Grades, Academic Performance, and Labor Market Outcomes*

Mathías Fondo Rodrigo Nicolau Joaquín Paleo

Job Market Paper

November 1, 2026

Click here for the most recent version

Abstract

Grades open doors to scholarships, graduate programs, and jobs, but distinguishing the effect of grades from the underlying ability they measure is hard. We study this question by implementing a Regression Discontinuity Design around grade-jump cutoffs at Uruguay's main higher education institution, where a deterministic rule converts continuous course scores into coarser transcript grades. Marginally earning a higher first-semester grade raises next-semester GPA by 0.23 standard deviations. Because students observe both the score and the grade, the response reflects the transcript label rather than anything students learn about their own ability. In the labor market, monthly earnings at the first job show no detectable change. In contrast, hourly wages are 17 to 21 percent higher, while weekly hours worked are 2.2 to 2.8 hours lower. The wage premium fades quickly, but the hours reduction persists. These findings seem to be driven by within-industry changes, rather than students sorting across industries with systematically different jobs. A higher grade also raises public-sector entry by 7 percentage points, where hiring rules score GPA explicitly. Finally, among students who start working after the grade is realized, we find evidence that academic performance improves in the semesters after the first job begins, relative to students who marginally earned a lower grade, which is consistent with them working fewer hours.

*Fondo: University of Bologna (mathias.fondo@unibo.it). Nicolau: University of California, Berkeley (rodrigonicolau@berkeley.edu). Paleo: University of California, Davis (jmpaleo@ucdavis.edu).

1 Introduction

Performance measures like course grades and GPA play a central role in students' academic lives and their transition into the labor market. Within higher education, grades can determine access to scholarships and financial aid, eligibility for majors and academic tracks, and admission to graduate programs. When searching for jobs, employers may use grades to screen and rank candidates who compete to fill vacancies.[1] In many settings, grades also enter formal hiring rules, particularly in public-sector employment.

Given that grades are meant to be informative about students' underlying ability, the association between academic performance and later outcomes reflects both what grades measure as well as, potentially, what they do. Isolating this causal component is therefore challenging, but fundamental for the design of grading systems and, more broadly, for understanding how educational credentials translate into relevant economic outcomes.

This paper provides causal evidence on the effects of grades along two dimensions: students' subsequent academic performance, and their transition into the labor market. Our setting is the *Universidad de la República* (UdelaR), the main higher education institution in Uruguay, where courses are usually scored on a continuous 0–100 scale, and a deterministic rule, fixed and common to every course, converts that score into a coarser 0–12 transcript grade. For example, a student who obtains a score of 81 is given a grade of 9, while a student who obtains an 82 is given a 10, even though the two are presumably very similar in terms of their underlying course performance. The student sees both the grade and the score, but only the grade is recorded on the transcript. Our main empirical strategy implements a Regression Discontinuity Design that consists of comparing students on either side of these cutoffs at which the transcript grades "jump" to estimate the causal effect of earning a higher grade.[2] Throughout this paper, we will focus on the effect of marginally earning a higher transcript grade during first-semester courses.

We follow students from enrollment into their first jobs and beyond by linking each student's complete academic record at the School of Economics and Administration of UdelaR to matched employer–employee records covering the universe of formal employment in Uruguay. The academic records contain every course attempt, with both the continuous score and the transcript grade, and run from 2014 through 2023. The employment records report, for every worker, their earnings, hours worked, employing firm, and the

[1] In the United States, roughly four in ten employers report screening applicants by GPA during hiring decisions (NACE, 2025).

[2] An alternative design, which uses the variation coming from grade jumps as exogenous shocks to GPA itself, in the Regression Discontinuity Aggregation fashion of Borusyak et al. (2026), corroborates our findings.

sector and industry of employment at a monthly level, so we are able to observe earnings at their first job, along with the hours the position involves and relevant characteristics of their employer.

We find that marginally earning a higher grade in a first-semester course raises subsequent academic performance. GPA in the following semester increases by 0.56 points, or 0.23 standard deviations, with the effect persisting into cumulative GPA over later semesters and being concentrated at the lower grade cutoffs. This is largely driven by higher-graded students choosing a less risky path going forward: an alternative way of passing a course consists of taking one high-stakes final exam, which will be recorded as a zero on the transcript if failed, while standard coursework has no impact on GPA if the passing minimum is not met.[3] In the first exam period after first-semester grades are released, marginally higher-graded students take 0.37 to 0.40 fewer final exams, roughly 30 percent of the control mean, and are 13 percentage points less likely to have any zero recorded, off a base of 29 percent. These differences disappear at the one- and two-year horizons, so the response is a deferral of risky attempts rather than a permanent change in how students clear their courses and behave in college. Somewhat more than half of the GPA effect operates through this mechanism, with evidence suggesting that the remainder likely reflects better performance on the exams students do take. Importantly, we find no reduction in the number of course credits attempted or obtained in future periods, so this avoidance comes at no cost in terms of degree progression.

On the labor-market side, we find that the first job itself differs. Marginally higher-graded students are paid 17 to 21 percent more per hour and work 2.2 to 2.8 fewer weekly hours, or 12 to 15 percent of the mean, with the probability of working more than twenty hours a week falling by 11 to 13 percentage points. On the other hand, we find no effect on monthly earnings, with the higher wage and the fewer hours offsetting each other. This wage premium seems concentrated at the beginning of the employment spell and fades rapidly, while the reduction in hours worked persists through the first year and a half. On the extensive margin, we find no evidence that a higher grade changes whether or when students start working, with no difference in the months until first employment since enrollment, and no effect on the duration of the first job. The wage and hours differences therefore do not seem to come from a different selection of students into work, but from differences in the first jobs obtained.

In terms of the characteristics of these first jobs, we find that earning a higher grade

[3]Section 2.2 describes the rules in greater detail. Courses are passed through either i) reaching a passing minimum in the standard coursework (typically a midterm and a final), or ii) through a single final exam that decides the course and is taken in one of the exam periods of the student's choosing. The two routes are graded on different scales, with the standard coursework path being more generous.

raises the probability that the first job is in the public sector by 7.0 to 7.5 percentage points, against a base of 14.5 percent. Given our institutional context, this result is not surprising. Hiring in the public sector in Uruguay operates through a formal competitive process, in which the GPA enters explicitly either as an eligibility threshold or as a scored item when evaluating the candidates. From the point of view of the employers, since the official transcript reports grades and GPA, and not the underlying scores, their responses are due to the credential itself. A marginal change in the transcript grade can therefore move a hiring decision, even if it does not reflect differences in students' underlying human capital. This shift toward public-sector employment is persistent. The wage premium itself is observed within industries, so it is not explained by marginal students moving into higher-paying ones. The between-industry component of the effect is null, and the premium survives within the private sector alone. Higher-graded students also sort toward higher-paying firms, and this sorting consolidates over their first year of employment. Retention at the first firm, on the other hand, is unaffected, so a higher grade seems to matter for which match forms in the first place, and not whether the match survives longer.

These results speak to a classic debate in labor economics. Education is associated with higher earnings, and the two canonical explanations differ in why: human capital theories hold that schooling makes workers more productive (Becker, 1964; Mincer, 1974), while signaling theories propose that educational credentials reveal ability that employers cannot observe directly (Spence, 1973), with employers gradually shifting away from the credential as they observe performance on the job (Farber and Gibbons, 1996; Altonji and Pierret, 2001; Lange, 2007). However, most of the evidence related to this debate looks at the effects of years of schooling and degree attainment (Card, 1999; Clark and Martorell, 2014), which can be thought of as measures of education quantity. Some notable exceptions are the literatures on the effects of attending elite institutions (Dale and Krueger, 2002; Hoekstra, 2009; Abdulkadiroğlu et al., 2014; Zimmerman, 2019; Chetty et al., 2026), and of earning honors distinctions (Freier et al., 2015; Feng and Graetz, 2017; Khoo and Ost, 2018). Similarly, grades speak to the quality of the education obtained, and despite being ubiquitous in academic life, having the potential of affecting all students and not just the higher achieving ones, and usually being available for employers to screen on, they have received far less attention than educational attainment itself.

In practice, it is difficult to directly separate what a credential signals from what schooling produces. A common test is an indirect one based on dynamics, since the wage premium of a credential that works as a signal should fade as employers observe output (Farber and Gibbons, 1996; Altonji and Pierret, 2001; Lange, 2007). In Denmark, this premium has been found to fade in the years after graduation (Hansen et al., 2024). We find

the same fading of the wage premium over a much shorter horizon, within months of hiring, even though the hours gap and the public-sector shift persist.

The academic results contribute to the literature on the effects of grades on future academic performance. Existing evidence on what a grade itself does to subsequent outcomes is mixed: a regression-discontinuity study of college letter grades found that the letter grade on a first exam moved performance on the second, while leaving course-taking and major choice unaffected (Main and Ost, 2014), letter grades shift the choice of major (Li and Xia, 2024), and information about relative performance can move achievement in either direction (Azmat and Iriberri, 2010; Azmat et al., 2019). The interpretation in several of these papers is a self-signaling story, in which students see only the coarse grade, so this grade is new information about their own ability, and the response runs through updated beliefs about comparative advantage or expectations. Since students at UdelaR see the underlying score as well as the transcript grade, this explanation seems less likely to be driving our results if students are updating their beliefs using the most informative signal, because the scores are only marginally different.

The quasi-experimental literature on the labor market returns to grades finds overall positive effects: a better letter grade raises monthly earnings six months after graduation at a Singapore university (Tan, 2023), a Danish reform that induced exogenous variation on transcript grades raised earnings at entry, with the gains decaying over time (Hansen et al., 2024), and high-school GPA carries substantial labor-market value, even where the diploma itself signals little (Landaud et al., 2024; Clark and Martorell, 2014). These estimates rest on earnings alone, measured in post-graduation surveys or administrative yearly income data, and the absence of work hours means that monthly earnings cannot be separated into an hourly wage and hours worked. An earnings-based analysis in our setting would conclude that there is no effect, even though the hourly wage and the hours worked are both affected significantly.

A separate literature asks how working while in school might affect academic performance, generally finding that working more hours negatively affects students' academic performance (Stinebrickner and Stinebrickner, 2003), even though credible exogenous variation is rare (Neyt et al., 2019). In work related to our setting, a randomized work-study lottery in Uruguay assigned formal jobs among more than 90,000 student applicants, with participants earning 8 percent more two years later (Le Barbanchon et al., 2023). We reverse the direction of that question, and explore instead how academic performance might affect employment outcomes (both at the extensive and intensive margin) for workers who are still enrolled in college degrees. For a worker who has finished schooling, hours worked are a margin to adjust between work and leisure, while

for an enrolled college student hours worked compete directly with time invested in further human capital accumulation. If students work to cover roughly fixed living costs while studying, then a higher hourly wage maps naturally into fewer hours worked. This is what we observe, with students earning a higher hourly wage and working fewer hours. The fewer hours spent in the labor market are then available for additional study or leisure. We also find evidence suggesting that there are gains in academic performance in the semesters immediately after the first job begins. One implication of these findings is that student employment and academic performance are jointly determined, so that estimates running in a single direction would usually be difficult to interpret.

The remainder of the paper proceeds as follows. Section 2 describes the institutional setting and the data. Section 3 presents the research design and the evidence supporting its validity. Sections 4 and 5 report the academic and labor-market results. Section 6 discusses the potential mechanisms behind these results, and Section 7 concludes.

2 Institutional Background and Data

2.1 Higher education in Uruguay and *UdelaR*

Uruguay is a high-income country of 3.4 million people, with about 40 percent of its population living in Montevideo, its capital city. In terms of regional standards, it is an outlier along several dimensions. Informal employment is rare,[4] with the large majority of workers registered with the Banco de Previsión Social (BPS), the principal social-security administrator, which runs the general scheme covering private-sector and most public-sector employment. Its welfare state is also broad by regional standards, with Uruguay having the largest middle class and the lowest poverty rate in Latin America,[5] and income inequality that is low for the region even though high relative to the OECD.[6]

The education system shares this institutional character. Public education in Uruguay is tuition-free at all levels, and the only requirement for enrolling in the university is the completion of a high school degree.[7] Notably, there is no admissions screening besides

[4]Harmonized estimates for 2023 place Uruguay's informality rate below a fifth of the rates of Mexico and Peru and well below those of Argentina, Brazil, Colombia, and Chile. Among salaried workers, informality had fallen to approximately 7 percent by 2013 (Bergolo et al., 2026).

[5]Over two-thirds of the population is middle class (defined as US$13–70 per day at 2011 PPP), the largest share in the region, and the poverty rate, at 8.8 percent in 2019, is the region's lowest (World Bank, 2022).

[6]Uruguay's Gini index was 40.0 in 2024, below Argentina's 42.4 and Chile's 43.0 and the lowest among the region's high-income economies (World Bank, 2026).

[7]For some degrees, completing a particular track in the final year of high school is also required. For the accounting and administration degrees that constitute the majority of our sample, a high school diploma suffices.

high school completion: no entrance examination, no high-school GPA requirement, and no selection criterion of any kind based on past performance. Within the institutions providing higher education, the Universidad de la República (UdelaR) accommodates the vast majority of students, admitting around 80 percent of all students enrolling in any university degree in the country.[8] Our data come from the School of Economics and Administration (*Facultad de Ciencias Económicas y de Administración*, FCEA), which offers undergraduate and graduate degrees in Accounting, Business Administration, Economics, and Statistics. FCEA is the largest of UdelaR's schools, receiving between 12 and 15 percent of all new entrants to the university in each year from 2014 to 2018, and it had close to 26,000 active undergraduate students in 2018.[9]

2.2 Degree structure, grading, and the information environment

Students interested in earning an undergraduate degree from FCEA enroll directly into a degree program.[10] In order to graduate, students must earn 360 credits through mandatory and elective courses. Typically, courses are worth 10 credits,[11] so roughly 36 approved courses are required for graduation. The suggested path to graduation requires passing 9 courses per year, so nominal time to completion is four years. Appendix D and Tables A9 to A13 reproduce the full curriculum of each degree in the sample, listing the mandatory courses by semester together with their credit values and English translations.

Our analysis focuses on the effect of first-semester grades only for several reasons. First, the choice of courses and grades in later semesters could potentially be affected by past performance, while the first semester is free of this selection by construction. Second, the first semester consists solely of four mandatory introductory courses common to all degrees at FCEA: Introduction to Economics, Accounting Concepts, Administration and Organizations, and Calculus I (see Tables A9 to A13 for details), and these are the only courses first-semester students can take. Newly enrolled students face the same set of

[8] For example, in 2019, UdelaR accounted for 82 percent of students enrolling in degrees related to Economics, Business, or Administration, with the remaining 18 percent spread across several smaller private institutions. This setting is therefore closer to studying the universe of college students in these fields, rather than one institution among many that might be highly selected.

[9] Entrants by school are from the university's basic statistics (Universidad de la República, 2020): in 2018, 3,311 of the 26,973 students entering a UdelaR school for the first time entered FCEA, ahead of Law (2,591), Psychology (2,341), and Medicine (2,079). Active students, defined as those with any course or exam activity in the previous two years plus the year's entrants, are from the planning office's statistical summary for the school (Universidad de la República, 2019). The university-wide count of 137,757 active undergraduates in 2018 comes from the same office's continuous student survey.

[10] This is similar to US systems that, in some cases, require declaring majors at enrollment.

[11] Exceptions to this 10-credit rule exist but are rare. One example is the undergraduate thesis, which is worth 25 credits instead.

courses, with no choice of instructor, section, or difficulty level. Even though students may differ in how many of the four courses they end up taking, they make this choice before receiving any first-semester grade. Finally, because these are the first courses of a student's academic career in university, the grades they produce are the earliest possible performance signal they receive, realized before any degree-specific specialization or accumulated track record.

The grading environment in these courses leaves little room for the kind of instructor discretion that would threaten our research design.[12] First-semester courses are large, with each of the four recording scores for between roughly 2,200 and 5,700 students in each year from 2017 to 2023, and for about 4,000 in the median course-year,[13] so that contact with instructors is limited and a substantial share of the grading is automated by scanner. Most courses are evaluated through two in-person written tests, a midterm and an end-of-course test, each typically worth around half of the total score. The result is a continuous score on a 0–100 scale, which a university-wide predetermined mapping, reproduced in Table 1, converts into the 0–12 grade recorded on the transcript.[14] The cutoffs at which the grade changes are fixed by regulation, identical across courses, instructors, and years, and no curve is ever applied. The GPA shown on the transcript is the simple average of these 0–12 grades.

When results are released through the university's system, students see both their score and the corresponding grade. Since the grade is simply a coarsening of the score, in principle it offers no additional information to the student about their performance or ability beyond what is already revealed by the score. This suggests that two students with marginally different scores but different grades should not update their beliefs about themselves in a substantially different way. The official transcript (*escolaridad*), the document that employers, scholarship committees, and graduate programs might see, reports for each course only the grade, together with its credits, the date of approval, and the number of failed attempts. The transcript's last page displays the student's GPA prominently in a highlighted box, both as a general average over all graded activities, including failed attempts, and as an average over passed courses only. The information environment is therefore asymmetric by design, with students having little reason to revise their beliefs at a cutoff and employers never seeing anything other than the coarse grade.

To pass a course through the coursework route, students need to reach a minimum

[12] Section 3 presents formal evidence against score manipulation.

[13] Counts are distinct students with a recorded score in the course's first-semester roster, with Calculus I counted across its two course codes.

[14] UdelaR replaced the 0–12 scale with an even coarser conceptual scale in January 2025, after our sample period ends, so every grade in this paper is on the 0–12 scale.

of 50 points in the continuous score. Students who do not reach this minimum but still wish to pass the course have two options. First, they can enroll in the course again the following year and attempt the coursework route once more. Second, they can take the course's final exam in one of the dedicated exam periods that follow each semester, with the performance on this high-stakes exam alone deciding whether the course is passed or failed. These two routes are not treated symmetrically on the transcript. Falling short of the 50-point minimum through the standard coursework leaves no grade on the record, whereas a failed final exam is recorded as a zero and counts towards the GPA. Additionally, final exams are graded on a harsher scale than the coursework route, with an exam passed at 50 percent receiving only a grade of 3, while the same 50 points earned through the in-course tests receive a 6. These high-stakes final exams are therefore important activities through which a student's GPA can worsen, which is an important feature of the grading rules that matters for the interpretation of the academic results we present in Section 6.1.

2.3 Labor-market institutions and the role of GPA in hiring

Formal employment and working while enrolled in college. The matched employer–employee records of the BPS cover the universe of formal employment. Because informality among salaried workers is only around 7 percent (Bergolo et al., 2026), and we are additionally conditioning on individuals enrolled in college, the first formal job we observe in the BPS records is essentially their first job. In this respect our setting differs from most of Latin America, where administrative employment records might miss a substantial informal margin. For the same reason, null effects on the timing of entry are unlikely to reflect substitution between formal and informal work, so we can interpret them as nulls on the extensive margin.

Working while enrolled in college is very common in this setting. Figure 5 plots the distribution of months passed between enrollment and the start date of the first job, with the median entrant starting within two years of enrolling. For most students, their first job begins well before their expected graduation, a fact that shapes both the choice of the first job as our labor-market margin and the importance of being able to observe hours worked as an additional outcome, since for an enrolled student hours worked compete directly with hours of study.

GPA and hiring decisions. Hiring in the public sector in Uruguay operates through formal, merit-based competitive processes, administered by the national civil-service office

(ONSC) through a centralized portal (*Uruguay Concursa*). For positions aimed at college students and graduates, academic records enter this process explicitly. The GPA can appear either as an eligibility requirement (a minimum GPA to be considered for the position) or as a specific item in the formula used to score and rank candidates. Under such a rule, a marginally higher transcript grade can move a hiring decision directly, holding human capital fixed. This institutional feature is connected to the public-sector results of Section 5. Private-sector hiring, on the other hand, is less transparent on how it uses candidates' information during the process, and we are more limited in what we can document. Because transcripts from UdelaR are easy to request and verify electronically from the university, employers can ask applicants for them when screening, although nothing requires them to do so.

