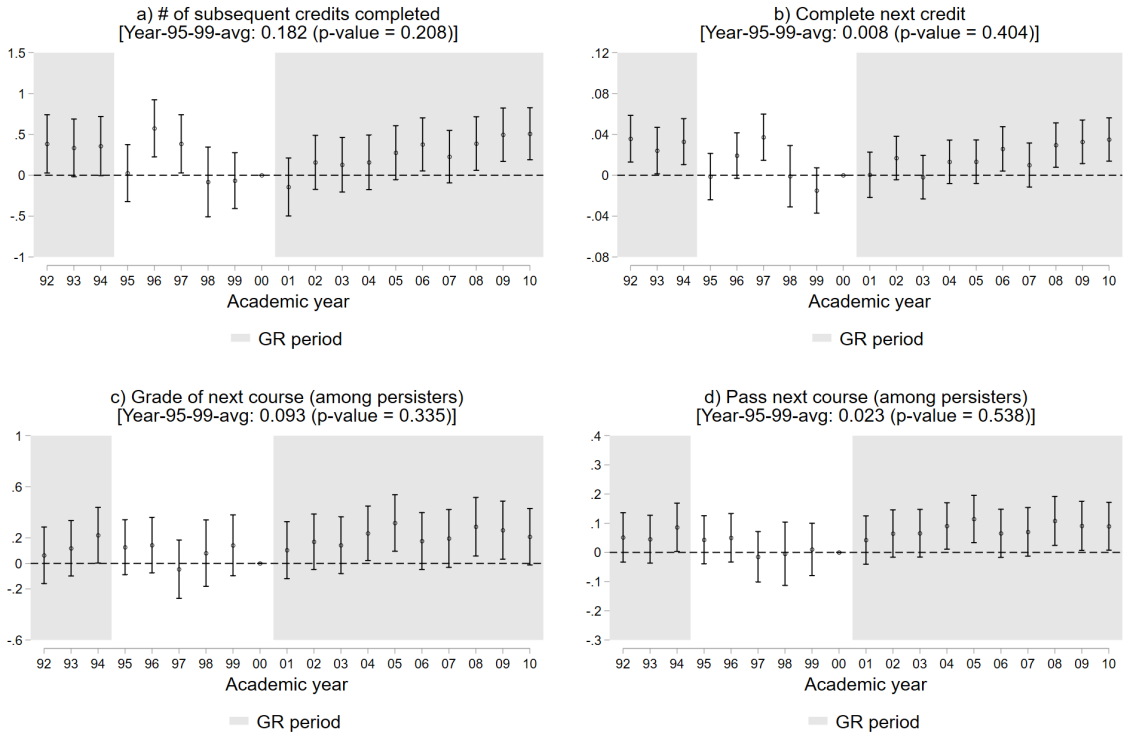
Notes: This figure plots unconditional retaking rates by academic year for first-time undergraduate students in the full student–course-level sample. The series compare students who received an initial D/F in a letter-graded course with students who received an initial C in a letter-graded course.

Figure 2: Event-study estimates of course retaking, Years 1–2 sample: D/F vs. C



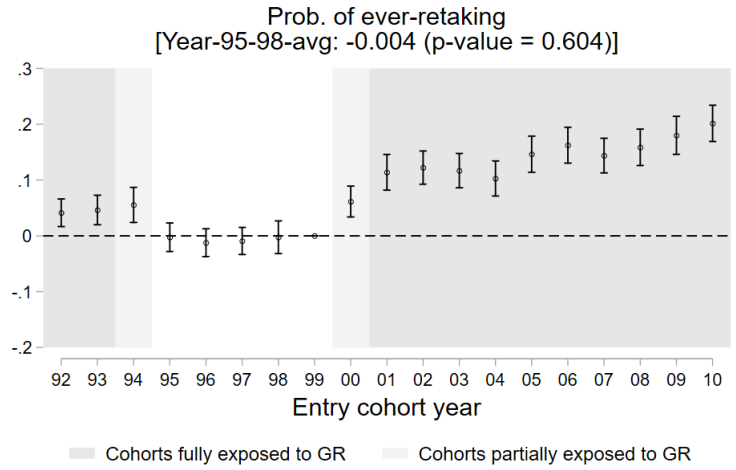
Notes: This figure reports event-study estimates, with 95% confidence intervals, for the D/F–C gap in course-retaking rates using the Years 1–2 student–course-level sample from 1991 to 2016, based on Eq. (5). The sample is restricted to first-time students who received an initial grade in a letter-graded course during the relevant observation window. Year 2000, the last grade-averaging year before the reinstatement of grade replacement in 2001, is the omitted reference year. Standard errors are clustered at the student level. "year-95-99-avg" reports the average of the event-study coefficients for the grade-averaging years 1995–1999.

Figure 3: Event-study estimates of same-subject persistence and performance, Years 1-2 sample: D/F vs. C



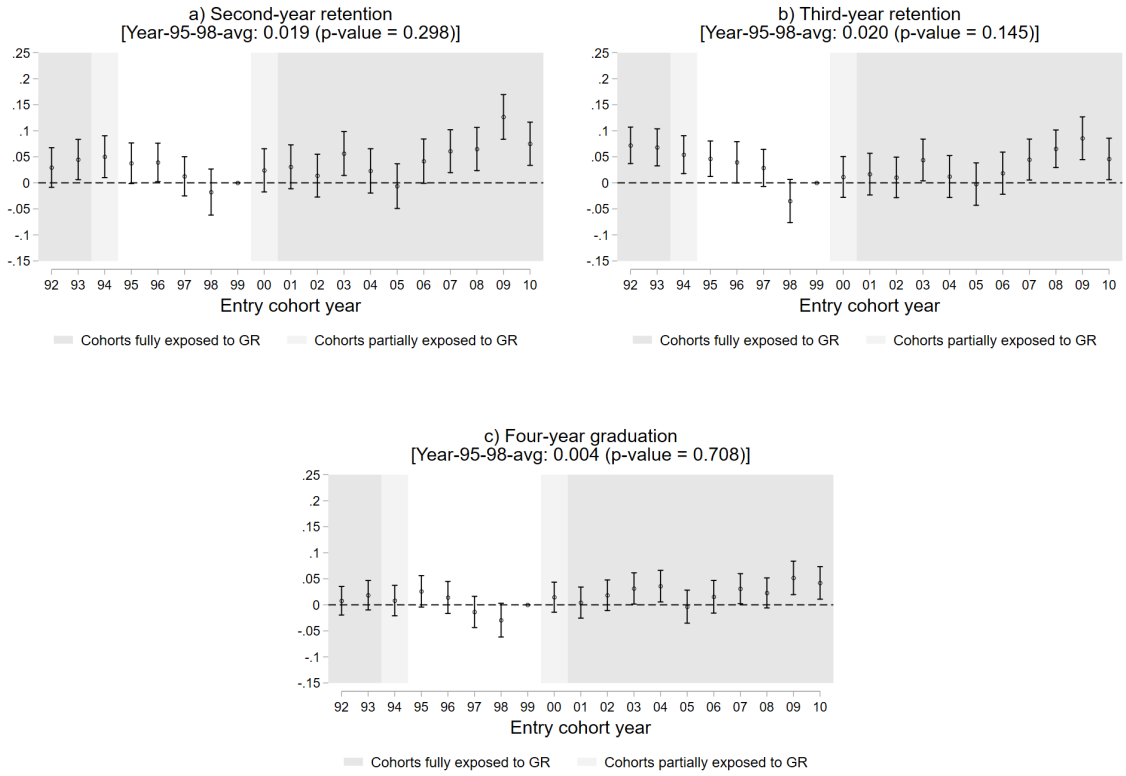
Notes: This figure reports event-study estimates, with 95% confidence intervals, for same-subject persistence and performance using the Years 1–2 student–course-level sample from 1991 to 2016, based on Eq. (6). Panels (a) and (b) use the sample of first-time students who received an initial grade in a letter-graded course during the relevant observation window. Panels (c) and (d) further restrict the sample to students who enrolled in a subsequent letter-graded course in the same subject; therefore, the performance outcomes are conditional on same-subject persistence. Year 2000, the last grade-averaging year before the reinstatement of grade replacement in 2001, is the omitted reference year. Standard errors are clustered at the student level. "year-95-99-avg" reports the average of the event-study coefficients for the grade-averaging years 1995–1999.