2.4 Data sources

We use administrative data linked at the individual level that allow us to follow students during their academic career and into the labor market.

Academic records (2014–2023). The registrar data track the progress and performance of each student within the FCEA. This includes information on the time and degree of enrollment, along with every course attempted since, with both the precise 0–100 score and the 0–12 transcript grade. This allows us to build a panel that tracks, for each student, their outcomes in each semester since enrollment: GPA for each semester, and cumulative GPA until each semester, number of course credits attempted, passed, and failed, along with which field the courses taken belong to. Additionally, this provides information on students' basic demographic characteristics at the time of enrollment.

Student surveys (2018–2023). Enrolled students must complete a survey that collects information on their individual characteristics each year. Failing to complete this survey results in students not being able to register for classes on the following semester, so compliance is effectively universal among continuing students. This survey collects a rich set of demographic variables, along with information relating to prior education (including the type of high school they attended before enrolling in college), family background (including parental education and occupation), scholarships received to attend college, and whether they participate in the labor market. We use this set of variables to characterize our population for external-validity discussions, and importantly, to conduct balance tests in support of our research design.

Initial assessments and personality inventory (2018–2023). Before newly enrolled students begin their first semester, they take diagnostic assessments in Mathematics and Reading, and complete a Spanish-language Big Five personality questionnaire (BFI-44). Because these are measured before any university grade is realized, they are pre-determined with respect to our treatment, and we use them as an additional set of covariates for validity tests, since this allows us to check that cognitive skills and non-cognitive personality traits are also balanced at the grade cutoffs (Tables 4 and 5). Unfortunately, these are only available for the later cohorts in our sample, and so are used mainly to perform balance tests of the RD design.

Social security records (2010–2023). We obtain matched employer–employee records from BPS, covering the universe of formal employment. These include, for every worker and month, employment status, earnings, hours worked, sector (public or private) and industry of employment, and an identifier for the employing firm. This allows us to construct firm-level average wages as a proxy for firm pay policy, which underlies the firm-sorting results. Two industry classifications can be assigned to each job. Every record has an economic-activity code following ISIC. Additionally, the records carry the firm's bargaining group under the *Consejos de Salarios*, Uruguay's industry-level collective bargaining system. Appendix C documents details on the differences between classifications, and shows every industry-based result under both alternatives.

2.5 Sample construction and descriptive statistics

Sample restrictions. We restrict the sample to students who enrolled in FCEA at ages 17–19. This allows us to focus on the population for which our research question is most relevant: students enrolling shortly after finishing high school, for whom this is their first experience in college, and for whom their academic achievements are possibly their only signals of productivity going into the labor market. This excludes returning adult students, whose grades would be neither their first performance signal, nor a plausible input into a first job that likely started before enrollment. Table 2 shows the contrast between the two groups. The full enrollee population averages 23.5 years of age at enrollment and 4.5 years elapsed since high school, with 46 percent working during their first enrollment year, while the restricted sample averages 18.1 years of age and 1.1 years since high school, with 11 percent working at entry.

Definition of the first job. We define a student's first job by looking at the first month with positive earnings in the 2010–2023 BPS panel, without imposing a minimum spell length. Because the labor panel begins four years before the earliest cohort enrolls, we are able to observe potential pre-enrollment work histories, and can distinguish genuine new labor-market entrants from students who begin their degree already having some work experience. When estimating the effects on first-job outcomes, our sample consists of students whose first recorded job began after enrollment. Students with pre-enrollment employment are excluded by this restriction, and do not enter the sample if they switch jobs to a different firm. When a worker holds jobs at multiple firms in the same month, the firm paying the highest earnings is taken as the reference employer, with earnings and hours added across all jobs.[15]

Descriptive statistics. Table 2 reports descriptive statistics for the full enrollee population along with the age-restricted sample. In terms of students' characteristics for the restricted sample, 53 percent are women, 83 percent identify as white, 74 percent are first-generation university students, 51 percent live in Montevideo, and 30 percent attended a public institution during their last high school year. As for the degrees that they are enrolled in, 64 percent are pursuing the accountancy degree, with the remainder spread across economics (12 percent), administration (8 percent), and the administration-technician track (9 percent). Given that a high school degree is the only requirement for enrolling at UdelaR (Section 2.1), it is not surprising that the share of first-generation students in the sample is high.

At the first job, the median entrant earns a real monthly income of around 13,700 Uruguayan pesos, roughly 330 US dollars,[16] works roughly 19 hours per week at the beginning of their first job, and starts that job a median of 23 months after enrollment. The public sector represents 14.5 percent of these jobs.

2.6 Labor-market outcomes and measurement conventions

We focus on the first formal job for two reasons. First, grades should matter most at the school-to-work transition, when employers observe few alternative signals of productivity. Second, working while enrolled is very common in our setting (Section 2.3), so for

[15] Robustness checks including these observations or changing how the handling of multi-employment show qualitatively similar results.

[16] Monetary figures throughout the paper are real Uruguayan pesos in the units defined in Section 2.6. The dollar figure uses the average interbank exchange rate of October 2022, the base month of the deflator, of 41.1 pesos per dollar (Instituto Nacional de Estadística, 2023).

most students the first job begins during their studies rather than after graduation.

Outcomes are organized in three blocks relating to the extensive margin, the intensive margin, and the characteristics of the match, each estimated at the first job, and in months after the start date of the first job.

(i) *Extensive margin*: months from enrollment to first job, length of the first employment spell, non-employment indicators at 6 and 12 months, and the month-by-month employment profile over 36 months, which are the outcomes on which any search, delay, or selection response would appear.

(ii) *Intensive margin*: real monthly earnings, weekly hours worked, and real hourly wages.

(iii) *Match characteristics*: indicator for the hiring firm being in the public sector, industry of the firm, average wage paid by the employing firm, and retention at the first firm.

Beyond the fixed horizons, Section 5.4 also traces the wage, hours, and earnings responses month by month over the two years following the start of the first job.

Measurement conventions. Hours worked are recorded in the social security data through the monthly payroll forms that firms are required to file, with employers reporting each worker's weekly hours worked and the days worked within the month. Firms found misreporting are subject to penalties, and validation exercises on these records suggest that the quality of the reported hours is high (Bergolo et al., 2026). Earnings are real monthly Uruguayan pesos, deflated by the Uruguayan National Institute of Statistics (INE) consumer price index with base on October 2022.

3 Empirical Strategy

3.1 Course-level grade effects: pooled regression discontinuity

To identify the effect of marginally receiving a higher grade in a first-semester course, we construct a student-by-course dataset, assign each 0–100 score (s_{ict}) to its closest grade-jump cutoff (cutoff_{ict}), and recenter the running variable at that cutoff. Let r_{ict} denote the running variable: $r_{ict} = s_{ict} - \text{cutoff}_{ict}$.

As our main specification, we estimate the following equation:

$$Y_{ict} = \alpha + \beta z_{ict} + \gamma_1 r_{ict} + \gamma_2 r_{ict}^{+} + X_{ict}'\theta + \varepsilon_{ict}, \tag{1}$$

where $z_{ict} = \mathbf{1}(r_{ict} \geq 0)$ and $r^{+}_{ict} = z_{ict} \cdot r_{ict}$. The design pools observations over the six grade-jump cutoffs, with X_{ict} including cutoff-level fixed effects,[17] so that β is identified from within-cutoff variation. Specifications adding individual characteristics as controls are shown alongside the main results, but they don't affect any of the conclusions. Under the assumptions discussed in Section 3.2, the coefficient β identifies the effect of receiving a higher transcript grade for students at the margin, holding their underlying score fixed. It is a course-level estimand, namely the effect of one grade step in one first-semester course.[18]

Estimation follows the modern RD toolkit: the local-linear specification was chosen following Pei et al. (2022), triangular kernel, and MSE-optimal bandwidths selected following Calonico et al. (2020), with standard errors clustered at the individual level.[19] Depending on the outcome of interest, optimal bandwidths are typically between 0.8 and 2.1 score points. Alternative specifications with fixed 1, 1.5, and 2 point bandwidths don't affect the main results. Throughout the paper we report both the conventional local-linear estimate and the robust bias-corrected confidence interval of Calonico et al. (2020).

3.2 Validity of the course-level design

A causal interpretation of β requires that nothing other than the transcript grade changes discontinuously at the cutoffs, and we test this assumption by looking for manipulation of the running variable and for discontinuities in predetermined characteristics and in placebo outcomes, including selection into the labor-market sample.

Manipulation of the running variable. An important threat to any grade-cutoff design is score manipulation that would lead to students selecting into the higher grades for reasons unobserved to us and also related to the outcomes of interest.

The institutional argument against this comes first. A large share of grading in first-semester courses is automated by scanner and inaccessible to discretion. The cutoffs apply to the total score, which aggregates components that are graded separately, typically a midterm and a final (Section 2.2). To place the total just above a threshold, a grader marking any single one of these components would have to know the student's scores on the others and anticipate the aggregation. Importantly, a generalized behavior that

[17] These are indicators that identify which is the closest cutoff to that observation's score, with the lowest cutoff omitted.

[18] Appendix A develops a GPA-level estimand using an aggregate design motivated by Borusyak et al. (2026).

[19] Since the dataset is at the student-by-course level, there is a chance that an individual enters the RDD multiple times, and therefore clustering at the individual level is necessary in this setup.

simply bumps all students' grades indiscriminately would not threaten our identification strategy; what would be problematic is that instructors target specific students for some unobservable reason that relates to our outcomes of interest. However, in this setting first-semester courses are mass courses with thousands of students and essentially no individual instructor contact, which leaves little room for relationship-based favoritism.

Figure 1 shows the distribution of 0–100 scores of all first-semester courses. The score density exhibits clear excess mass at 50, which is the minimum passing threshold. Since we are interested in estimating the effect of earning a higher grade, and not in passing a course, we do not use the observations that are near this passing threshold in our analysis. In any case, standard RD identification fails at this threshold, so it could not be used for estimation. The bunching at 50 seems reassuring about the remaining cutoffs, since manipulation by graders, where it exists, is visible in the density and, as we show below, detectable by a formal manipulation test. On the other hand, we don't observe a similar behavior at the grade-jump cutoffs that we do use for estimation.

The local polynomial density test of Cattaneo et al. (2020a) does not reject continuity at the pooled cutoff (robust $p = 0.371$). However, the running variable is discrete and heaped, with mass points at integers, which are the most common scores given on a 0–100 scale. Half of student×course observations carry fractional scores from partial-credit and scanner-penalty conventions, with 14 percent ending in half points and 36 percent in finer fractions. The density test assumes a continuously distributed running variable, and our scores, heaped at mass points of unequal size, do not meet that assumption.

We complement the local polynomial density test with the rounding-aware manipulation test of Andarge et al. (2025), which is built for running variables that mix continuously generated values with deterministic rounding conventions. In our setting, this means interpreting the integer values of the running variable as "rounded" values of an underlying continuous score. The idea is that we infer what "normal" rounding looks like from integers that do not change the grade, and then compare how often scores land in a roundable interval around each grade-jump cutoff with how often they would under that normal rounding. Excess rounding at the cutoffs that assign higher grades would be suggestive of manipulation beyond normal rounding. At the 50-point passing threshold, where discretionary passing is visible to the eye in the density, the test detects excess rounding of $+0.93$ percentage points ($t \approx 9.5$), so that the method finds manipulation at this point. At the six grade-jump cutoffs the design actually uses, estimated excess rounding ranges from -0.13 to $+0.06$ percentage points and is never statistically distinguishable from zero. Appendix B details the method, the implementation choices, and the full set of estimates. Together with the covariate and placebo smoothness documented next, we

disregard manipulation as a threat to our research design.

Balance and placebo outcomes. Table 4 examines whether a set of predetermined characteristics are smooth through the cutoffs: we focus on indicator variables for female, first-generation college students, white, and finished high school at a public institution, along with age at enrollment. Additionally, we test for discontinuities in pre-enrollment Mathematics and Reading assessment scores for the post-2018 diagnostic subsample. Table 5 extends the test to Big Five personality traits measured at enrollment as well. Overall, we find balance on both sides of the grade-jump cutoffs across all of these variables; Figures 2 and 3 display the corresponding RD plots.

Table 6 reports RD results on a set of placebo outcomes of interest. The first column tests for discontinuities in a variable that captures the same-semester GPA in the students' other first-semester courses (i.e. excluding from the computation of the GPA the performance on the course for which the student is entering the RDD, which would mechanically show a positive effect by being to the right of the cutoff in that course). Any contemporaneous shock correlated across courses, or individual level unobserved characteristic that would make the same student systematically earn higher grades, shows no detected discontinuity at the cutoff (robust -0.037, s.e. 0.116; 95% CI $[-0.264, 0.190]$). Additionally, we look at outcome variables that capture how likely it is that the student in question lands just above a cutoff again in the following semester, which persistent luck, score targeting, or grader favoritism in general would move. We detect no difference between students earning a marginally higher grade and those earning the lower one in the likelihood of landing on the right side of a cutoff in the following semester or in later semesters (robust 0.043, s.e. 0.049; 95% CI $[-0.053, 0.139]$; mean 0.279).

3.3 Estimation sample for labor-market outcomes

Our labor-market estimates come from the same pooled specification as Equation (1), with an identical running variable, cutoffs, fixed effects, kernel, individual-level clustering, and MSE-optimal bandwidth selection, applied to outcomes measured at the student's first formal job and later in their labor history. The treatment is therefore identical throughout the paper, which keeps the estimand transparent and makes the validity evidence of Section 3.2 directly applicable.

We impose the following conditions on the labor-market sample. Observations are student-by-course records from (i) first-semester courses of (ii) students who enrolled at ages 17–19, restricted to (iii) students observed taking a first formal job after enrollment,

with months from enrollment to first job strictly positive, and with (iv) first jobs in the top 1% of the real-earnings distribution trimmed. Conditions (i)–(iii) leave 11,641 student-by-course observations (4,955 students), of which 7,950 (3,414 students) have observed, untrimmed first-job earnings, which is the estimation sample of the core labor tables, with Table 3 detailing the per-table variations. Condition (iii) drops students whose first BPS employment spell predates university, for whom the effect of grades on the first job is undefined, and students never observed in formal employment by the end of the data. Because entry into this sample is smooth through the cutoff (Section 3.2), and because the extensive-margin estimates of Section 5.1 show no effect on the timing or incidence of first employment, conditioning on it does not induce selection bias at the cutoff.

Because each labor-market outcome is observed once per student but enters the estimation once per student-by-course observation, a student contributes the same first-job outcome several times. This duplication is mechanical, in the sense that it affects weighting rather than identification, and standard errors clustered at the individual level account for it, as they do for any within-student correlation.

4 Results: Academic Outcomes

4.1 Academic performance

Figure 4 displays the RD plot for next-semester GPA, and Table 7 reports the estimates. Marginally earning a higher grade in a first-semester course increases GPA in the following semester by 0.563 points, or 0.23 standard deviations, off a left-of-cutoff mean of 6.20. The point estimate moves slightly to 0.604 with the inclusion of individual controls.

Next-semester GPA is computed as the simple average of the grades of all courses during the second semester alone, and therefore no part of this effect is mechanical. Whatever moves next-semester GPA seems to point to a change in the student's subsequent behavior. Additionally, this response is large relative to the shock that caused it. On average, during their first semester students earn a passing grade on between 2 and 3 courses, and only one in five students are scored in all four of them. As a consequence, this next-semester response is of the same order as that mechanical shock to their first-semester GPA, so the behavioral response in the following semester is comparable to the mechanical shock that caused it. This effect is persistent: cumulative GPA over all subsequent semesters, excluding the first, increases by 0.268 points, or 0.13 standard deviations.

Even though students observe their continuous scores, subsequent performance is still affected by their grades. Section 6 discusses potential mechanisms behind this result.

4.2 Attainment and persistence

In contrast to outcomes that measure students' future performance, those related to degree completion seem to be unaffected. For each student, we look at the effect of earning a higher grade on the probability of graduating 5 and 6 years after enrollment, along with the probability of ever graduating and the number of years until they complete their degree. Table 9 reports the results for these outcomes under the same RD research design specified before.[20]

We find no effects on the probability of graduating within five years of enrollment. Within six years, the point estimate is positive and marginally significant. However, the MSE-optimal bandwidth is unusually narrow, and the conventional estimate declines steadily as we widen it, so we read this result as suggestive. The indicator for ever graduating and years-to-degree among graduates show null results as well.

Since we don't observe when a student effectively drops out, we define a dropout in semester t as a student that has no academic activity recorded during the semesters between $t-2$ and t. Table 8 shows the results for indicators of dropping out between semesters 5 and 8. Point estimates are uniformly negative, and marginally significant at semester 6.

The overall pattern shown in these results points to higher grades having an important effect on students' future performance, with any response in whether they continue in school and how fast they complete their degree being modest and not robust.

4.3 Heterogeneity

Tables 10–12 report heterogeneous treatment effects on cumulative future GPA, estimated jointly by group with group-specific MSE-optimal bandwidths and robust bias-corrected inference implemented following Calonico et al. (2025). We find no systematic difference by gender, race, or first-generation status.

In terms of where in the grade distribution this response happens, the effect seems concentrated at low-grade cutoffs. Next-semester GPA displays the same pattern, with less precision. We explore this cutoff-level heterogeneity further under the multi-cutoff framework of Cattaneo et al. (2016). Table 13 reports the results. The point estimates decline essentially monotonically as we move towards higher grade cutoffs, even though cutoff-level inference is weak.[21] Additionally, the sample-share-weighted average of cutoff-

[20] Each "in-x-years" outcome is estimated only on cohorts observed for at least x years. Robustness checks looking at outcomes defined as indicators for graduating at each year starting in their 4th year after enrollment also support the null interpretation of the results.

[21] Robust confidence intervals span roughly a full GPA point on either side, with only the lowest cutoff

specific effects is smaller than the pooled estimate, which is the divergence expected when effects genuinely vary across cutoffs (Cattaneo et al., 2016).

Given these limitations, we treat the low/high binary split performed in estimating the results of Table 12 as the finest level of aggregation that our data can support. Section 6 considers what can account for these differential effects along the score distribution.

5 Results: Labor-Market Outcomes

Throughout this section, treatment is defined in the same way as in Section 4: marginally earning a higher transcript grade in one first-semester course, with the treatment effect estimated via Equation (1) on the labor sample of Section 3.3. The section is organized around three main questions. Does earning a higher grade change whether and when students start working? Does it change how much they earn? And does it change how much they work?

5.1 Effects on employment and the timing of entry

We find that marginally earning a higher grade has no extensive-margin effects. To assess how long it takes for students to enter the labor market, we look at how many months passed since enrollment until they start their first job. Also, to understand how "sticky" this first labor-market experience is, we compute the monthly length of this first employment spell. Table 15 shows that the point estimates for these outcomes are both statistically and economically insignificant.

Next, we look at the probability of being employed. We again find null results on employment during the year after they start their first job. Figure 6 further extends this analysis by plotting the month-by-month RD coefficients on employment probabilities, which suggests a null result at every horizon following the first job. These nulls provide evidence contrary to a search interpretation of our next results: marginally higher-graded students do not seem to be searching longer, being more selective, or delaying entry, since they enter the same labor market at the same time and hold jobs of the same duration.

Methodologically, this helps address sample selection issues from looking at labor outcomes. Our intensive-margin analysis conditions on observing a first job, so an effect on who enters the labor market would make that conditioning problematic in interpreting

being statistically significant, and the top cutoff having almost no identifying variation at all, with six effective observations on the left side of the threshold.

the results as a pure causal effect of grades.[22]

A potential additional concern is movement between formal and informal employment as another margin of adjustment, which our data cannot identify. In interpreting our results, we lean on our setting's low informality among salaried workers (Section 2.3), which limits the scope of this margin.[23]

5.2 Effects on monthly earnings

Table 16 reports effects on total real monthly earnings at the first job and the months following. At the first job, the estimate is -206 pesos on a control mean of 13,721, with a robust 95% confidence interval between -17% and $+11\%$ of the mean. The estimates on the following months are also small relative to the sample mean and statistically insignificant. Overall, we find no detectable change in monthly earnings, even though the null is not a precise zero, with the confidence intervals including moderate effects in either direction.

Without further analysis, the first row of Table 16 would lead us to say that grades do not have an effect on the transition into the labor market. Together with the previous subsection, it says that a higher grade changes neither whether a student works, nor when, nor for how long, nor how much money they take home. However, the results in the rest of this section show that this conclusion would give an incomplete picture of what grades might do.

5.3 Effects on hours worked and hourly wages

We find that marginally earning a higher grade changes the first job that students obtain, in terms of its hourly wage and hours, as well as the employer that provides it. The effects on wages and hours are reported in this subsection, and those on the employer in Section 5.5. Because the wage and hours effects go in opposite directions, they offset each other in monthly earnings, so the earnings null of the previous subsection is consistent with students holding different first jobs.

[22] A more direct test of this effect, which is the RD on the probability of appearing in the labor sample estimated on the full academic sample, is a precise zero (see Section 3.2).

[23] Informality also declines steeply with education. As a reference, in 2006 household-survey estimates, informal employment was 18.7 percent among workers with tertiary education against 41.2 percent nationally (Benedetti, 2007), and among salaried workers informality had fallen to approximately 7 percent by 2013 (Bergolo et al., 2026). Since the students in our sample are enrolled in university and hold salaried jobs, they likely belong to the group of workers with the lowest informality.

Hours worked. Table 19 reports the effect on hours worked. At the first job, marginally earning a higher grade reduces weekly hours worked by 2.23, on a control mean of 19.0 weekly hours. Additionally, the probability of working more than 20 hours per week, which is roughly the boundary between a full and a part-time job, falls by 10.7 percentage points off a base of 46%. This implies that marginal students appear to be shifted out of long-hours first jobs.

Looking at the same outcome later on shows similar results. The estimated effect on weekly hours worked 6 months after starting the first job is -2.83 weekly hours on a mean of 28.5, which is roughly a 10% reduction in hours worked. By twelve months, the point estimate attenuates to -1.16 , which is part of the dynamic pattern documented in Section 5.4.

Hourly wages. Table 18 reports the effect on hourly wages. Marginally earning a higher grade raises log real hourly wages by 0.155 at the first job. The result holds when taking into account different sampling restrictions discussed in Section 3.3. This increase in hourly wages, in contrast to the hours worked result, seems to be a feature of the start of the working relationship. At six months, the point estimate falls to 0.013, and at twelve months to 0.017, both of which are small, insignificant, and well below the entry-level estimate. The estimates over the first two years in Section 5.4 trace this decline in more detail. Since retention at the first firm is unaffected (Section 5.5), and so is the probability of being employed (Section 5.1), the decline does not seem to come from marginally higher-graded students losing their first jobs or leaving the labor market. This suggests that the wage itself is what adjusts, as predicted by employer-learning models in which firms rely on the credential when setting the entry wage, but not once they have further evidence of worker productivity (Farber and Gibbons, 1996; Altonji and Pierret, 2001; Lange, 2007).

Monthly earnings, wages, and hours on a common sample. The two effects above are not independent of the null in Section 5.2. Monthly earnings are the hourly wage multiplied by the hours worked,

$$\ln(\text{monthly earnings}) \;=\; \ln(\text{hourly wage}) \;+\; \ln(\text{monthly hours}),$$

so a wage increase of roughly the size of the hours reduction leaves earnings relatively unchanged. Table 17 shows the results of estimating all three terms on one identical sample by using a fixed bandwidth (1.93 score points, the MSE-optimal choice for the earnings

equation). The effect on log monthly earnings is 0.028, and it splits into an effect on the hourly wage of $+0.140$ and an effect on hours worked of -0.112. Marginally earning a higher grade therefore raises what a student is paid per hour by roughly 14 percent and reduces the hours they work by roughly 11 percent, with the two responses cancelling almost exactly. In general, a zero earnings effect could mean that grades genuinely have no effect. Here, since we are able to decompose earnings into its hourly components, we can see that it hides important compositional effects moving in opposite directions.

5.4 Effects over the first two years

The estimates above describe the job at its start. In this section, we study the dynamics of these effects. For each outcome, we estimate Equation (1) on the same outcome averaged within three-month windows, alongside the first-job estimate itself, while holding the bandwidth fixed across horizons at the first-job optimum so that the profile reflects the outcome's evolution and not sampling differences.[24] Figure 7 plots the estimated coefficients. The estimates at later horizons are less precise, because the effective sample falls by roughly 40 percent from entry to the final window, with some spells ending and others being censored. We therefore focus on the overall patterns more than on individual coefficients.

In terms of the wage premium, these results point to it being an entry phenomenon: a significant increase at the start, still visible until half a year after beginning the first job, but converging to zero after that. In contrast, the difference in weekly hours worked appears to be more permanent, with a steadily negative coefficient lasting at least until 18 months after the first job begins. Lastly, earnings follow the implied difference of the two previous results, being indistinguishable from zero while the wage advantage offsets the hours gap, and drifting slightly negative later on, reaching a difference of $-4{,}874$ pesos at months 15–17 (roughly 14 percent of the control mean).

5.5 Effects on employer characteristics and match survival

The results discussed so far cannot say whether a marginally higher-graded student has different outcomes because they are matched to different employers, or because their position within similar employers has different characteristics. We address this in this section by looking at average characteristics of the employers as outcomes.

We first look at the average hourly wage paid by the student's first firm as a measure to

[24] Per-horizon-optimal versions, showing qualitatively similar results, are in Figure A8.

approximate firm quality. Students who marginally earned a higher grade start their first jobs at firms that pay, on average, between 10 and 15 percent higher wages. Furthermore, Table 20 shows that a year after starting this job, these students are working at higher-paying firms, with firm average wages being 14 percent higher.

We next study whether these students are working at different types of firms. Interestingly, we find that the probability of working in the public sector increases by 7.0 percentage points off a base of 14.5%, a roughly 50 percent increase (Table 21). This initial difference in sector of employment is also persistent, with the estimates at six and twelve months being of the same order.[25] This result is in line with what we would expect given our institutional setting regarding public-sector hiring conditions (Section 2.3), in which academic records enter the hiring decision directly. In Section 6 we further decompose the wage and hours effects in a between and a within industries component, and not just sector of employment.

Finally, we look at the effect that higher grades have on the survival of these initial matches. To do this, we use as outcomes monthly indicators of being matched to the initial firm to assess changes in the probability that the worker remains employed at their first firm k months after starting the job. We find that grades affect which match is formed, but not whether it survives for longer (Figure 9).

6 Mechanisms and Interpretation

This section explores some of the potential mechanisms through which marginally earning a higher transcript grade affects students' future academic achievements and labor-market outcomes. On the academic side, students' future grades could improve either because students perform better in future courses, or because they make choices regarding what/how many courses to take that affect their future GPA. We present evidence that students respond by protecting their now higher GPA by taking less risky choices that could negatively impact their grades. On the labor-market side, higher grades could directly lead to better jobs because workers find employment in higher-paying industries, or because they sort into higher-paying firms. Our results point to the effect being mostly driven by within-industry differences, suggesting firm-level sorting as the main cause behind our findings.

[25] The month-by-month profile in Figure 8 shows similar results.

6.1 Interpreting the academic response

Given that students observe their underlying continuous score along with their grade, it is unlikely that the academic response reflects learning about their own ability, but rather a response to the transcript label itself (and their now-higher first-semester GPA). The heterogeneous pattern across the score distribution discussed before (Table 12), with the effect being concentrated at low-grade cutoffs, points to what kind of response this might be. A potential explanation is that, when students begin their second semester, they care about protecting a GPA that is now higher. The gradient can come from a marginally higher grade near the bottom of the scale setting a standard a student can plausibly achieve in the following semesters, while this is harder to do at the top of the distribution.[26]

This type of optimizing behavior has a concrete margin through which it could operate in our setting. As described in Section 2.2, a failed final exam is recorded as a zero on the transcript and thus hurts the GPA significantly, while falling short of the standard coursework path leaves the transcript unchanged. A student concerned with protecting their record should therefore attempt to avoid these risky exams.

In the first exam period after first-semester grades are posted, marginally higher-graded students attempt 0.37 to 0.40 fewer exams, roughly 30 percent of the control mean. Among those who attempt any exam, higher-graded students fail 0.23 to 0.25 fewer of them, roughly half of it, and are 13 percentage points less likely to have any zero recorded in their transcript, off a base of 29 percent (Table 14, Panel A). These differences disappear over the following years, which suggests that this response is not a permanent change in how students choose to earn their course-credits. Importantly, we find no effect on the number of courses passed, credits obtained, or credits attempted on the next semester (Table 14, Panel A), which suggests that this deferral does not slow progress towards their degree. A summary index of the protective margins moves by 0.12 to 0.14 control-side standard deviations over the next semester and is concentrated at the low cutoffs (Table 14, Panel C), which are the cutoffs at which the GPA response itself is concentrated.[27]

This mechanism carries a substantial share of the main academic effects, with somewhat more than half of it operating through zeros that are avoided from students' transcripts (Table 14, Panel B). Conditional on sitting these risky final exams, exam grades are higher, particularly at the low cutoffs, which is consistent with students deferring until they are prepared to pass.

[26] A standard reference-dependent model with convex effort costs predicts this behavior.

[27] The low/high estimates for the exam margins use split-sample estimation with within-group cutoff fixed effects rather than the joint estimator of Table 12.

The academic response is therefore not purely an improvement in students' future performance. A substantial part of it is a change in which activities students attempt, and as a consequence which make it into their transcript, along with better outcomes on the ones they do decide to take.

6.2 Interpreting the wage and sector effects

As we discussed before, the academic response is concentrated at the bottom of the grade distribution. This is not the case for the labor-market results.

Estimating heterogeneous effects for every first-job outcome by low versus high grade cutoffs yields statistically indistinguishable effects (see Table A7 for the main outcomes). This suggests that, while the effects of a higher grade depend on where on the distribution the grade jump happens for outcomes that are exclusively determined by the students' choices, for outcomes that potentially depend on how outsiders react to differences in students' grades, earning a higher grade seems to have a uniform return over the grade distribution.

Employers only observe grades and GPA, with no information on the underlying scores. A marginally higher grade raises first-job hourly wages, shifts entry toward the public sector, and raises the pay level of the employing firm over the first year (Section 5.5). Taken at face value, this reads as credential evidence, since the research design is shifting transcript grades while holding the underlying performance fixed at the threshold.

One issue with interpreting our findings this way is that the design does not hold subsequent human capital fixed.[28] As discussed before, a higher grade itself changes behavior, increasing future academic performance. By the time a student enters the labor market and reaches an applicant pool, the marginally higher grade from the first semester is already a bundle of the mechanical transcript jump from that semester, along with the improved subsequent performance, which could operate as an additional signal or, potentially, human capital accumulation. The labor-market effects discussed here are, in principle, the effects of this bundle, and the design cannot separate them entirely.

However, we are able to say something about where these differences seem to be coming from. Since we can observe an industry of employment for essentially every job (Ap-

[28] This issue is not unique to our setting. It applies to any design based on grade cutoffs in which the grade arrives before students finish their schooling, because the marginal grade can then affect their subsequent academic choices and the human capital they accumulate. The grade-cutoff estimates of Tan (2023), for example, capture both the effect of the grade itself and the effect of these later changes. In designs based on terminal credentials, as in Clark and Martorell (2014), this issue does not arise.

pendix C), we can decompose our labor-market results by assigning to each worker the (control-side) average of the outcome of interest for the industry in which they are employed, and estimating the same RD on this outcome. This treatment effect can then be interpreted as the part of the labor-market effect that we can attribute to differences between industries in which workers begin their careers. We can also compute the difference between each individual's realized outcome and this industry-average outcome, which we interpret as the part of the effect that comes from differences within the same industry. Performing this decomposition exercise, we find that the effect on the between component is null, while the within component seems to be responsible for the entire estimated treatment effect (see Table A8).

Additionally, the wage premium results survive after conditioning on the first job being in the private sector, so it is not the case that higher entry wages in the public sector are driving these results. Its dynamic results also hold within industries. Overall, this suggests that the wage premium is paid within industries, and not by marginal students matching into industries with higher average pay.

6.3 Interpreting the reduction in hours worked

One of our main findings is that students who marginally earn a higher grade work fewer hours when they enter the labor market. This subsection suggests an interpretation for this result.

A first possibility is that students are choosing to work fewer hours. In our setting, jobs are held by currently enrolled college students. If work serves to finance consumption while studying, then a decrease in hours worked and an increase in hourly wages that compensate each other, which is what we find, could be an optimal response if a higher grade expands their opportunity set in terms of which jobs they can choose from.

The second possibility is that the jobs themselves that students manage to get come with fewer hours. Since the hours of a job usually come together with its wage and its other amenities, some firms may offer shorter standard hours than others, and the public sector, where entry rises sharply, is one plausible example. In that case, students could be working fewer hours as a consequence of moving to jobs in which shorter hours come with the position, and the public-sector shift documented in Section 5.5 is one potential channel.

Applying the same decomposition of Section 6.2 with weekly hours worked as our outcome, we find that the effect on the between-industry component is again null, so the effect on hours does not seem to come from industry-specific job types. Instead, the

reduction takes place within them, which points to different hours chosen within similar jobs.

Finally, for an enrolled student, hours of work imply giving up hours for study, which is an input that can affect the performance of working students (Stinebrickner and Stinebrickner, 2003; Neyt et al., 2019). Therefore, an important part of the return to a higher grade could be taken in time. To explore this channel, we use as outcomes, for each student, the change in GPA between the last semester before and the first semester after the job begins. We find that this measure of performance increases by 0.52, the number of activities they enroll in rises, and GPA two semesters after starting their first job is also 0.58 points higher (Table 22).

7 Conclusion

This paper studied the effect of marginally earning a higher grade in a first-semester course on students' subsequent academic performance and on their first steps into the labor market, exploiting the deterministic grade-jump cutoffs at UdelaR, Uruguay's main higher education institution. We show evidence that students who land just above a grade-jump cutoff are observably very similar to ones who land just below it across a rich set of covariates.

As a consequence of earning a marginally higher grade, students' GPA in the following semester increases by 0.23 standard deviations, with the effect persisting into cumulative GPA and concentrated at cutoffs at the lower part of the score distribution. The response operates in substantial part through how students clear the courses that follow: marginally higher-graded students defer taking high-stakes exams that can have a heavy negative impact on their transcript, with somewhat more than half of the next-semester GPA effect coming from this channel.

On the labor-market side, marginally earning a higher grade changes the characteristics of the first job: hourly wages at the first job are 17 to 21 percent higher, weekly hours worked are 2.2 to 2.8 hours lower, and entry into public-sector jobs, where hiring rules score GPA explicitly, rises by 7 percentage points off a base of 14.5 percent. While the reduction in hours persists through the first year and a half, the wage premium fades quickly. The effect on earnings is null, with the increase in wages and decrease in hours offsetting each other. We find no evidence of higher grades affecting the extensive margin, with no changes in the timing of their first job or duration of the employment spell.

Earnings-based estimates of the returns to academic performance can understate them for working students. In our setting, an important part of the returns to grades at the

first job is taken in time rather than in money. Additionally, this reduction in hours worked is plausibly productive. We find increases in students' academic performance in the semesters immediately after they begin their first job, relative to students who also start working but earn a marginally lower grade.

Because students observe their continuous underlying scores, the academic response is unlikely to come from students learning about their own ability. We find that the grade itself can change behavior, and it does so unevenly, raising performance mostly at the bottom of the grade distribution. This change in performance is a direct consequence of higher-graded students opting into a less risky path in the semesters that follow, avoiding choices that could significantly hurt their GPA.

The higher grade also affects labor-market outcomes at the moment when employers know least about the potential workers, with the significant increase in public-sector employment, where GPA is used for hiring explicitly, and a wage premium paid within industries and fading over time forming a consistent picture of a transcript grade serving as a signaling device at hiring.

References

Atila Abdulkadiroğlu, Joshua D. Angrist, and Parag A. Pathak. The elite illusion: Achievement effects at Boston and New York exam schools. *Econometrica*, 82(1):137–196, 2014.

Joseph G. Altonji and Charles R. Pierret. Employer learning and statistical discrimination. *Quarterly Journal of Economics*, 116(1):313–350, 2001.

Tihitina Andarge, Dalia Ghanem, David A. Keiser, and Gabriel E. Lade. Quantifying threshold manipulation in the presence of rounding: The case of lead monitoring in U.S. drinking water. *American Economic Review: Insights*, 2025. Forthcoming.

Ghazala Azmat and Nagore Iriberri. The importance of relative performance feedback information: Evidence from a natural experiment using high school students. *Journal of Public Economics*, 94(7–8):435–452, 2010.

Ghazala Azmat, Manuel Bagues, Antonio Cabrales, and Nagore Iriberri. What you don't know…can't hurt you? A natural field experiment on relative performance feedback in higher education. *Management Science*, 65(8):3714–3736, 2019.

Gary S. Becker. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. Columbia University Press, New York, 1964.

Enrico Benedetti. Empleo informal en el Uruguay. Informe temático, encuesta nacional de hogares ampliada 2006, Instituto Nacional de Estadística, Montevideo, 2007.

Marcelo L. Bergolo, Rodrigo Ceni, Mathias Fondo, and Damián Vergara. Minimum wages and the distribution of firm wage premia. Working Paper 34188, National Bureau of Economic Research, 2026. Issued August 2025, revised July 2026.

Kirill Borusyak, Matan Kolerman-Shemer, and Lihua Lei. Regression discontinuity aggregation, with an application to union effects on inequality. Working paper, 2026.

Sebastian Calonico, Matias D. Cattaneo, and Max H. Farrell. Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *Econometrics Journal*, 23(2):192–210, 2020.

Sebastian Calonico, Matias D. Cattaneo, Max H. Farrell, Filippo Palomba, and Rocío Titiunik. Treatment effect heterogeneity in regression discontinuity designs. Working paper, July 2025, 2025.

David Card. The causal effect of education on earnings. In Orley C. Ashenfelter and David Card, editors, *Handbook of Labor Economics*, volume 3A, pages 1801–1863. Elsevier, Amsterdam, 1999.

David Card, Jörg Heining, and Patrick Kline. Workplace heterogeneity and the rise of West German wage inequality. *Quarterly Journal of Economics*, 128(3):967–1015, 2013.

David Card, Ana Rute Cardoso, and Patrick Kline. Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women. *Quarterly Journal of Economics*, 131(2):633–686, 2016.

Matias D. Cattaneo, Luke Keele, Rocío Titiunik, and Gonzalo Vázquez-Baré. Interpreting regression discontinuity designs with multiple cutoffs. *Journal of Politics*, 78(4):1229–1248, 2016.

Matias D. Cattaneo, Michael Jansson, and Xinwei Ma. Simple local polynomial density estimators. *Journal of the American Statistical Association*, 115(531):1449–1455, 2020a.