Figure 4: Event-study estimate of ever-retaking, Years 1-2 sample: D/F vs. C



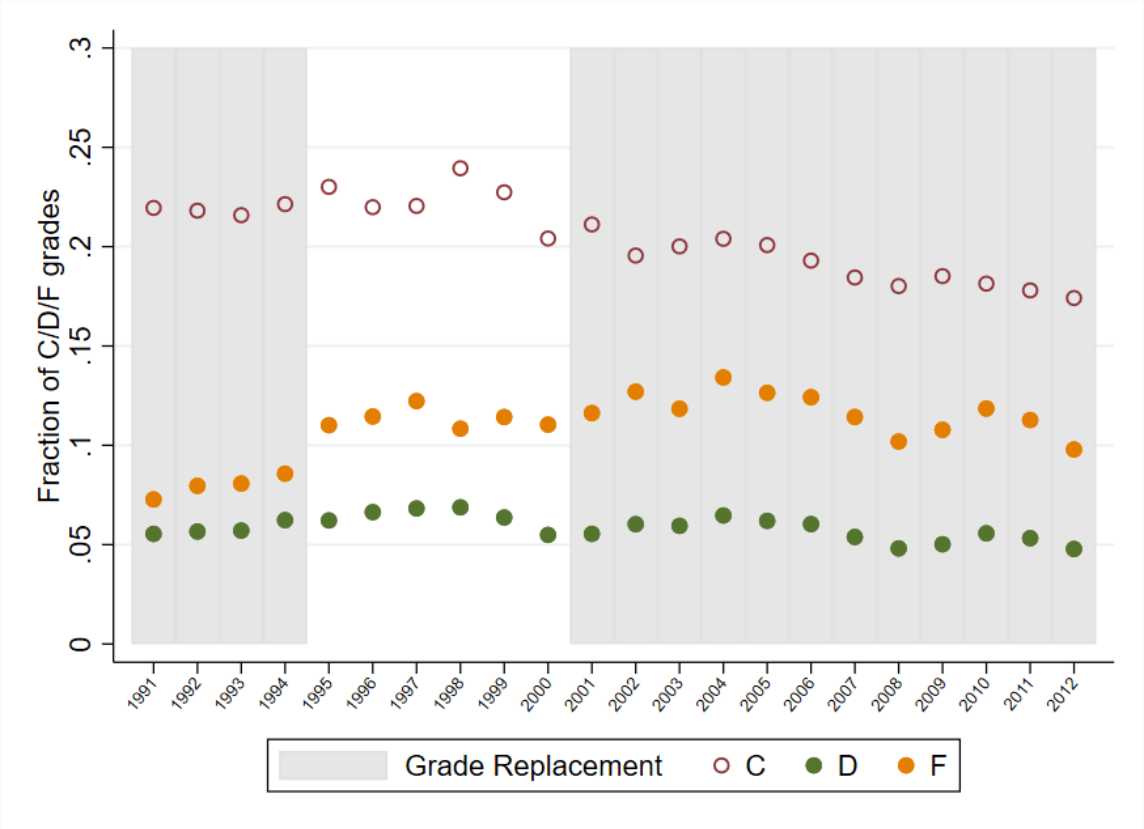
Notes: This figure reports event-study estimates, with 95% confidence intervals, for the D/F–C gap in ever-retaking rates using the Years 1–2 student-level sample from 1991 to 2016, based on Eq. (7). The sample is restricted to first-time students who completed at least one letter-graded course during the relevant observation window. The 1999 college-enrollment cohort is the last fully unexposed cohort before the reinstatement of grade replacement in 2001 and serves as the omitted reference cohort. "year-95-98-avg" reports the average of the event-study coefficients for the clean grade-averaging cohort years 1995–1998.