Matias D. Cattaneo, Rocío Titiunik, and Gonzalo Vázquez-Baré. Analysis of regression discontinuity designs with multiple cutoffs or multiple scores. *The Stata Journal*, 20(4): 866–891, 2020b.

Raj Chetty, David J. Deming, and John N. Friedman. Diversifying society's leaders? The determinants and causal effects of admission to highly selective private colleges. *Quarterly Journal of Economics*, 141(1):51–145, 2026.

Damon Clark and Paco Martorell. The signaling value of a high school diploma. *Journal of Political Economy*, 122(2):282–318, 2014.

Stacy Berg Dale and Alan B. Krueger. Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables. *Quarterly Journal of Economics*, 117(4):1491–1527, 2002.

Henry S. Farber and Robert Gibbons. Learning and wage dynamics. *Quarterly Journal of Economics*, 111(4):1007–1047, 1996.

Andy Feng and Georg Graetz. A question of degree: The effects of degree class on labor market outcomes. *Economics of Education Review*, 61:140–161, 2017.

Ronny Freier, Mathias Schumann, and Thomas Siedler. The earnings returns to graduating with honors: Evidence from law graduates. *Labour Economics*, 34:39–50, 2015.

Anne Toft Hansen, Ulrik Hvidman, and Hans Henrik Sievertsen. Grades and employer learning. *Journal of Labor Economics*, 42(3):659–682, 2024.

Mark Hoekstra. The effect of attending the flagship state university on earnings: A discontinuity-based approach. *Review of Economics and Statistics*, 91(4):717–724, 2009.

Instituto Nacional de Estadística. Anuario Estadístico Nacional 2023. Technical report, Instituto Nacional de Estadística, Montevideo, 2023. Table 9.1.9, cotización interbancaria de las principales monedas por mes. https://www.gub.uy/instituto-nacional-estadistica/comunicacion/publicaciones/anuario-estadistico-nacional-2023-volumen-n-100, accessed 2026-09-24.

Pauline Khoo and Ben Ost. The effect of graduating with honors on earnings. *Labour Economics*, 55:149–162, 2018.

Fanny Landaud, Éric Maurin, Barton Willage, and Alexander Willén. The value of a high school GPA. *Review of Economics and Statistics*, pages 1–24, 2024.

Fabian Lange. The speed of employer learning. *Journal of Labor Economics*, 25(1):1–35, 2007.

Thomas Le Barbanchon, Diego Ubfal, and Federico Araya. The effects of working while in school: Evidence from employment lotteries. *American Economic Journal: Applied Economics*, 15(1):383–410, 2023.

Hongyan Li and Xing Xia. Grades as signals of comparative advantage: How letter grades affect major choices. *Journal of Economic Behavior & Organization*, 227:106717, 2024.

Joyce B. Main and Ben Ost. The impact of letter grades on student effort, course selection, and major choice: A regression-discontinuity analysis. *Journal of Economic Education*, 45 (1):1–10, 2014.

Jacob Mincer. *Schooling, Experience, and Earnings*. Columbia University Press, New York, 1974.

NACE. Job outlook 2026. Technical report, National Association of Colleges and Employers, Bethlehem, PA, 2025.

Brecht Neyt, Eddy Omey, Dieter Verhaest, and Stijn Baert. Does student work really affect educational outcomes? A review of the literature. *Journal of Economic Surveys*, 33 (3):896–921, 2019.

Ronald Oaxaca. Male-female wage differentials in urban labor markets. *International Economic Review*, 14(3):693–709, 1973.

Zhuan Pei, David S. Lee, David Card, and Andrea Weber. Local polynomial order in regression discontinuity designs. *Journal of Business & Economic Statistics*, 40(3):1259–1267, 2022.

Michael Spence. Job market signaling. *Quarterly Journal of Economics*, 87(3):355–374, 1973.

Ralph Stinebrickner and Todd R. Stinebrickner. Working during school and academic performance. *Journal of Labor Economics*, 21(2):473–491, 2003.

Brandon Joel Tan. The consequences of letter grades for labor market outcomes and student behavior. *Journal of Labor Economics*, 41(3):565–588, 2023.

Universidad de la República. Síntesis estadística por servicio universitario: Facultad de Ciencias Económicas y de Administración. Technical report, Dirección General de Planeamiento, Universidad de la República, Montevideo, December 2019. https://planeamiento.udelar.edu.uy/wp-content/uploads/sites/33/2020/02/S%C3%ADntesis-estad%C3%ADstica-CEyA-2019.pdf, accessed 2026-09-24.

Universidad de la República. Estadísticas Básicas 2018 de la Universidad de la República. Technical report, Dirección General de Planeamiento, Universidad de la República, Montevideo, 2020. Chapter 4.1, Ingresos de estudiantes a servicio, serie 2010–2018. https://planeamiento.udelar.edu.uy/portal/wp-content/uploads/sites/33/2020/02/web_Estadisticas-basicas_2018_20200203.pdf, accessed 2026-09-24.

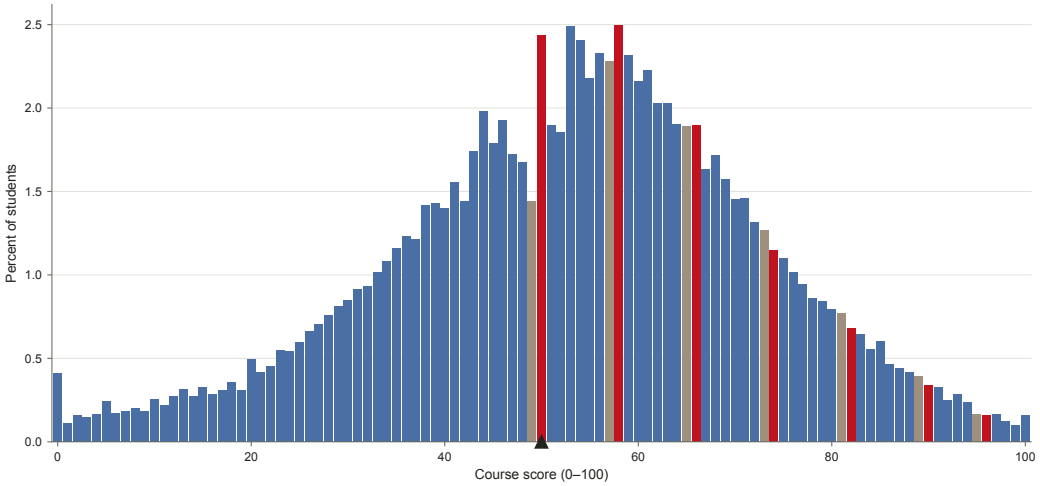
World Bank. Uruguay systematic country diagnostic update. Technical report, World Bank Group, Washington, DC, March 2022.

World Bank. World development indicators: Gini index (si.pov.gini). World Bank Open Data, 2026. Accessed 2026-08-10 via the World Bank API.

Seth D. Zimmerman. Elite colleges and upward mobility to top jobs and top incomes. *American Economic Review*, 109(1):1–47, 2019.

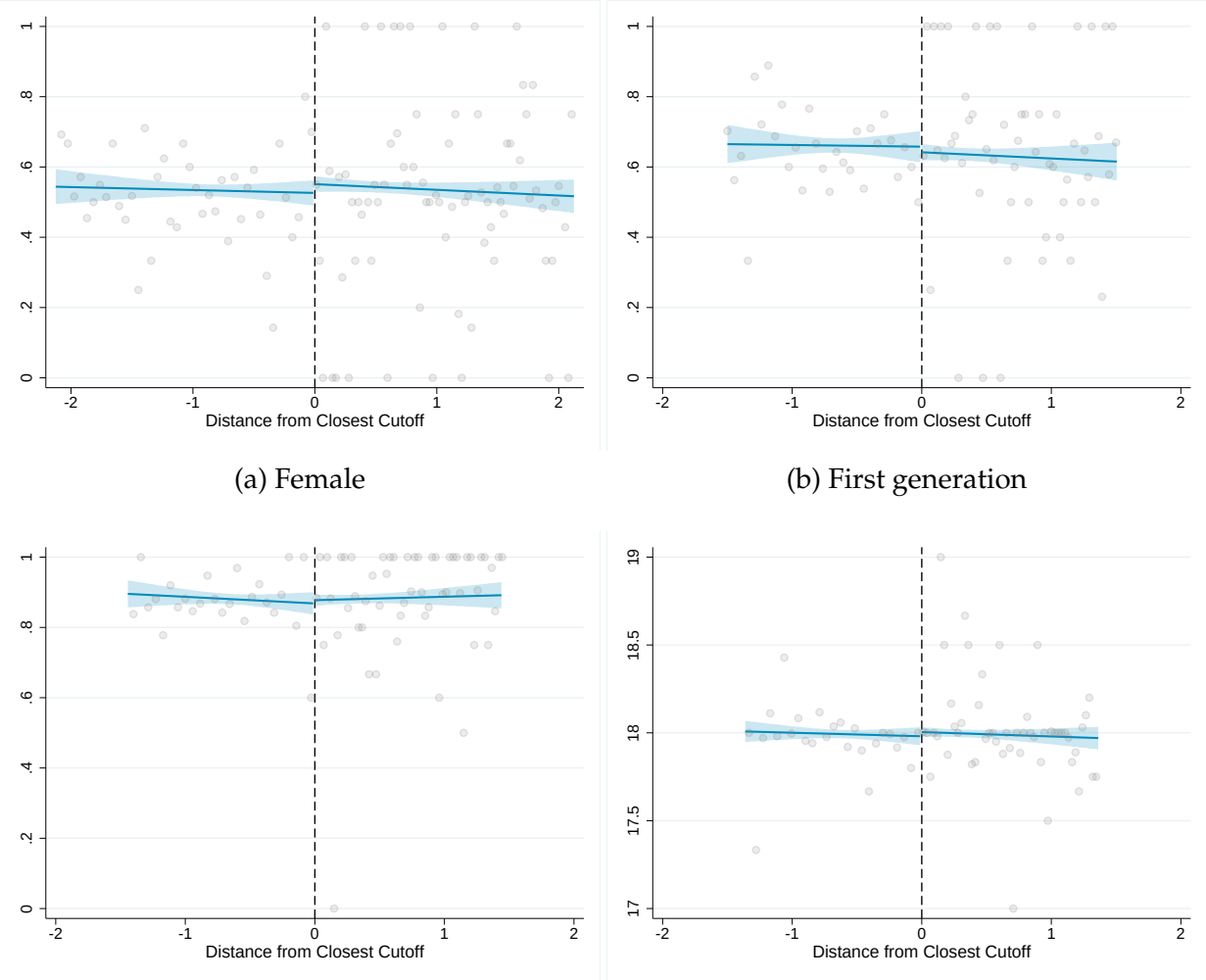
Figures and Tables

Figure 1: Distribution of course scores ↩ p. 14



Notes: Raw 0–100 scores of students who sat both midterm examinations in the five first-semester courses, 2013–2024, taken from the acta records (n = 71,564). Each bin is one point wide and centred on an integer. The bins at the pass mark of 50 and at the six grade-jump cutoffs (58, 66, 74, 82, 90, and 96) are shown in red, and the bin immediately below each of them in grey; the caret marks the pass mark. The pass margin is excluded from the design because of the visible excess mass at exactly 50 (Section 3.1). The rounding-aware manipulation test of Appendix B measures this excess mass and finds no similar excess at the six cutoffs used in the design.

Figure 2: Covariate balance at the grade cutoffs: demographics ↩ p. 15



(a) Female

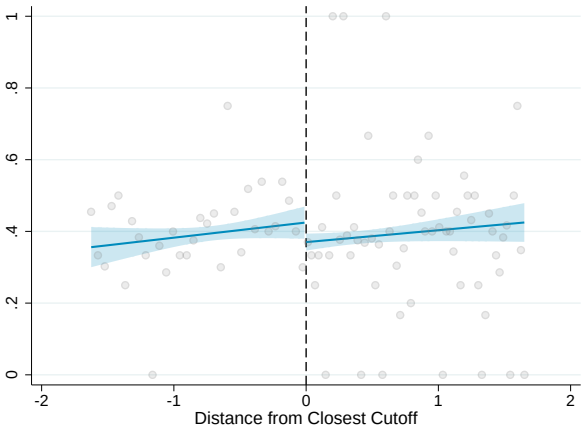
(b) First generation

(c) White

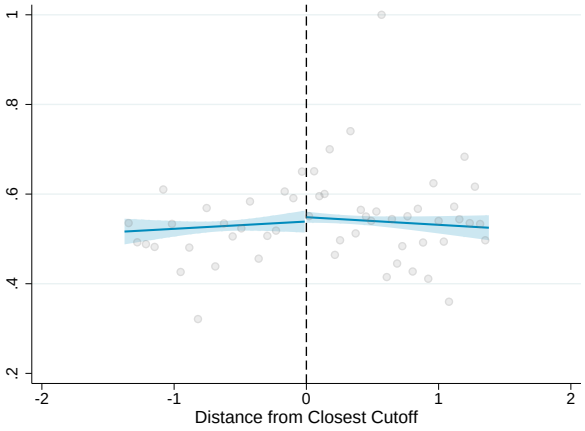
(d) Age at enrollment

Notes: RD plots of predetermined covariates against the running variable, pooled across the six grade-jump cutoffs. Gray circles are bin averages; lines are kernel-weighted linear fits on each side of the cutoff with 95 percent confidence bands, within the MSE-optimal bandwidth of each covariate. The corresponding estimates are in Table 4.

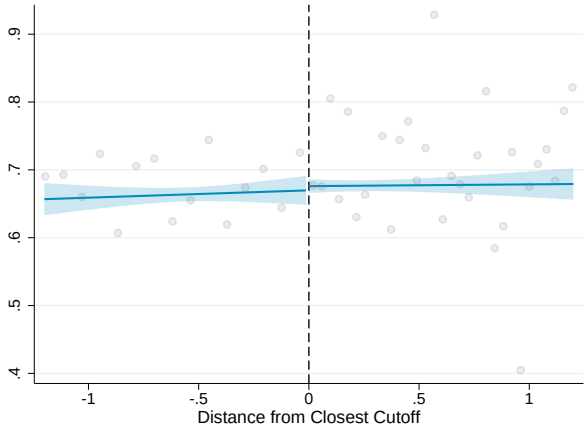
Figure 3: Covariate balance at the grade cutoffs: schooling and diagnostics ↩ p. 15



(a) Public high school



(b) Mathematics diagnostic



(c) Reading diagnostic

Notes: As in Figure 2. The diagnostic assessments are available for the post-2018 cohorts.

Figure 4: RD: next-semester GPA ↩ p. 16

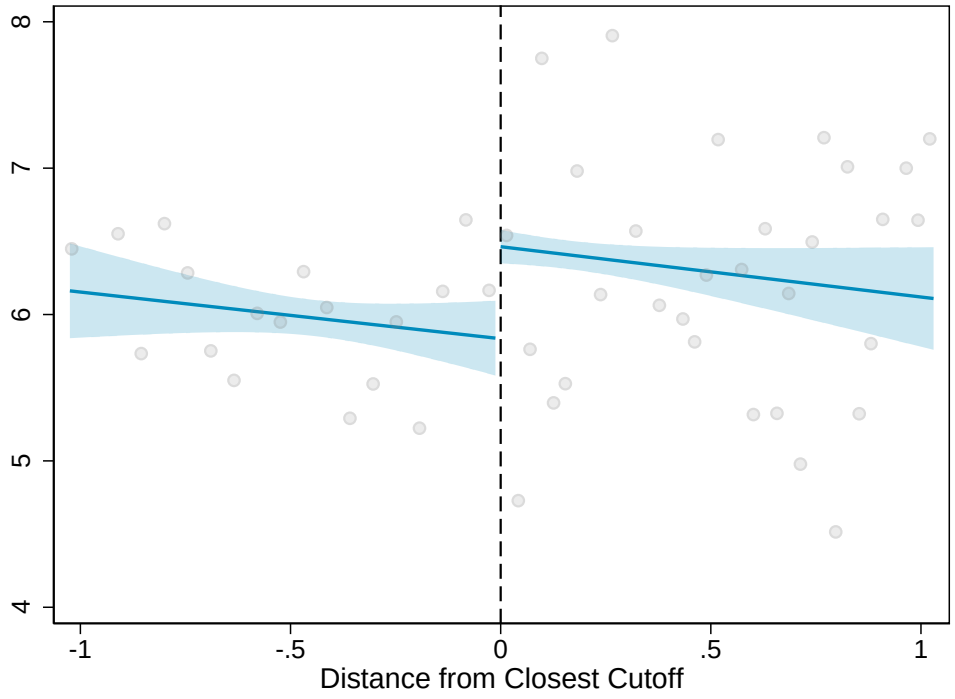
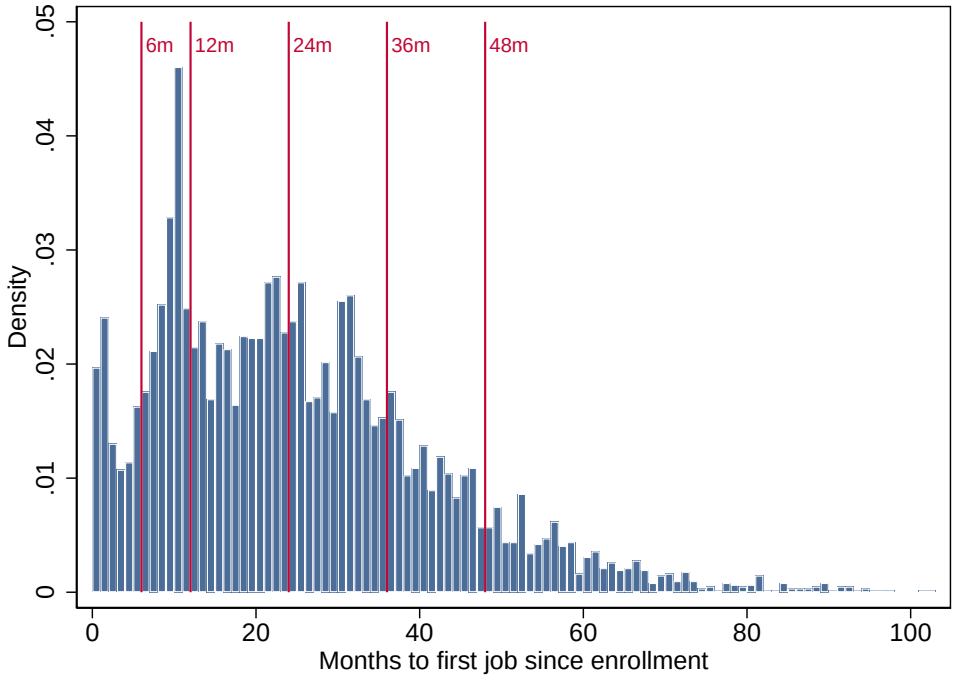
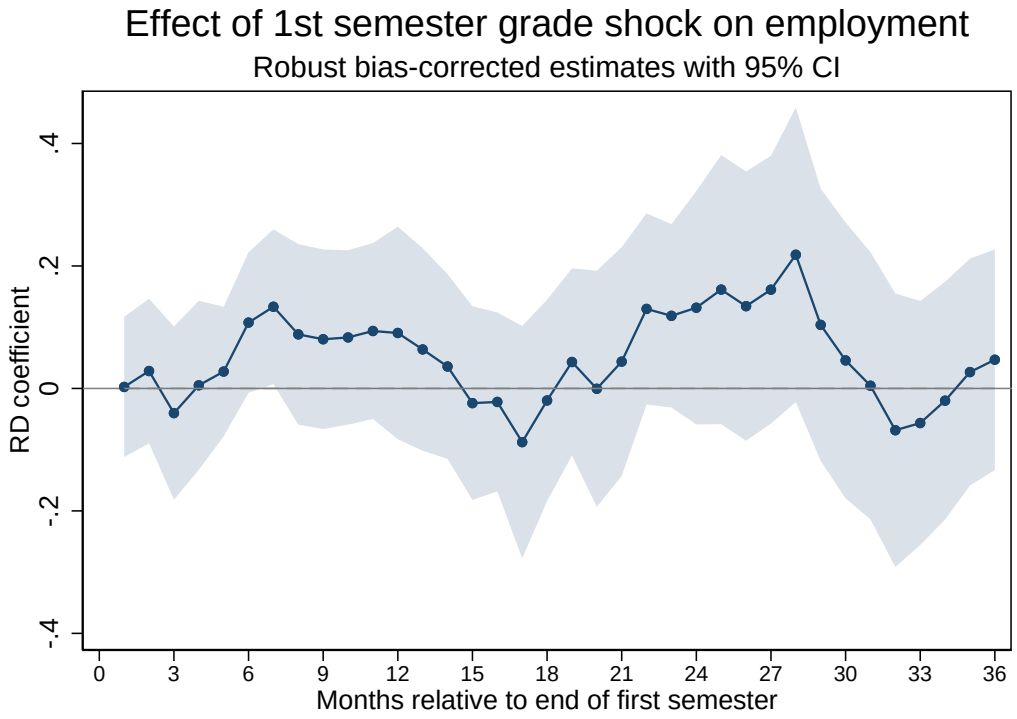


Figure 5: Months from enrollment to first formal job ↩ p. 8



Notes: Labor sample. Median 23 months; 54% of first jobs begin within 24 months of enrollment, which for most students is before graduation.

Figure 6: Employment probability by month since first semester ↩ p. 18



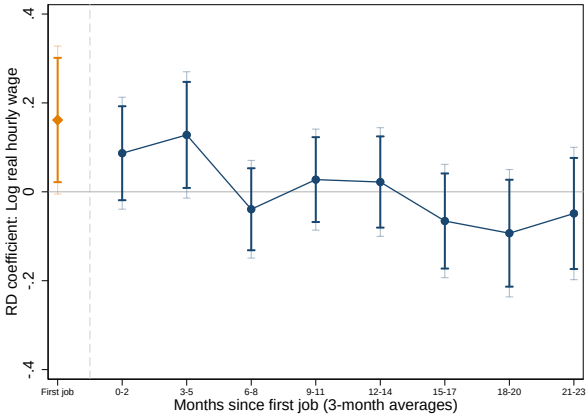
Notes: Month-by-month course-level RD coefficients on an employment indicator, months 1–36 after the first semester. 95% confidence intervals, standard errors clustered at the student level.

Table 1: Score-to-grade mapping ↩ p. 7

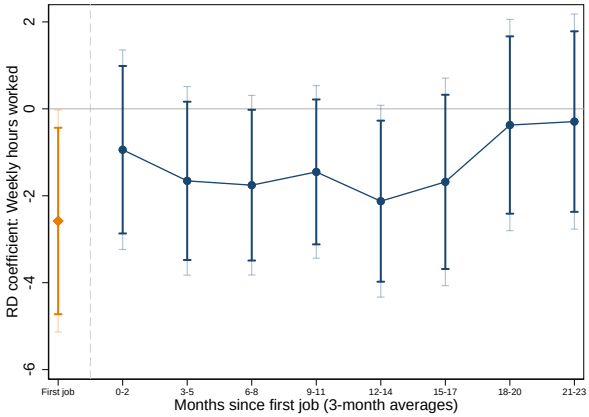
Raw score	Transcript grade
$[0, 50)$	Fail
$[50, 58)$	6
$[58, 66)$	7
$[66, 74)$	8
$[74, 82)$	9
$[82, 90)$	10
$[90, 96)$	11
$[96, 100]$	12

Notes: The mapping is set by university regulation and is the same for every course, instructor, and year, with no curve applied. The design uses the six cutoffs at which the grade jumps (58, 66, 74, 82, 90, and 96) and excludes the pass/fail threshold at 50 (Section 3.1).

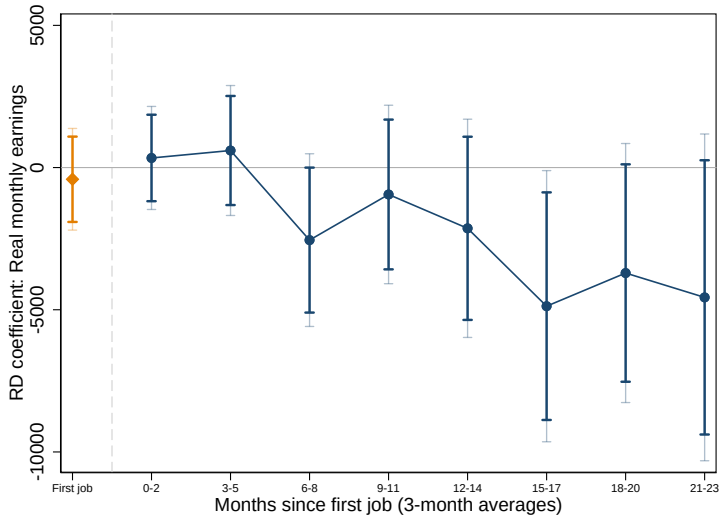
Figure 7: Dynamics at the first job: wage, hours, and earnings ↩ p. 21



(a) Log real hourly wage



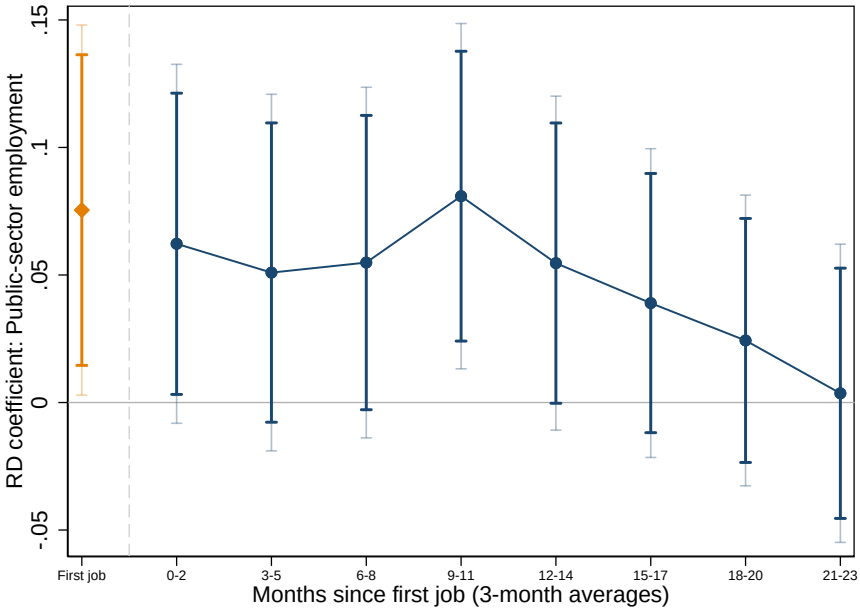
(b) Weekly hours



(c) Real monthly earnings

Notes: RD estimates on outcomes averaged within three-month windows of time since the first-job month, with the stand-alone first-job estimate at the left of each panel (separated by the dashed line). Bandwidth fixed across horizons at the first-job MSE-optimal value for each outcome. Whiskers: dark = 90% CI, light = 95% CI, clustered at the student level. Per-horizon-optimal-bandwidth versions in Figure A8.

Figure 8: Dynamics of public-sector employment ↩ p. 22



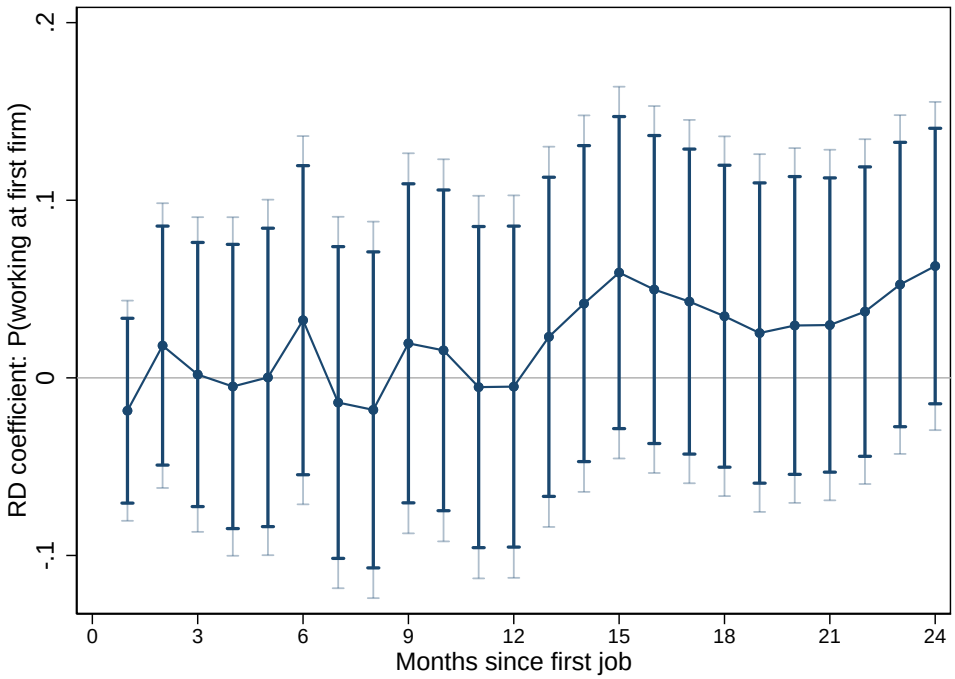
Notes: Design as in Figure 7.

Table 2: Descriptive statistics ↩ p. 10

	(1) All enrollees		(2) Ages 17–19	
	mean	sd	mean	sd
Woman	0.57	0.49	0.53	0.50
Age	23.46	7.09	18.12	0.61
Years since high school	4.50	5.92	1.07	0.51
Native born	0.96	0.19	0.98	0.13
Montevideo	0.58	0.49	0.51	0.50
White	0.83	0.38	0.83	0.38
Father's education	5.08	2.25	5.44	2.37
Mother's education	5.48	2.28	5.92	2.36
First generation	0.81	0.40	0.74	0.44
Public high school	0.22	0.41	0.30	0.46
Accountancy	0.47	0.50	0.64	0.48
Economics	0.08	0.27	0.12	0.33
Administration	0.13	0.34	0.08	0.28
Administration technician	0.24	0.43	0.09	0.29
Works in first enrollment year	0.46	0.50	0.11	0.31
Observations	34183		13937	

Notes: Column (1): full enrollee population. Column (2): students enrolling at ages 17–19, the analysis population (Section 2.5). Parental education in years.

Figure 9: Retention at the first firm ↩ p. 22



Notes: RD estimates on the probability of remaining employed at the first firm k months after the match begins, unconditional (leavers count as zeros, and months beyond the panel end are censored). The bandwidth is fixed at every horizon at the MSE-optimal value at month 6. Control-group retention is 76, 63, 47, 32, and 23% at months 3, 6, 12, 18, and 24. Whiskers as in Figure 7.

Table 3: Analysis samples and estimation counts ↩ p. 16

	Observations	Students
Panel A. Academic course-level sample (student×course)		
First-semester scores near a grade cutoff, ages 17–19	13,626	5,888
of which: next-semester GPA observed	12,915	5,414
of which: cumulative future GPA observed	13,216	5,623
Panel B. Labor course-level sample		
+ first formal job begins after enrollment	11,641	4,955
of which: first-job earnings observed	8,055	—
after top-1% earnings trim (core labor tables)	7,950	3,414
weekly-hours table (both 1% tails trimmed)	7,887	3,382
hourly-wage table (first jobs beginning by Dec. 2021)	6,224	2,688
hourly-wage table, triangular kernel, no date restriction	7,950	3,414
Panel C. RDA student-level sample (Appendix A)		
Students in the first-job build	—	9,747
IV conditions (first job >6 months after enrollment; ≥1 scored course)	—	5,144
of which: wage outcome observed (*N* of the IV tables)	—	3,824

Notes: Each indented row adds one condition to the row above it, and Sections 2.5 and 3.3 describe each condition and why we impose it. The estimation *N* in each results table corresponds to one of these rows. Panel C is at the student level, so it reports no course-level observation counts.

Table 4: Covariate balance at grade cutoffs ↩ p. 10

	Female	First Gen	White	Age	Public HS	Math	Reading
Robust	0.033	-0.008	0.015	0.029	-0.065	0.003	0.004
	(0.036)	(0.042)	(0.031)	(0.045)	(0.041)	(0.022)	(0.021)
Bandwidth	2.130	1.512	1.448	1.363	1.654	1.383	1.221
Obs.	13626	13267	13445	13626	13439	8106	8086
Eff. obs.	7260	4914	4484	4413	5090	2712	2496
Clusters	4334	3504	3246	3190	3562	3224	3027

Notes: Robust bias-corrected RD estimates of eq. (1) on predetermined covariates, academic sample. Math and Reading are pre-enrollment diagnostic scores (post-2018 subsample). Standard errors, clustered at the student level, in parentheses; MSE-optimal bandwidths. Stars: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Balance: Big Five personality traits and diagnostic assessments ↩ p. 10

	Extraversion	Agreeableness	Conscientiousness	Neuroticism	Openness	Math	Reading
Conventional	-0.0466	-0.0635	-0.0725	0.197	0.0569	0.00595	0.00823
	(0.121)	(0.117)	(0.117)	(0.135)	(0.124)	(0.0205)	(0.0178)
Robust	-0.0741	-0.0839	-0.0807	0.237	0.0692	-0.000523	0.00305
	(0.145)	(0.139)	(0.144)	(0.162)	(0.150)	(0.0241)	(0.0224)
Bandwidth	1.5332	1.5779	1.3830	1.4132	1.6315	1.4848	1.1796
Obs.	4782	4782	4786	4783	4775	6583	6562
Eff. obs.	1794	1797	1607	1614	1831	2196	1998
Clusters	2797	2811	2569	2578	2836	2597	2410

Notes: Big Five factors (BFI-44, z-scored) measured at enrollment, labor sample; Math and Reading diagnostics as in Table 4. Conventional and robust bias-corrected estimates; standard errors, clustered at the student level, in parentheses. Stars: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Placebo outcomes ↩ p. 15

	(1)	(2)	(3)	(4)
Robust	-0.0352	0.0354	0.0438	0.0549
	(0.116)	(0.0673)	(0.0485)	(0.155)
Control mean	8.041	0.009	0.283	1.630
Bandwidth	1.241	1.218	1.189	1.183
Obs.	11515	13262	13262	13262
Eff. obs.	3579	4026	4023	4013
Clusters	3028	3027	3026	3023

Notes: Column (1): same-semester GPA in the student's other first-semester courses. Columns (2)–(4): indicators/counts for landing within one point above a cutoff in the following semester and in all future semesters. Robust bias-corrected estimates; standard errors, clustered at the student level, in parentheses. Stars: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Main academic results: next-semester and future GPA ↩ p. 16

	(1) GPA next sem.	(2) GPA next sem.	(3) Future GPA	(4) Future GPA
Robust	0.563**	0.604**	0.268*	0.336**
	(0.239)	(0.240)	(0.148)	(0.150)
Control mean	6.197	6.197	5.822	5.823
Bandwidth	1.036	1.032	1.748	1.645
Obs.	12915	12634	13216	12925
Eff. obs.	3777	3710	5074	4905
Clusters	2958	2958	3599	3562

Notes: Robust bias-corrected RD estimates of eq. (1). Columns (2) and (4) add individual controls (female, white, first-generation, age at enrollment). Outcome SDs: next-semester GPA 2.47; future GPA 2.08. Standard errors, clustered at the student level, in parentheses; MSE-optimal bandwidths; cutoff fixed effects throughout. Stars: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: Dropout ↩ p. 17

	(1) Semester 5	(2) Semester 6	(3) Semester 7	(4) Semester 8
Robust	-0.0160	-0.0565*	-0.0418	-0.0442
	(0.0209)	(0.0304)	(0.0377)	(0.0380)
Control mean	0.0861	0.1197	0.2314	0.2619
Bandwidth	1.879	1.393	1.484	1.510
Obs.	13262	13262	13262	13262
Eff. obs.	5401	4391	4445	4913
Clusters	3764	3226	3260	3504

Notes: Indicators for having no course activity by the stated semester. Conventional and robust bias-corrected estimates; standard errors, clustered at the student level, in parentheses. Stars: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 9: Graduation ↩ p. 17

	Grad. in 5y	Grad. in 6y	Ever graduates	Years to degree
Conventional	-0.00296	0.136*	0.0634	-0.0434
	(0.0446)	(0.0753)	(0.0737)	(0.155)
Robust	0.00726	0.154*	0.0682	-0.0565
	(0.0542)	(0.0885)	(0.0880)	(0.186)
Control mean	0.487	0.556	0.609	4.485
Bandwidth	1.237	0.683	0.752	1.390
Obs.	8020	6379	6379	3874
Eff. obs.	2424	933	1103	1341
Clusters	1811	833	965	913

Notes: Graduations observed through 2024; each within-x-years column is estimated on cohorts with at least x years of follow-up (within-5: enrolled $\leq$2019; within-6 and the remaining columns: enrolled $\leq$2018). Years to degree is measured only for graduates, a selected group, and is descriptive. For the within-6 column the MSE-optimal bandwidth is narrow (0.69, with 972 effective observations), and the conventional estimate falls at each wider fixed bandwidth, to 0.087 at $h = 1$, 0.056 at $h = 1.5$, and 0.031 at $h = 2$ (Section 4). Conventional and robust bias-corrected estimates; standard errors, clustered at the student level, in parentheses. Stars: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 10: Heterogeneity by gender ↩ p. 17

Group	RD effect	95% robust CI	Eff. obs.	h
Women	0.346	$[-0.158,\ 0.993]$	2,644	1.63
Men	0.220	$[-0.263,\ 0.780]$	2,284	1.73

Notes: Outcome: cumulative future GPA. RD heterogeneous treatment effects estimated jointly by group (Calonico et al., 2025): covariate-adjusted, group-specific MSE-optimal bandwidths, triangular kernel, standard errors clustered at the student level. Confidence intervals are robust bias-corrected and therefore not centered at the conventional point estimates. Effective observations are those within the group's bandwidth.

Table 11: Heterogeneity by first-generation status

Group	RD effect	95% robust CI	Eff. obs.	h
Not first-generation	0.312	$[-0.299,\ 1.094]$	1,732	1.64
First-generation	0.305	$[-0.138,\ 0.793]$	3,174	1.66

Notes: Outcome: cumulative future GPA. RD heterogeneous treatment effects estimated jointly by group (Calonico et al., 2025): covariate-adjusted, group-specific MSE-optimal bandwidths, triangular kernel, standard errors clustered at the student level. Confidence intervals are robust bias-corrected and therefore not centered at the conventional point estimates. Effective observations are those within the group's bandwidth.

Table 12: Heterogeneity by low vs. high cutoff ↩ p. 17

Group	RD effect	95% robust CI	Eff. obs.	h
Low cutoffs (jumps into 7, 8, 9)	0.449	$[\ 0.011,\ 1.119]$	3,272	1.46
High cutoffs (jumps into 10, 11, 12)	-0.457	$[-1.788,\ 0.105]$	494	0.84

Notes: Outcome: cumulative future GPA. RD heterogeneous treatment effects estimated jointly by group (Calonico et al., 2025): covariate-adjusted, group-specific MSE-optimal bandwidths, triangular kernel, standard errors clustered at the student level. Confidence intervals are robust bias-corrected and therefore not centered at the conventional point estimates. Effective observations are those within the group's bandwidth. Low cutoffs include the thresholds at 58, 66, and 74, and high cutoffs those at 82, 90, and 96. Effects on next-semester GPA show the same pattern with less precision.