Figure 5: Event-study estimates of retention and four-rear graduation, Years 1–2 sample: D/F vs. C



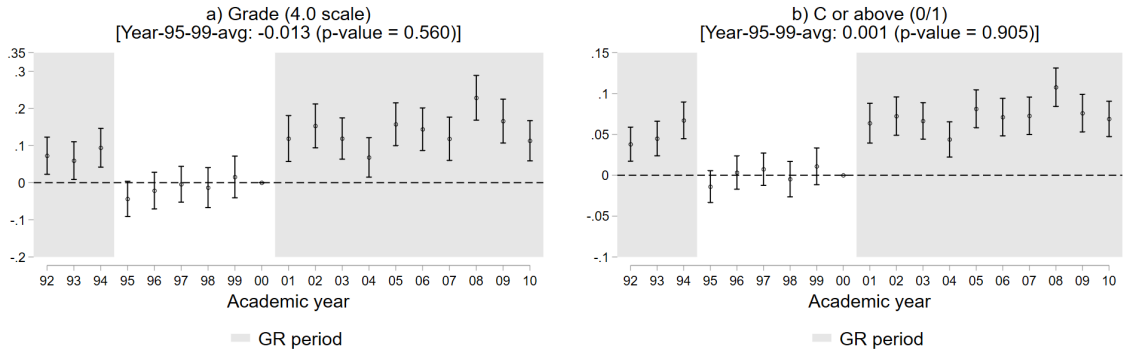
Notes: This figure reports event-study estimates, with 95% confidence intervals, for year-to-year retention and four-year graduation using the Years 1–2 student-level sample from 1991 to 2016, based on Eq. (8). The sample is restricted to first-time students who completed at least one letter-graded course during the relevant observation window. The 1999 college-enrollment cohort is the last fully unexposed cohort before the reinstatement of grade replacement in 2001 and serves as the omitted reference cohort. The marker labeled "year-95-98-avg" reports the average of the event-study coefficients for the clean grade-averaging cohort years 1995–1998.

Figure 6: Fraction of C, D, and F grades over time, Years 1–2 student-–course-level sample



Notes: This figure plots, by academic year, the fractions of initial grades that are C, D, and F among first-time college students in letter-graded courses in the Years 1–2 student—course-level sample. The displayed shares do not sum to one because A and B grades are not shown.

Figure 7: Event-study estimates of final course performance in potentially retaken courses, Years 1-2 sample: D/F vs. C



Notes: This figure reports event-study estimates, with 95% confidence intervals, for final course performance in the potentially retaken course using the Years 1–2 student–course-level sample from 1991 to 2016. Final course performance is measured either as the numeric grade on the 4.0 scale or as an indicator for earning at least a C. For retakers, final performance is measured using the retake grade; for non-retakers, it is measured using the grade on the initial attempt. Year 2000, the last grade-averaging year before the reinstatement of grade replacement in 2001, is the omitted reference year. Standard errors are clustered at the student level. "year-95-99-avg" reports the average of the event-study coefficients for the grade-averaging years 1995–1999.

Appendix

Table A1: Impacts of ever-retaking on year-to-year retention and four-year graduation (excluding partially exposed cohorts)

	Year 1 (1)	Years 1–2 (2)	Years 1–3 (3)
Panel A: First stage			
Ever-retake (0/1)	0.114***	0.164***	0.191***
	(0.005)	(0.005)	(0.006)
F-statistic	621	935	985
Panel B: Reduced form			
First-year retention (0/1)	0.001	–	–
	(0.008)	–	–
Second-year retention (0/1)	−0.006	0.045***	–
	(0.007)	(0.007)	–
Third-year retention (0/1)	−0.002	0.030***	0.053***
	(0.006)	(0.007)	(0.007)
Four-year graduation (0/1)	−0.003	0.018***	0.023***
	(0.005)	(0.006)	(0.006)
Panel C: IV			
First-year retention (0/1)	0.009	–	–
	(0.072)	–	–
Second-year retention (0/1)	−0.052	0.274***	–
	(0.066)	(0.044)	–
Third-year retention (0/1)	−0.017	0.184***	0.277***
	(0.060)	(0.042)	(0.038)
Four-year graduation (0/1)	−0.029	0.107***	0.119***
	(0.043)	(0.034)	(0.033)
N	73,721	68,595	62,244

Notes: This table reports estimates based on Eqs. (3)–(4), using student-level data from 1991 to 2016. The sample is restricted to first-time students who completed at least one letter-graded course during the relevant retaking window. Columns (1)–(3) track ever-retaking status through the end of Year 1, Years 1–2, and Years 1–3 after college enrollment, respectively. Partially exposed cohorts, including cohorts 1994 and 2000 from the Years 1–2 sample, and cohorts 1993, 1994, 1999, and 2000 from the Years 1–3 sample, are excluded. Robust standard errors are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A2: Impacts of course retaking on same-subject persistence (1991-2008)

	Year 1	**Years 1-2**	**Years 1-3**
	(1)	**(2)**	**(3)**
Panel A: First stage			
Retake (0/1)	0.039***	0.076***	0.100***
	(0.002)	(0.003)	(0.003)
F-statistic	250	525	734
Panel B: Reduced form			
# of subsequent credits completed	0.097	0.164**	0.168**
	(0.089)	(0.081)	(0.078)
Complete next credit (0/1)	0.012**	0.010**	0.012***
	(0.005)	(0.005)	(0.004)
Panel C: IV			
# of subsequent credits completed	2.502	2.151**	1.669**
	(2.290)	(1.061)	(0.776)
Complete next credit (0/1)	0.303**	0.130**	0.122***
	(0.136)	(0.061)	(0.043)
N	111,141	151,205	179,198

Notes: This table reports estimates based on Eqs. (1)–(2), using student–course-level data from 1991 to 2008. The sample is restricted to first-time students who completed a letter-graded course during the relevant retaking window. Subsequent same-subject credit measures exclude the retaken course itself. Columns (1)–(3) track course-retaking status through the end of Year 1, Years 1–2, and Years 1–3 after college enrollment, respectively. All specifications control for gender, linear terms in most-recent-semester GPA and most-recent-cumulative GPA, initial-grade fixed effects, course-by-semester-year fixed effects, and entry-semester-year fixed effects, where entry semester-year refers to the semester-year of first enrollment at BSU. For first-semester students, both GPA measures are set to zero, and an indicator flags these cases. Standard errors, reported in parentheses, are clustered at the student level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A3: Impacts of course retaking on same-subject performance among persisters (1991-2008)

	Year 1	Years 1-2	Years 1-3
	(1)	(2)	(3)
Panel A: First stage			
Retake (0/1)	0.140***	0.195***	0.213***
	(0.012)	(0.013)	(0.013)
F-statistic	128	214	263
Panel B: Reduced form			
Grade of next course	0.069	0.088*	0.092*
	(0.065)	(0.052)	(0.047)
Pass next course (0/1)	0.036	0.053***	0.045**
	(0.025)	(0.020)	(0.018)
Panel C: IV			
Grade of next course	0.495	0.452*	0.431**
	(0.461)	(0.266)	(0.220)
Pass next course (0/1)	0.261	0.271***	0.212**
	(0.180)	(0.102)	(0.084)
N	13,019	20,107	25,234

Notes: This table reports estimates based on Eqs. (1)–(2), using student–course-level data from 1991 to 2008. The sample is restricted to first-time students who completed a letter-graded focal course and subsequently enrolled in a letter-graded course in the same subject; therefore, performance measures are conditional on same-subject persistence. The retaken course itself is excluded from the subsequent same-subject performance measures. Columns (1)–(3) track course-retaking status through the end of Year 1, Years 1–2, and Years 1–3 after college enrollment, respectively. All specifications control for gender, linear terms in most-recent-semester GPA and most-recent-cumulative GPA, initial-grade fixed effects, course-by-semester-year fixed effects, entry-semester-year fixed effects, and next-course-by-semester-year fixed effects, where entry semester-year refers to the semester-year of first enrollment at BSU. For first-semester students, both GPA measures are set to zero, and an indicator flags these cases. Standard errors, reported in parentheses, are clustered at the student level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: Impacts of ever-retaking on year-to-year retention and four-year graduation (1991-2008)