Table 13: Cutoff-by-cutoff estimates ↩ p. 17

	GPA next semester			Future GPA		
Cutoff	RD effect	95% robust CI	Eff. obs.	RD effect	95% robust CI	Eff. obs.
58 (jump into 7)	1.488	[0.61, 2.81]	659	0.391	[−0.25, 1.26]	1,058
66 (jump into 8)	0.578	[−0.41, 1.77]	992	0.536	[−0.24, 1.45]	641
74 (jump into 9)	0.486	[−0.53, 1.76]	458	0.375	[−0.37, 1.10]	938
82 (jump into 10)	−0.103	[−1.20, 0.80]	526	−0.327	[−1.47, 0.52]	304
90 (jump into 11)	−0.724	[−2.02, 0.07]	279	−0.291	[−1.42, 0.63]	338
96 (jump into 12)	0.513	[−0.57, 1.37]	48	−1.212	[−2.48, −0.23]	48
Weighted average	0.522	[0.07, 1.10]	2,962	0.256	[−0.11, 0.66]	3,327
Pooled	0.674	[0.29, 1.30]	3,738	0.473	[0.18, 0.85]	4,831

Notes: Multi-cutoff RD following Cattaneo et al. (2016), estimated with `rdmc` (Cattaneo et al., 2020b) under the paper's baseline specification at each cutoff: local linear, triangular kernel, per-cutoff MSE-optimal bandwidths, individual controls, standard errors clustered at the student level. Point estimates are conventional and confidence intervals are robust bias-corrected. The weighted average uses each cutoff's share of the sample as its weight. The pooled row normalizes and pools the score without cutoff fixed effects, so it differs somewhat from Table 7. Cutoff 96 has almost no observations to the left of the threshold within the bandwidth (six effective observations on the control side), so it does not identify an effect.

Table 14: Exam taking and the role of failed-exam zeros ↩ p. 23

Outcome	(1) No controls	(2) With controls	Control mean
Panel A. Exam taking in the first exam period after grades post			
Exam sittings	−0.366**	−0.401**	1.31
	(0.171)	(0.175)	
Exams failed	−0.234**	−0.246**	0.48
	(0.111)	(0.110)	
Any zero recorded	−0.136**	−0.129**	0.29
	(0.053)	(0.053)	
Courses passed, next semester	0.279*	0.302*	3.41
	(0.168)	(0.173)	
Credits attempted, next semester	0.805	0.948	39.5
	(1.666)	(1.579)	
Credits obtained, next semester	2.513	2.650	33.2
	(1.672)	(1.697)	
Summary index, next semester	0.124*	0.140**	—
	(0.066)	(0.066)	
Panel B. Next-semester GPA with and without failed-exam zeros (fixed bandwidth)			
GPA next semester	0.465**	0.496**	6.39
	(0.222)	(0.228)	
GPA excluding failed-exam zeros	0.204	0.214	7.31
	(0.152)	(0.157)	
Exams failed, next semester	−0.205*	−0.206*	0.66
	(0.109)	(0.111)	
Panel C. Concentration at the low cutoffs (split-sample estimates)			
	Low cutoffs	High cutoffs	
Summary index, next semester	0.219***	−0.104	
	(0.081)	(0.101)	
GPA excluding failed-exam zeros, next semester	0.332**	−0.073	
	(0.158)	(0.225)	
Exam grade, first full exam cycle	1.068**	—	
	(0.431)		

Notes: Robust bias-corrected RD estimates of eq. (1) with cutoff fixed effects; standard errors clustered at the student level; column (2) adds individual controls. Panel A: the first exam period is July–August of the enrollment year, with exams for the student's other first-semester courses; MSE-optimal bandwidths per outcome (2,252–4,413 effective observations). The summary index averages control-side z-scores of (−) exams failed, (+) courses passed through the course track, and (−) the exam share of graded activities in the next semester, in control-side standard deviations. Courses passed count both routes, and credits are those of the courses attempted or passed. Over the one and two years after first-semester grades are released, courses passed (+0.14 and +0.25; control means 9.1 and 16.8) and credits obtained (+1.1 and +1.7; 89.8 and 165.8) do not differ either, with 95% confidence intervals excluding reductions larger than about a tenth of the mean (with controls; not shown). Panel B: the three rows share one sample (3,691 and 3,622 effective observations) and the GPA equation's MSE-optimal bandwidth ($h = 1.04$); the second row drops failed-exam zeros from the numerator and the denominator of next-semester GPA. Exam counts at one and two years do not differ (not shown). Panel C: low cutoffs are jumps into 7–9 and high cutoffs jumps into 10–12, each group with its own cutoff fixed effects and MSE-optimal bandwidth and no individual controls. The exam-grade row conditions on sitting at least one exam in the first full exam cycle (December–March) and is shown only for the low cutoffs, where the conditional sample is large enough. * $p < 0.10$, ** $p < 0.05$,

Table 15: Extensive margin: timing, duration, and incidence of first employment ↩ p. 18

	Months to first job	Length first job	Not emp. (m6)	Not emp. (m12)
Conventional	-0.545	0.438	-0.0182	0.0123
	(1.373)	(0.993)	(0.0385)	(0.0467)
Robust	-0.783	0.283	-0.0258	0.0236
	(1.676)	(1.226)	(0.0463)	(0.0583)
Control mean	24.786	13.900	0.268	0.346
Bandwidth	1.576	2.159	1.764	1.256
Obs.	7950	7950	7950	7950
Eff. obs.	2947	4227	3165	2562
Clusters	2049	2523	2162	1839

Notes: Labor sample (Table 3, Panel B). Months to first job measured from enrollment; spell length in months; non-employment indicators measured 6 and 12 months after the first-job month. Conventional and robust bias-corrected estimates; standard errors, clustered at the student level, in parentheses. Stars: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 16: Total real monthly earnings ↩ p. 19

	First job	Month 1	Month 6	Month 12
Conventional	-205.7	1417.4	-949.1	-411.6
	(819.8)	(979.2)	(1285.2)	(2037.5)
Robust	-420.0	1702.5	-1235.2	27.92
	(998.8)	(1165.8)	(1577.8)	(2264.2)
Control mean	13721.101	22828.518	27133.223	30556.570
Bandwidth	1.3056	1.3402	1.3142	0.9142
Obs.	7950	7300	5796	5135
Eff. obs.	2567	2367	1892	948
Clusters	1841	1846	1845	1222

Notes: Real monthly pesos, summed across simultaneous jobs; top 1% of first-job earnings trimmed. Horizon columns condition on employment in the stated month. Conventional and robust bias-corrected estimates; standard errors, clustered at the student level, in parentheses. Stars: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 17: Monthly earnings, hourly wages, and hours worked at the first job ↩ p. 20

	ln(earnings)	ln(earnings/hours)	ln(hours)	ln(constructed wage)
Conventional	0.0387	0.114*	-0.0748	0.111*
	(0.0771)	(0.0632)	(0.0861)	(0.0633)
Robust	0.0278	0.140*	-0.112	0.137*
	(0.0937)	(0.0765)	(0.104)	(0.0766)
Control mean	9.258	5.169	4.089	5.171
Bandwidth (fixed)	1.933	1.933	1.933	1.933
Obs.	7789	7789	7789	7789
Eff. obs.	3184	3184	3184	3184
Clusters	2156	2156	2156	2156

Notes: All columns use the same sample, the 7,789 student×course observations with all three outcomes defined, and the same fixed bandwidth ($h = b = 1.93$, the MSE-optimal choice for the earnings equation). The accounting identity therefore holds in the estimates, with column (1) equal to column (2) plus column (3) up to rounding in both the conventional and the robust rows. Column (2) is ln(earnings/hours), and column (4) is the log hourly wage variable constructed separately, reported as a consistency check. Cutoff fixed effects; triangular kernel; standard errors clustered at the student level. Conventional and robust bias-corrected estimates (Calonico et al., 2020). Stars: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 18: Log real hourly wages ↩ p. 20

	First job	Month 6	Month 12
Conventional	0.155**	0.0130	0.0169
	(0.0781)	(0.0627)	(0.0592)
Robust	0.187**	0.0308	0.0204
	(0.0935)	(0.0705)	(0.0675)
Control mean	5.144	5.374	5.489
Bandwidth	1.260	0.805	1.248
Obs.	6224	5037	5135
Eff. obs.	2046	863	1563
Clusters	1471	935	1401

Notes: Uniform kernel; first jobs beginning by December 2021 (Section 3.3). With a triangular kernel and no date restriction ($N = 7{,}950$), the estimates are 0.137 (conventional, s.e. 0.072) and 0.162 (robust, s.e. 0.085) at the first job. Horizon columns condition on employment in the stated month. Standard errors, clustered at the student level, in parentheses. Stars: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 19: Hours worked ↩ p. 20

	Hours worked	>20 hs worked	Hours worked (m6)	Hours worked (m12)
Conventional	-2.234**	-0.107**	-2.832***	-1.161
	(1.122)	(0.0492)	(1.061)	(1.148)
Robust	-2.764**	-0.133**	-3.306**	-0.992
	(1.327)	(0.0581)	(1.293)	(1.395)
Control mean	18.989	0.461	28.507	28.953
Bandwidth	1.339	1.266	1.094	1.162
Obs.	7887	7887	5771	5107
Eff. obs.	2554	2542	1686	1546
Clusters	1832	1825	1698	1732

Notes: Weekly hours = monthly hours ÷ 52/12 (Section 2.6); entry-month hours reflect partial months, so month-6 and month-12 columns are the cleaner intensive-margin measurements. This table trims both 1% tails of first-job earnings; under a top-trim-only specification the first-job estimates are −1.71 (conventional) and −2.17 (robust, both insignificant) while the month-6 estimates are −2.96/−3.46 (significant at 1%). Standard errors, clustered at the student level, in parentheses. Stars: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 20: Firm pay: firm-average hourly wages ↩ p. 22

	First job	Month 6	Month 12
Conventional	0.107	0.0382	0.140**
	(0.0651)	(0.0670)	(0.0677)
Robust	0.132*	0.0520	0.161**
	(0.0767)	(0.0795)	(0.0772)
Control mean	6.042	6.068	6.107
Bandwidth	1.431	1.412	1.453
Obs.	7950	5796	5135
Eff. obs.	2636	1935	1748
Clusters	1880	1879	1894

Notes: The outcome is the log average hourly wage paid by the employing firm, over all of the firm's workers. Conventional and robust bias-corrected estimates; standard errors, clustered at the student level, in parentheses. Stars: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 21: Public-sector employment ↩ p. 22

	First job	Month 6	Month 12
Conventional	0.0705**	0.0555*	0.0638**
	(0.0308)	(0.0287)	(0.0290)
Robust	0.0755**	0.0636*	0.0684*
	(0.0370)	(0.0341)	(0.0352)
Control mean	0.145	0.138	0.124
Bandwidth	1.7691	1.8251	1.7006
Obs.	7950	7950	7950
Eff. obs.	3165	3202	3034
Clusters	2162	2177	2090

Notes: Indicator for the reference employer being public-sector, at the first job and at 6- and 12-month horizons. Conventional and robust bias-corrected estimates; standard errors, clustered at the student level, in parentheses. Stars: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 22: Academic outcomes around the start of the first job ↩ p. 26

Outcome	Conventional	Robust	Control mean	Eff. obs.
GPA, pre-job to first post-job semester	0.520**	0.606**	−0.80	2,265
	(0.260)	(0.307)		
Activities enrolled, same change	0.373**	0.423**	−0.92	2,807
	(0.155)	(0.185)		
GPA, two semesters after job start	0.576**	0.678**	5.90	2,195
	(0.288)	(0.330)		

Notes: Students whose first formal job begins after the grade shock. The first two rows are within-student changes from the last semester before the job to the first full semester after it begins; the third row is the level two semesters after. RD estimates as in eq. (1), with cutoff fixed effects and MSE-optimal bandwidths; standard errors, clustered at the student level, in parentheses. Stars: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix

A Alternative specification: aggregating grade shocks into an instrument for GPA

This paper estimates the effect of marginally earning a higher grade in a single first-semester course by implementing an RDD. Given that scholarship rules, graduate admissions, and hiring formulas use GPA as the summary measure of a student's grades, a complementary question is how a change in GPA itself affects our outcomes of interest. This appendix outlines the Regression Discontinuity Aggregation (RDA) framework of Borusyak et al. (2026) and shows that it reproduces the conclusions of Section 5 in GPA units. The idea is to use as exogenous variation of GPA only that which comes from courses in which a student marginally earns a higher grade in an Instrumental Variables setting. We treat this as an alternative specification rather than the main research design.

A.1 Specification

For student i taking first-semester courses J_i, let $r_{ij} = s_{ij} - \text{cutoff}_{ij}$ and $z_{ij} = \mathbf{1}(r_{ij} \geq 0)$, and assign equal importance weights $\omega_{ij} = 1/|J_i|$. The aggregate instrument is

$$Z_i = \sum_{j \in C_i} \omega_{ij} z_{ij}, \qquad C_i = \{ j \in J_i : |r_{ij}| \leq h \}, \tag{2}$$

the (weighted) share of student i's first-semester courses in which they landed just above a grade threshold, among courses within bandwidth h of a cutoff. We instrument first-semester GPA with Z_i in

$$Y_i = \beta \, \text{GPA}_i + \gamma_0 \sum_{j \in C_i} \omega_{ij} + \gamma_1 \sum_{j \in C_i} \omega_{ij} r_{ij} + \gamma_2 \sum_{j \in C_i} \omega_{ij} r_{ij}^{+} + \tilde{\gamma}' \tilde{W}_i + \varepsilon_i, \tag{3}$$

where the three RDA control terms play the role of the standard local-linear controls, absorbing, respectively, the total weight of courses near a cutoff, their average running variable, and its interaction with the above-cutoff indicator. We set $h = 1$ at baseline, in line with the optimal bandwidths in Section 3.1, and construct all RDA variables from first-semester performance only. The estimation sample conditions on a first job beginning more than six months after enrollment and at least one scored first-semester course (Table 3, Panel C). Unlike with the course-level design, each student enters the RDA estimation exactly once, so inference here is done at the student level since there is no

within-student duplication to account for. We report heteroskedasticity-robust standard errors throughout, with first-stage strength reported per column as the Kleibergen–Paap F statistic.

A.2 Instrument validity: first stage and exogeneity

Table A1 reports the first stage. The instrument is strong: a one-unit increase in Z raises recorded first-semester GPA by 2.98 points (3.34 on the IV estimation sample), with a Kleibergen–Paap F of 36.7. In addition to a strong first stage, identification requires that, conditional on the RDA controls, Z_i is as good as random. Table A2 regresses Z_i on baseline demographics: without RDA controls the covariates are jointly significant ($F = 9.48$, $p < 0.001$), reflecting the mechanical correlation between the density of scores near cutoffs and student characteristics; including the RDA controls, joint significance disappears ($F = 1.18$, $p = 0.256$).

A.3 Results

Table A3 reports the IV estimates on the $N = 3{,}824$ students with both first-semester records and an observed post-enrollment first job. A one-point increase in first-semester GPA raises log real hourly wages at the first job by 0.229 log points (s.e. 0.083), raises the log average hourly wage paid by the initial firm by 0.129 (s.e. 0.066), reduces monthly hours worked by 8.7 (s.e. 5.2, significant at 10%), and increases the probability of public-sector employment by 8.7 percentage points (s.e. 3.5), which is the same qualitative pattern as the course-level estimates of Section 5. Table A4 includes the graduation outcomes of Section 4 under the same IV specification (per-horizon cohort windows as in Table 9). All four estimates are statistically indistinguishable from zero, corroborating the course-level conclusion that graduation responds far less than performance, even though first-stage strength is lower on these samples (F between 10.7 and 16.1), making the IV nulls less informative than the reduced-form ones from the RDD.

B A rounding-aware test for manipulation of the score distribution

Standard density tests assume a continuously distributed running variable. Our scores are discrete and heaped, with half of them being integers, 14% ending in half points, and the rest carrying finer fractions from partial-credit and scanner-penalty conventions

(Table A5). A continuous-density test may therefore reject smoothness because of this granularity alone (Section 3.2). Andarge et al. (2025) develop a manipulation test built for a similar environment to ours: running variables generated by continuous underlying performance and then subjected to deterministic rounding conventions, where the question is not whether the density is smooth but whether scores are rounded *up into* policy-relevant intervals more often than the setting's ordinary rounding behavior predicts (in our setting, this would mean comparing how much "excess rounding" there is at integers that grant a jump in the coarse grade, versus other nearby integers that do not grant this reward, and are therefore not manipulated but simply rounded).

The test works as follows. For each policy threshold c, define the roundable interval $[c - 0.5, c + 0.5)$, which is the set of scores that ordinary rounding to the nearest integer could produce from underlying values near c. The counterfactual frequency of landing in such an interval is estimated from the rounding behavior observed at non-policy integers (integers that trigger no grade change), with the underlying continuous distribution fitted by censored maximum likelihood, namely a generalized beta of the second kind (GB2), censored on manipulation windows $(c - 3, c + 1)$ around every policy threshold so that the fit is never contaminated by the very intervals under test. The estimand η_c is the excess rounding at c, measured as the gap between the observed and the counterfactual probability of landing in the roundable interval. It is point-identified when rounding is homogeneous across integers, and bounded otherwise. For inference we use a nonparametric bootstrap with 200 replications, clustered at the student level.

One difficulty in implementing the test is that scores below 50 do not enter our sample, so the support of the scores is effectively $[50, 100]$, and the GB2 family cannot represent a distribution truncated at an interior point. We therefore fit transformed scores, under two separate transformations: a reflection $(101 - s)$ and a shift $(s - 50)$. Each transformation implies a different parametric family, so comparing the two shows whether the estimates depend on this parametric choice. Even though the passing threshold at 50 is excluded from the main RDD, its visible bunching makes it a relevant control for this test. A manipulation test that missed the bunching there would give us a reason to doubt it at the other cutoffs.

The results from this exercise are reported in Table A6. At the pass/fail threshold, the test detects excess rounding of $+0.93$ percentage points ($t \approx 9.5$), so that the method finds the manipulation we know exists. At the six grade-jump cutoffs used by the RD design, estimated excess rounding ranges from -0.15 to $+0.08$ percentage points and is never statistically distinguishable from zero. Figure A1 displays the fitted and observed distributions under both transformations. We conclude that the rounding behavior at the

grade-jump cutoffs is indistinguishable from the rounding behavior at the other integers that do not grant a higher grade. The one threshold with a visible manipulation in the density of the scores fails the same test by an order of magnitude.

C Industry of employment and index decompositions

C.1 Two industry classifications

Industry of employment is measured two independent ways, and every result in this appendix is reported under both, with results being qualitatively identical across the two. The first classification uses the firm's bargaining group under the wage bargaining system, Uruguay's structure of industry-level collective bargaining, which BPS records carry as a firm-level attribute. We aggregate the 24 groups into six private cells with usable sample sizes — professional and technical services, commerce, finance, education, other services, and industry and primary activities — with public-sector jobs, which sit outside the wage bargaining system, as their own cell. This classification is observed directly for 61% of first jobs and is imputed for most of the remainder from the firm's modal classification elsewhere in the panel (99.2% agreement with observed values where both exist) or from the firm's activity code (90.1% agreement); 1.8% of first jobs remain unclassifiable, and failure to be classified by this first method does not exhibit a discontinuity at the cutoff. The second classification requires no imputation: every employer record carries a five-digit economic-activity code following ISIC rev. 4, which we aggregate into parallel cells (sections M–N professional and administrative services; G commerce; K finance; P education; H, I, J, Q–T other services; the remainder industry and other), with public sector again separate.