	Year 1	**Years 1-2**	**Years 1-3**
	(1)	**(2)**	**(3)**
Panel A: First stage			
Ever-retake (0/1)	0.105***	0.109***	0.117***
	(0.005)	(0.006)	(0.006)
F-statistic	512	364	361
Panel B: Reduced form			
First-year retention (0/1)	–0.004	–	–
	(0.008)	–	–
Second-year retention (0/1)	–0.010	0.002	–
	(0.008)	(0.007)	–
Third-year retention (0/1)	–0.004	0.004	0.020***
	(0.007)	(0.006)	(0.006)
Four-year graduation (0/1)	–0.005	–0.002	0.006
	(0.005)	(0.005)	(0.005)
Panel C: IV			
First-year retention (0/1)	–0.033	–	–
	(0.075)	–	–
Second-year retention (0/1)	–0.092	0.014	–
	(0.078)	(0.066)	–
Third-year retention (0/1)	–0.041	0.040	0.169***
	(0.062)	(0.062)	(0.055)
Four-year graduation (0/1)	–0.043	–0.000	0.051
	(0.045)	(0.005)	(0.041)
N	59,277	60,092	60,497

Notes: This table reports estimates based on Eqs. (3)–(4), using student-level data from 1991 to 2008. The sample is restricted to first-time students who completed at least one letter-graded course during the relevant retaking window. Columns (1)–(3) track ever-retaking status through the end of Year 1, Years 1–2, and Years 1–3 after college enrollment, respectively. All specifications control for gender, an indicator for ambiguous degree-seeking status, and first-enrollment-year fixed effects, where first-enrollment year refers to the academic year of first enrollment at BSU. Robust standard errors are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Reduced-form effects of grade replacement for students enrolled in Pass/-Fail courses, Years 1–2 student–course-level sample

Outcome Variable	**Estimate (SE)**
Retake (0/1)	0.107 (0.080)
Panel A: Same-subject persistence	
Number of subsequent credits completed	−0.403 (0.391)
Complete next credit (0/1)	−0.068 (0.032)**
N	77,396
Panel B: Same-subject performance *(conditional on persistence)*	
Grade of next course	−0.183 (0.398)
Pass next course (0/1)	−0.080 (0.109)
N	6,783

Notes: This table reports reduced-form estimates based on a variant of Eq. (2), replacing $\text{Retake}_{jc,t_{-1}}$ with the interaction $\text{GR}_t \times F_{jc,t_{-1}}$, where $F_{jc,t_{-1}}$ indicates that student j received a Fail in Pass/Fail course c at t_{-1}. The estimates use student–course-level data for first-time students who took Pass/Fail courses during the Years 1–2 window after college enrollment, 1991–2016. Standard errors, reported in parentheses, are clustered at the student level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A6: Characteristics of D/F vs. C students with and without grade replacement, Years 1–2 sample

	D/F students		**C students**	
Grade Replacement (0/1)	**No**	**Yes**	**No**	**Yes**
Fraction of D/F vs. C in the sample	0.51	0.48	0.49	0.52
	(0.50)	(0.50)	(0.50)	(0.50)
White (0/1)	0.81	0.77	0.84	0.79
	(0.39)	(0.42)	(0.37)	(0.41)
Hispanics (0/1)	0.07	0.09	0.05	0.07
	(0.25)	(0.28)	(0.21)	(0.26)
Age at college enrollment	20.99	20.64	24.95	21.82
	(5.76)	(5.13)	(9.98)	(8.47)
In-state residence (0/1)	0.93	0.86	0.92	0.85
	(0.26)	(0.34)	(0.27)	(0.35)

Notes: This table reports means, with standard deviations in parentheses, using the Years 1–2 student-level sample from 1991 to 2016 (N = 74,628). White, Hispanic, age at college enrollment, and in-state residency are observed only for post-1998 students; statistics for these variables are based on the corresponding subsample (N = 47,163). "Grade Replacement (0/1)" indicates whether the grade-replacement policy was in effect during the relevant retaking window; it does not indicate whether the student retook a course.