C.2 The index decomposition

For an outcome y, write each worker's value as an identity,

$$y_i \;=\; \bar{y}_{s(i)} \;+\; \big(y_i - \bar{y}_{s(i)}\big), \tag{4}$$

where $\bar{y}_s$ is the mean of y among *control-side* workers in industry cell s inside the outcome's MSE-optimal bandwidth. The first term, which we call the industry index, is the value a worker would record if their wage or hours were exactly at the average of their industry. Since the local linear estimator is linear in the outcome, estimating the design separately on the index and on the residual, with the same bandwidth, bias bandwidth,

and sample, splits the total effect exactly into two parts. The *between* component is the effect that operates through the industry of the first job, and the *within* component is the effect that operates inside industries. Equivalently, the between component is equal to the sum over cells of the effect on cell membership multiplied by the demeaned norm of the cell. This decomposition is based on Oaxaca (1973) and on the sorting decompositions used in the firm-effects literature (Card et al., 2013, 2016), and we apply it at the industry level. Table A8 reports the resulting decompositions under both classifications. The first column presents the main results as estimated throughout, while the remaining columns report the decomposition, estimating the main equation with the between and within components as outcomes, each under the two industry classifications.

C.3 Inference and reading

A cluster bootstrap by student over the entire two-step procedure (re-drawing students, recomputing the cell norms, and re-estimating the components) addresses the fact that the index is itself being estimated in a first step. Bootstrap standard errors match the conventional ones for every component except the month-6 between terms, whose cell norms come from the thinner sample employed at month six. We also run a placebo-threshold randomization inference, re-estimating both components at twenty one-sided fake cutoffs and rebuilding the index each time from the control side of the fake threshold. Among these placebos, the actual within components for the wage and for hours rank first or second (empirical p of 0.048 to 0.095), while the between components fall in the middle of their placebo distributions.

At entry, the between-industry component is a precisely estimated zero for the wage and hours margin under both classifications, and the earnings null itself decomposes into two nulls. The effect of a higher grade therefore seems to come from within-industry differences in labor-market outcomes. However, this claim requires some caveats. With roughly eight industry cells, reallocation toward better sub-jobs within an industry is estimated as a within effect.

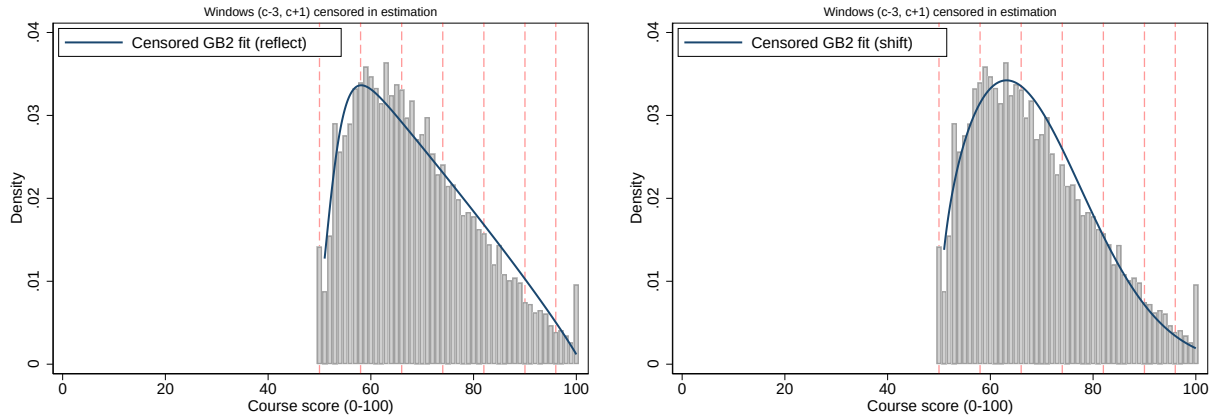
D Degree curricula

Tables A9 to A13 reproduce the curricula of the five undergraduate degrees offered by FCEA during our sample period, as published by FCEA: the mandatory courses of each degree by suggested semester, with their credit values and English translations. The plans shown are those in force for the cohorts in our sample, namely Plan 2012 for the

accountancy, economics, and administration degrees and Plan 2014 for the statistics and administration-technician degrees; elective requirements and minimum credits per subject area are summarized in the notes to each table. With the exception of the statistics degree, the first semester of every curriculum consists of the same four mandatory courses on which the design is built, with Cálculo I offered both as a single course and as two sequential half-courses.

Appendix Figures and Tables

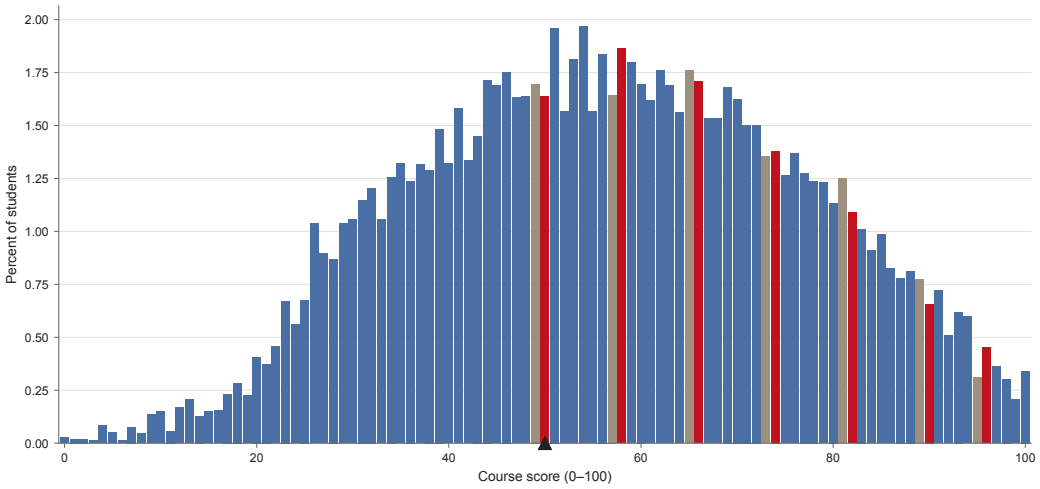
Figure A1: Andarge et al. test: fitted censored GB2 vs. observed score distribution ↩ p. 52



(a) Reflection transform

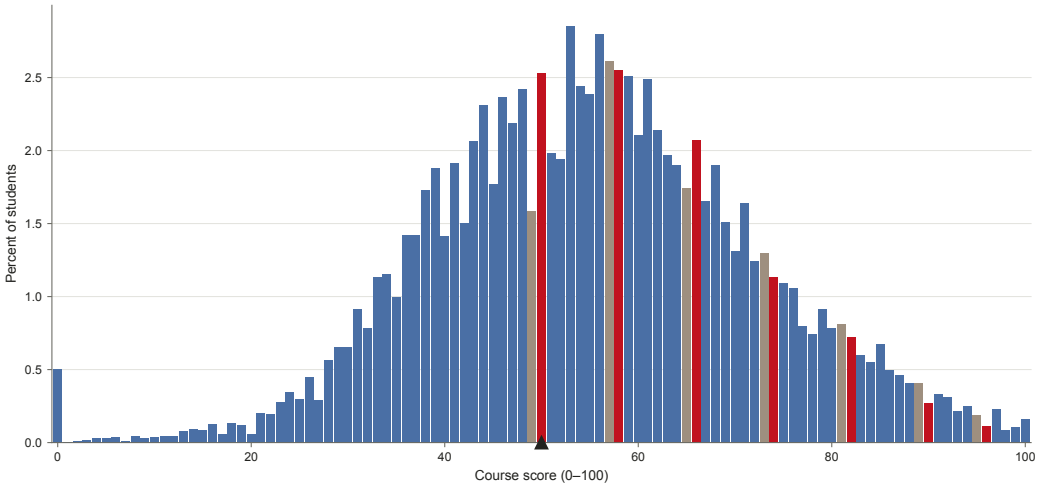
(b) Shift transform

Figure A2: Distribution of course scores: C10 ↩ p. 57



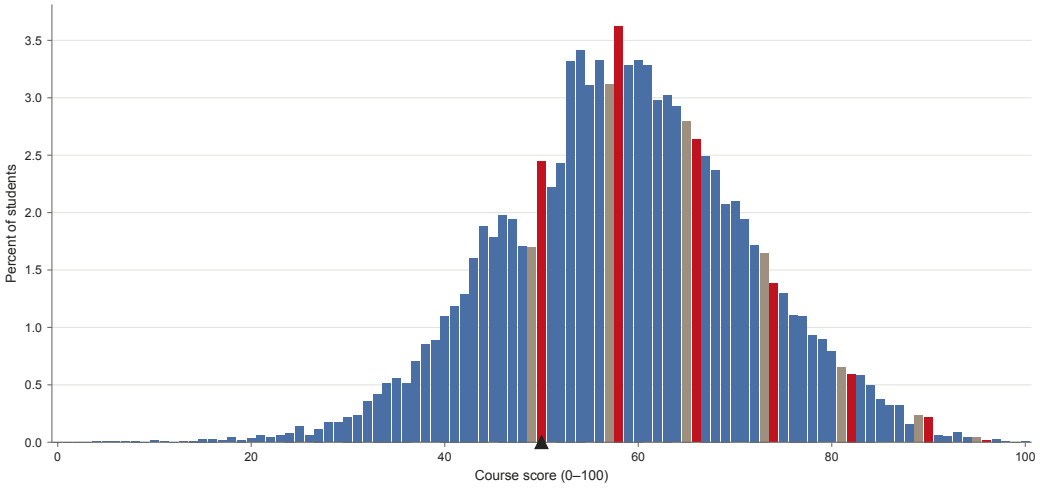
Notes: Sample and conventions as in Figure 1: raw 0–100 course scores, students who sat both midterm examinations ($n = 15{,}510$), one-point bins centred on integers. Bins sitting on an institutional threshold are shown in red, with the bin immediately below each shown in grey; the caret marks the pass mark at 50. Some of the courses in Figures A2 to A6 record whole numbers only, while others also record fractional scores from partial credit and scanner penalties. This heaping is the reason we use the rounding-aware test of this appendix.

Figure A3: Distribution of course scores: E10



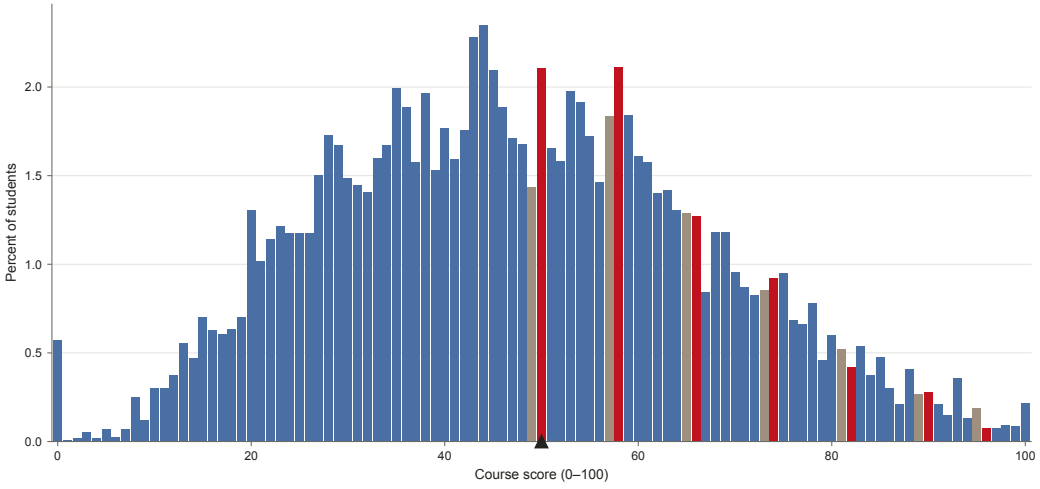
Notes: Sample and conventions as in Figure 1 ($n = 16{,}794$).

Figure A4: Distribution of course scores: A10



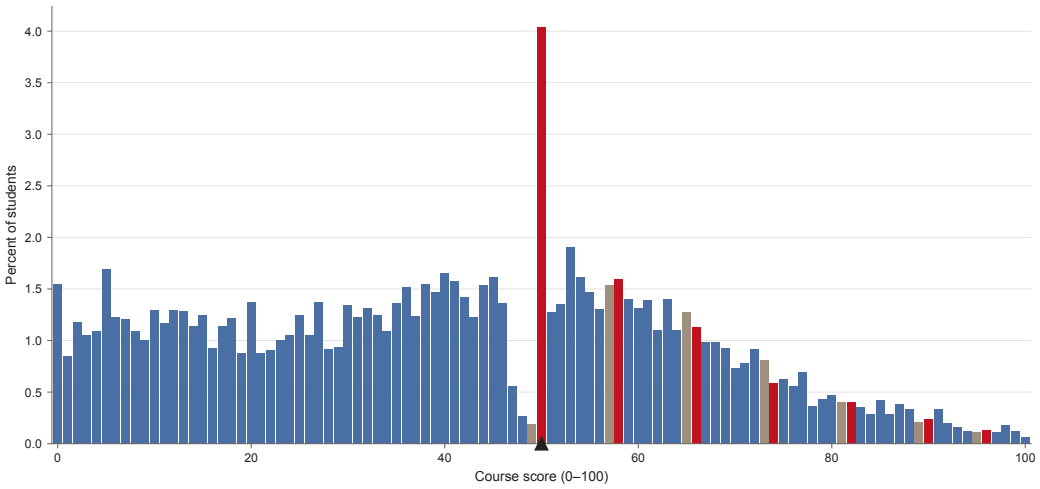
Notes: Sample and conventions as in Figure 1 ($n = 18{,}804$).

Figure A5: Distribution of course scores: 114A



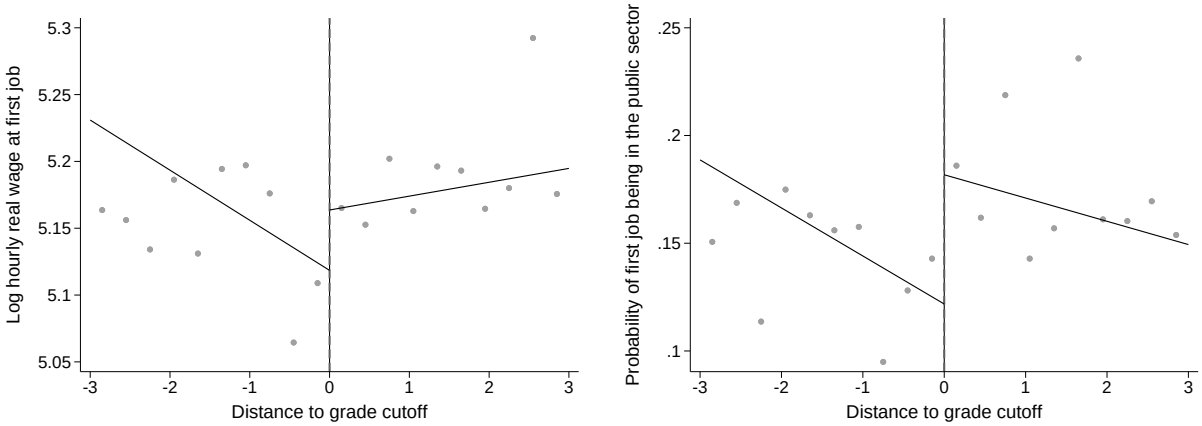
Notes: Sample and conventions as in Figure 1 ($n = 11{,}492$).

Figure A6: Distribution of course scores: MC10 ↩ p. 57



Notes: Sample and conventions as in Figure 1 ($n = 8{,}964$).

Figure A7: Seeing the labor-market discontinuities

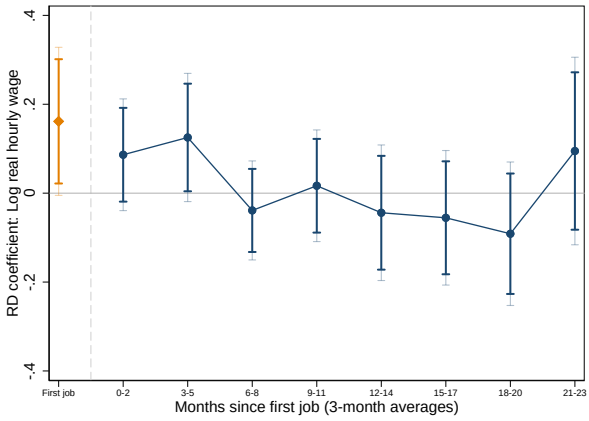


(a) Log real hourly wage at first job

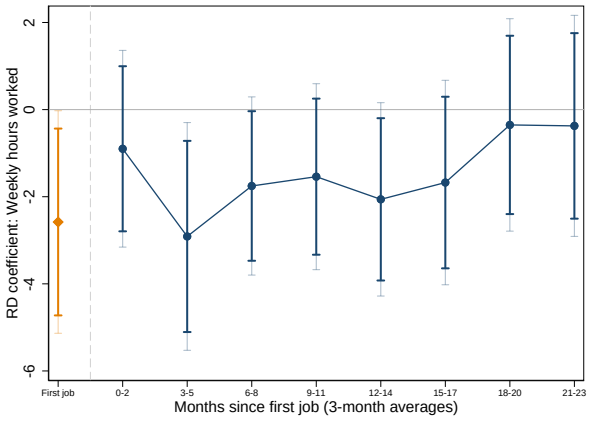
(b) Public-sector first job

Notes: Binned means of the raw outcome against distance to the nearest cutoff over a fixed ± 3 window, with linear fits on each side, which is a visual diagnostic rather than the estimator: the estimates in Tables 18 and 21 use MSE-optimal bandwidths ($\approx$1.2–1.8), cutoff fixed effects, and kernel weights, so fitted jumps here need not match the point estimates.

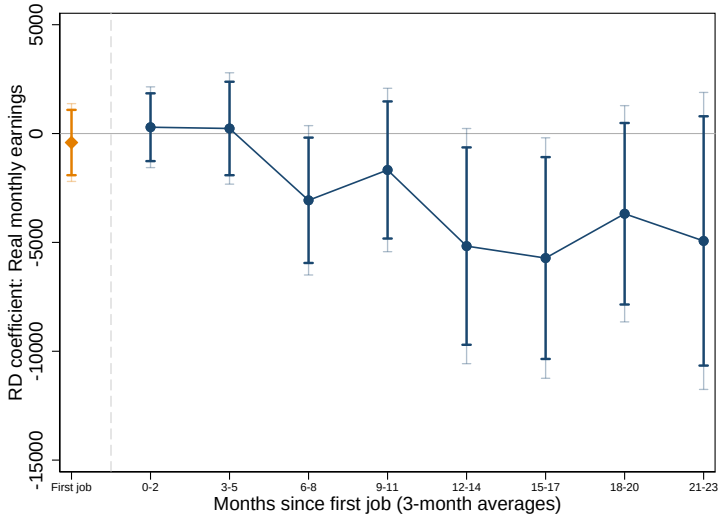
Figure A8: Dynamics with per-horizon-optimal bandwidths ↩ p. 21



(a) Log real hourly wage



(b) Weekly hours



(c) Real monthly earnings

Notes: As Figure 7, but with the MSE-optimal bandwidth re-selected at each horizon. The profiles have the same shape as in Figure 7. Individual windows move with the bandwidth, so the main text holds it fixed across horizons.