Table A7: Association of grade replacement with student characteristics, Years 1–2 sample

	D/F students	C students
Panel A: Student-course level data		
Female (0/1)	−0.013	0.001
	(0.020)	(0.079)
Most-recent-semester GPA	−0.035	−0.050
	(0.033)	(0.114)
Most-recent-cumulative GPA	0.013	0.023
	(0.027)	(0.095)
N	89,069	108,654
Panel B: Student level data		
Female (0/1)	0.129	−0.077
	(0.103)	(0.073)
White (0/1)	0.040	−0.034
	(0.085)	(0.059)
Hispanics (0/1)	0.009	0.002
	(0.060)	(0.031)
Age at entry	0.181	−1.047
	(1.056)	(1.191)
In-state status (0/1)	−0.062	−0.001
	(0.067)	(0.047)
N	36,328	38,300

Notes: This table reports the association between grade replacement and student characteristics using the Years 1–2 student–course-level sample and the Years 1–2 student-level sample. Panel A uses student–course-level data. Each characteristic is regressed on an indicator for grade replacement using the fixed effects from Eq. (2): initial-grade, course-by-semester-year, and entry-semester-year fixed effects. Regressions for most-recent-semester GPA and most-recent-cumulative GPA also include an indicator for first-semester students. Standard errors, reported in parentheses, are clustered at the student level. Panel B uses student-level data. Each characteristic is regressed on the grade-replacement indicator using first-enrollment-year fixed effects, as in Eq. (4). White, Hispanic, age at college enrollment, and in-state residency are observed only for post-1998 students; the regressions for these outcomes are based on subsamples of 23,757 D/F students and 23,406 C students, respectively. Robust standard errors are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A8: Alternative treatment group and alternative specifications, Years 1–2 sample

	Main estimates (Full sample)	Alternative treatment group: D students (Full sample)	Additional covariates (Full sample)	Additional covariates & course-section FEs (Post-1998 sample)
	(1)	(2)	(3)	(4)
Panel A: Same-subject persistence				
# of subsequent credits completed	2.416***	1.730*	2.184**	3.350***
	(0.886)	(0.994)	(0.821)	(1.016)
Complete next credit (0/1)	0.134***	-0.003	0.111**	0.124**
	(0.051)	(0.062)	(0.048)	(0.053)
F-statistic	779	439	906	800
N	197,723	138,915	197,723	126,721
Panel B: Same-subject performance				
(conditional on persistence)				
Grade of next course	0.485**	0.601***	0.472**	0.844**
	(0.214)	(0.269)	(0.207)	(0.376)
Pass next course (0/1)	0.272***	0.308**	0.263***	0.310**
	(0.082)	(0.104)	(0.079)	(0.138)
F-statistic	338	172	370	114
N	26,356	21,450	26,356	10,571
Panel C: Retention and graduation				
Second-year retention (0/1)	0.200***	0.230***	0.223***	–
	(0.049)	(0.078)	(0.046)	–
Third-year retention (0/1)	0.162***	0.203***	0.138***	–
	(0.046)	(0.079)	(0.044)	–
Four-year graduation (0/1)	0.094***	0.236***	0.089**	–
	(0.034)	(0.065)	(0.035)	–
F-statistic	637	193	827	–
N	74,628	47,263	74,628	–

Notes: This table reports IV estimates. Panels A and B use the Years 1–2 student–course-level sample from 1991 to 2016; Panel C uses the Years 1–2 student-level sample from 1991 to 2016. Column (1) reproduces the main IV estimates from Tables 4–6. Column (2) defines the treatment group as D students, rather than D/F students, in the student–course-level specifications, and Ever-D students, rather than Ever-D/F students, in the student-level specifications. Column (3) adds controls for White and Hispanic indicators, age at college enrollment, and in-state residency; because these variables are observed only for subsets of students, the IV regressions include a missing-value indicator to indicate this subset. Column (4) retains the Column (3) controls and, for the student–course-level outcomes in Panels A and B, replaces course-by-semester-year fixed effects with course-section fixed effects. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.