Table A1: RDA first stage and the mechanical benchmark ↩ p. 51

	(1) First-sem. GPA	(2) Share above cutoff (X)	(3) GPA (IV sample)	(4) X (IV sample)
Z	2.9762***	1.0365***	3.3415***	1.0882***
	(0.5049)	(0.0578)	(0.5518)	(0.0634)
Observations	5144	5144	3824	3824
R2	0.080	0.235	0.073	0.244

Notes: OLS of first-semester GPA (columns 1, 3) and of the mechanical scored-grade average X (columns 2, 4) on the instrument Z and the RDA controls C_1–C_3. Columns 3–4 restrict to the IV estimation sample. Standard errors in parentheses. Stars: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A2: Instrument balance ↩ p. 51

	(1) No RDA controls	(2) With RDA controls
Woman	0.006	0.003*
	(0.004)	(0.001)
Mother's education	0.002	-0.000
	(0.001)	(0.000)
Father's education	0.001	0.000
	(0.001)	(0.000)
White	-0.002	0.003
	(0.007)	(0.002)
Montevideo	-0.006	-0.000
	(0.005)	(0.002)
Public high school	0.002	-0.001
	(0.006)	(0.002)
Age	-0.001	0.002
	(0.004)	(0.001)
F-stat (joint sig.)	9.43	1.18
p-value (joint sig.)	0.000	0.253
Observations	4725	4725

Notes: Regressions of Z_i on baseline covariates, without (column 1) and with (column 2) the RDA control terms. Joint-significance F and p reported below the coefficients. Standard errors in parentheses. Stars: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A3: RDA IV: effects of first-semester GPA on first-job outcomes ↩ p. 51

	(1) log hourly wage	(2) log firm avg. wage	(3) monthly hours	(4) public sector
First-semester GPA	0.2293***	0.1290**	-8.6929*	0.0866**
	(0.0827)	(0.0657)	(5.1785)	(0.0354)
First-stage F (K-P)	36.67	36.67	36.67	36.67
Observations	3824	3824	3824	3824
RDA Controls	Yes	Yes	Yes	Yes
Ind. Controls	No	No	No	No

Notes: 2SLS estimates of eq. (3): first-semester GPA instrumented by Z. Hours in monthly units (-8.7 monthly ≈ -2.0 weekly). Robust standard errors in parentheses. Stars: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: RDA IV: graduation outcomes ↩ p. 51

	Grad. in 5y	Grad. in 6y	Ever graduates	Years to degree
First-semester GPA	0.0122	0.0671	0.0642	-0.0823
	(0.0669)	(0.0574)	(0.0579)	(0.193)
First-stage F (K-P)	10.72	13.31	13.31	16.10
Observations	6905	5787	5787	2029

Notes: 2SLS estimates of eq. (3) on the graduation outcomes of Table 9, on the student-level sample without labor conditioning; cohort windows per column as in Table 9. Years to degree is measured only for graduates and is descriptive. Robust standard errors in parentheses, with the Kleibergen–Paap F of each column below the estimates. Stars: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Fractional endings of the course score ↩ p. 52

Type of score	Share (%)
Integer (.000)	50.0
One half (.500)	13.8
One quarter (.250)	7.2
Three quarters (.750)	7.0
Eighths (.125, .375, .625, .875)	6.3
Finer fractions	15.7

Notes: Distribution of the non-integer part of the 0–100 course score (N = 13,626 student-by-course observations). Fractional endings come from partial-credit and scanner-penalty grading conventions.

Table A6: Rounding-aware manipulation test (Andarge et al., 2025): excess rounding at policy thresholds ↩ p. 52

	Reflection ($101 - s$)		Shift ($s - 50$)	
Threshold	$\hat{\eta} \times 100$	Interval test t	$\hat{\eta} \times 100$	Interval test t
50 (pass/fail; positive control)	0.93[a]	9.47	1.12[b]	30.21
	(0.09)		(0.02)	
58	−0.13	0.25	−0.02	1.63
	(0.10)		(0.11)	
66	0.06	2.49	−0.15	−0.30
	(0.11)		(0.10)	
74	0.00	0.71	−0.14	−1.45
	(0.10)		(0.11)	
82	−0.12	−0.82	−0.04	0.42
	(0.07)		(0.08)	
90	−0.09	−3.66	0.08	0.39
	(0.06)		(0.06)	
96	−0.03	−2.25	0.08	0.84
	(0.05)		(0.05)	

Notes: $\hat{\eta}$ is excess rounding into the roundable interval $[c - 0.5, c + 0.5)$, in percentage points of all observations, point-identified under homogeneous rounding; bootstrap standard errors in parentheses (200 replications, clustered at the student level). "Interval test t" is the t-statistic on the raw observed-minus-predicted interval frequency. Academic sample, $N = 13{,}626$. The reflection-based interval-test rejections at 66, 90, and 96 vanish under the shift transform, indicating parametric-fit artifacts; the excess-rounding estimates agree across transforms at every grade-jump cutoff.
[a] Reflection estimate quoted as the headline for the 50 threshold.
[b] Under the shift transform the 50 threshold sits at the support boundary and its bootstrap SE is artificially small.

Table A7: Discontinuities on labor market outcomes: low versus high cutoffs ↩ p. 24

	Low cutoffs		High cutoffs	
Outcome	RD effect	95% robust CI	RD effect	95% robust CI
Log hourly wage	0.177	[−0.55, 0.17]	0.081	[−0.79, 0.16]
Weekly hours	−2.46	[−6.80, 2.58]	−2.02	[−7.05, 4.09]
Works more than 20 hours	−0.098	[−0.35, 0.04]	−0.103	[−0.41, 0.15]
Monthly earnings	−1,076	[−5,818, 2,581]	−292	[−8,044, 5,116]
Public sector	0.076	[−0.03, 0.17]	0.042	[−0.11, 0.27]
Log firm-average wage	0.113	[−0.08, 0.40]	0.094	[−0.33, 0.30]

Notes: First-job outcomes estimated separately at the low (jumps into 7–9) and high (jumps into 10–12) cutoffs with the heterogeneous-effects estimator of Table 12 (Calonico et al., 2025): covariate-adjusted, uniform kernel, group-specific MSE-optimal bandwidths, standard errors clustered at the student level. Confidence intervals are robust bias-corrected and therefore not centered on the point estimates by construction. No low/high difference is statistically significant. Effective observations range from 1,132 to 1,972 at the low cutoffs and from 226 to 665 at the high cutoffs.

Table A8: Index decompositions of the first-job effects, by industry classification ↩ p. 25

	Total	Between		Within	
Outcome		Council	ISIC	Council	ISIC
Log hourly wage	+0.137	−0.010	+0.019	+0.147	+0.117
	(0.072)	(0.020)	(0.014)	(0.071)	(0.071)
Weekly hours, entry	−2.234	+0.151	−0.193	−2.385	−2.041
	(1.122)	(0.258)	(0.259)	(1.079)	(1.066)
Weekly hours, month 6	−2.832	−0.187	−0.555	−2.645	−2.277
	(1.061)	(0.179)	(0.161)	(1.033)	(1.033)
P(>20 hours/week)	−0.107	+0.012	−0.008	−0.119	−0.099
	(0.049)	(0.013)	(0.013)	(0.046)	(0.046)
Log monthly earnings	+0.045	+0.029	+0.019	+0.015	+0.026
	(0.076)	(0.017)	(0.016)	(0.074)	(0.073)

Notes: Conventional local linear estimates with cutoff fixed effects, clustered by student; each outcome at its own MSE-optimal bandwidth, held fixed with the bias bandwidth across its components so that between and within sum to the total by construction. Outcomes are trimmed as in their main tables, with hours outcomes trimming both 1% tails of first-job earnings and wage and earnings outcomes trimming only the top tail. For month-6 outcomes, the industry is that of the entry job, so the between component is the effect operating through that industry. Standard errors are conditional on the estimated index. A cluster bootstrap by student over the entire two-step procedure ($B = 500$) gives the same conclusions, with one exception noted in the text.

Table A9: Curriculum of the Public Accountant degree (Contador Público), Plan 2012
↩ p. 6

Code	Course	English translation	Credits
First semester			
C10	Conceptos Contables	Accounting Concepts	10
A10	Administración y Gestión de las Organizaciones I	Organizational Administration and Management I	10
E10	Introducción a la Microeconomía[b]	Introduction to Microeconomics	10
MC10	Cálculo I[a]	Calculus I	10
Second semester			
C20	Contabilidad General I	General Accounting I	10
J20TP	Derecho Civil	Civil Law	10
E20N	Economía Descriptiva	Descriptive Economics	10
Third semester			
C30	Contabilidad General II	General Accounting II	10
J30	Derecho Comercial	Commercial Law	10
A30	Procesos y Sistemas de Información	Processes and Information Systems	10
MC40	Introducción a la Estadística	Introduction to Statistics	10
Fourth semester			
C40	Contabilidad General III	General Accounting III	10
C41	Legislación Laboral y Seguridad Social	Labor Law and Social Security	10
J40TP	Derecho Societario	Corporate Law	10
MC30	Matemática Financiera	Financial Mathematics	10
Fifth semester			
C50N	Modelos y Sistemas de Costos	Cost Models and Systems	10
J50	Derecho Tributario	Tax Law	10
I52	Ética y Ejercicio Profesional	Ethics and Professional Practice	5
Sixth semester			
C60	Contabilidad Superior I	Advanced Accounting I	10
C61	Control Interno	Internal Control	10
C64	Contabilidad Gerencial	Managerial Accounting	10
C72	Tributaria I	Taxation I	10
Seventh semester			
C70	Auditoría I	Auditing I	10
C82	Tributaria II	Taxation II	10
A71	Finanzas Corporativas	Corporate Finance	10
Eighth semester			
C81	Informes Profesionales de Contador Público	Professional Reports of the Public Accountant	10
I80	Práctica Pericial en el Ámbito Judicial	Expert Accounting Practice in Judicial Proceedings	5
A67	Gestión Financiera del Estado	Public Financial Management	10

Notes: Mandatory courses of the Contador Público curriculum under Plan 2012, the plan in force for the cohorts in our sample, by suggested semester. The degree requires 360 credits in total, of which the mandatory courses above account for 270, and the remainder comes from electives, subject to minimum credits per area (Accounting and Taxes 130, Law 40, Administration 50, Quantitative Methods 40, Economics 30, Integrative Activities 20, Social Sciences and Humanities 10). The course codes are the ones used by the Faculty, and they are the same as in the administrative records. First-semester courses are offered in both semesters of the calendar year ("doble dictado"). Source: Facultad de Ciencias Económicas y de Administración, Universidad de la República.
[a] Cálculo I may alternatively be taken as two sequential half-courses of 5 credits each, Cálculo I/A (114A) in the first semester and Cálculo I/B (128A) in the second.
[b] Offered as Introducción a la Economía through 2017 and renamed Introducción a la Microeconomía in 2018, with the course code unchanged.

Table A10: Curriculum of the Bachelor in Economics (Licenciatura en Economía), Plan 2012

Code	Course	English translation	Credits
First semester			
E10	Introducción a la Microeconomía[b]	Introduction to Microeconomics	10
MC10	Cálculo I[a]	Calculus I	10
C10	Conceptos Contables	Accounting Concepts	10
A10	Administración y Gestión de las Organizaciones I	Organizational Administration and Management I	10
Second semester			
E34N	Introducción a la Macroeconomía[c]	Introduction to Macroeconomics	10
MC20	Álgebra Lineal	Linear Algebra	10
Third semester			
E30	Macroeconomía I	Macroeconomics I	10
MC31	Cálculo II	Calculus II	10
S40	Historia Económica Mundial	World Economic History	10
Fourth semester			
E31	Microeconomía I	Microeconomics I	10
E40	Macroeconomía II	Macroeconomics II	10
MC42	Estadística I	Statistics I	10
Fifth semester			
E41	Microeconomía II	Microeconomics II	10
E50	Macroeconomía III	Macroeconomics III	10
E51	Economía de América Latina	Economics of Latin America	10
MC51	Estadística II	Statistics II	10
I51E	Introducción a la Metodología	Introduction to Research Methodology	10
Sixth semester			
E60	Economía del Uruguay	Economy of Uruguay	10
E61	Microeconomía III	Microeconomics III	10
MC61	Econometría I	Econometrics I	10
Seventh semester			
E70	Economía Internacional	International Economics	10
E71	Teorías del Desarrollo Económico	Theories of Economic Development	10
MC70	Econometría II	Econometrics II	10
S82	Historia del Pensamiento Económico	History of Economic Thought	10

Notes: Mandatory courses of the Licenciatura en Economía curriculum under Plan 2012, by suggested semester. The eighth semester has no mandatory courses. The degree requires 360 credits in total, of which the mandatory courses above account for 240, and the remainder comes from electives, subject to minimum credits per area (Economics 150, Quantitative Methods 80, Social Sciences and Humanities 50, Integrative Activities 10, Accounting 10, Administration 10). First-semester courses are offered in both semesters of the calendar year ("doble dictado"). Source: Facultad de Ciencias Económicas y de Administración, Universidad de la República.

[a] Cálculo I may alternatively be taken as two sequential half-courses of 5 credits each, Cálculo I/A (114A) in the first semester and Cálculo I/B (128A) in the second.

[b] Offered as Introducción a la Economía through 2017 and renamed Introducción a la Microeconomía in 2018, with the course code unchanged.

[c] First offered in 2021, initially under code E34.

Table A11: Curriculum of the Bachelor in Management (Licenciatura en Administración), Plan 2012

Code	Course	English translation	Credits
First semester			
A10	Administración y Gestión de las Organizaciones I	Organizational Administration and Management I	10
C10	Conceptos Contables	Accounting Concepts	10
E10	Introducción a la Microeconomía[b]	Introduction to Microeconomics	10
MC10	Cálculo I[a]	Calculus I	10
Second semester			
A20	Administración y Gestión de las Organizaciones II	Organizational Administration and Management II	10
C20	Contabilidad General I	General Accounting I	10
E20N	Economía Descriptiva	Descriptive Economics	10
Third semester			
A30	Procesos y Sistemas de Información	Processes and Information Systems	10
E30	Macroeconomía I	Macroeconomics I	10
MC40	Introducción a la Estadística	Introduction to Statistics	10
J21	Derecho del Empresario	Business Law	10
Fourth semester			
A40	Administración de Recursos Humanos	Human Resource Management	10
MC30	Matemática Financiera	Financial Mathematics	10
Fifth semester			
A61	Marketing Básico	Basic Marketing	10
C31	Costos para la Gestión	Costs for Management	10
E53	Economía de las Decisiones Empresariales	Economics of Business Decisions	10
I50	Taller de Ética y Responsabilidad Social	Workshop on Ethics and Social Responsibility	5
Sixth semester			
C64	Contabilidad Gerencial	Managerial Accounting	10
S60	Administración de las Relaciones de Trabajo	Labor Relations Management	10
Seventh semester			
A70	Comportamiento Organizacional	Organizational Behavior	10
A71	Finanzas Corporativas	Corporate Finance	10
Eighth semester			
A80	Dirección Estratégica de Empresas	Strategic Management of Firms	10
A81	Proyectos de Inversión	Investment Projects	10
A83N	Cambio Organizacional	Organizational Change	10

Notes: Mandatory courses of the Licenciatura en Administración curriculum under Plan 2012, by suggested semester. The degree requires 360 credits in total, of which the mandatory courses above account for 235, and the remainder comes from electives, subject to minimum credits per area (Administration 120, Accounting and Taxes 60, Quantitative Methods 45, Economics 40, Integrative Activities 20, Social Sciences and Humanities 20, Law 15). First-semester courses are offered in both semesters of the calendar year ("doble dictado"). Source: Facultad de Ciencias Económicas y de Administración, Universidad de la República.

[a] Cálculo I may alternatively be taken as two sequential half-courses of 5 credits each, Cálculo I/A (114A) in the first semester and Cálculo I/B (128A) in the second.

[b] Offered as Introducción a la Economía through 2017 and renamed Introducción a la Microeconomía in 2018, with the course code unchanged.

Table A12: Curriculum of the Bachelor in Statistics (Licenciatura en Estadística), Plan 2014

Code	Course	English translation	Credits
First semester			
MC10	Cálculo I[a]	Calculus I	10
118E	Estadística Descriptiva	Descriptive Statistics	10
MC20	Álgebra Lineal	Linear Algebra	10
Second semester			
MC31	Cálculo II	Calculus II	10
1411	Computación I[b]	Computing I	10
E20N	Economía Descriptiva	Descriptive Economics	10
Third semester			
230	Probabilidad I	Probability I	15
Fourth semester			
MC43	Cálculo III	Calculus III	10
330	Inferencia I	Statistical Inference I	15
Fifth semester			
335	Muestreo y Planificación de Encuestas I	Sampling and Survey Design I	10
333	Modelos Lineales	Linear Models	10
Sixth semester			
334	Inferencia II	Statistical Inference II	10
MC61	Econometría I	Econometrics I	10
Seventh semester			
MC71	Análisis Multivariado I	Multivariate Analysis I	10
462	Series Cronológicas I	Time Series I	10

Notes: Mandatory courses of the Licenciatura en Estadística curriculum under Plan 2014, by suggested semester. The eighth semester has no mandatory courses, and the undergraduate thesis (Trabajo Final de Grado, 30 credits) is one of its electives. The degree requires 360 credits in total, of which the mandatory courses above account for 160, and the remainder comes from electives, subject to minimum credits per area (Mathematics 85, Statistical Inference 85, Graduate-Profile courses 60, Instrumental 30, Integrative Activities 20, General Knowledge 10). Several electives are cross-listed at other schools of the Universidad de la República (Engineering, Sciences, Social Sciences). Source: Facultad de Ciencias Económicas y de Administración, Universidad de la República.

[a] Statistics students may instead take Cálculo I/Est (MC11, 13 credits), and Cálculo I is offered in both semesters of the calendar year.

[b] Taught at the School of Engineering, and can be replaced by Programación Imperativa (1375, School of Engineering) or Computación (MA069, School of Sciences, 12 credits).

Table A13: Curriculum of the Associate Degree in Administration (Tecnicatura en Administración), Plan 2014 ↩ p. 6

Code	Course	English translation	Credits
First semester			
A10	Administración y Gestión de las Organizaciones I	Organizational Administration and Management I	10
C10	Conceptos Contables	Accounting Concepts	10
E10	Introducción a la Microeconomía[b]	Introduction to Microeconomics	10
MC10	Cálculo I[a]	Calculus I	10
Second semester			
A20	Administración y Gestión de las Organizaciones II	Organizational Administration and Management II	10
C20	Contabilidad General I	General Accounting I	10
J20TP	Derecho Civil	Civil Law	10
Third semester			
A30	Procesos y Sistemas de Información	Processes and Information Systems	10
C30	Contabilidad General II	General Accounting II	10
J30	Derecho Comercial	Commercial Law	10
MC40	Introducción a la Estadística	Introduction to Statistics	10
Fourth semester			
A40	Administración de Recursos Humanos	Human Resource Management	10
A42	Diseño Organizacional	Organizational Design	10
C40	Contabilidad General III	General Accounting III	10
C41	Legislación Laboral y Seguridad Social	Labor Law and Social Security	10
MC30	Matemática Financiera	Financial Mathematics	10
Fifth semester			
A70	Comportamiento Organizacional	Organizational Behavior	10

Notes: Mandatory courses of the Tecnicatura en Administración curriculum under Plan 2014, by suggested semester. This is a short-cycle degree of 225 credits over five suggested semesters, of which the mandatory courses above account for 170, and the remainder comes from electives, subject to minimum credits per area (Administration 60, Accounting and Taxes 50, Quantitative Methods 30, Law 20, Integrative Activities 15, Economics 10) and 30 credits of free distribution. First-semester courses are offered in both semesters of the calendar year ("doble dictado"). Source: Facultad de Ciencias Económicas y de Administración, Universidad de la República.

[a] Cálculo I may alternatively be taken as two sequential half-courses of 5 credits each, Cálculo I/A (114A) in the first semester and Cálculo I/B (128A) in the second.

[b] Offered as Introducción a la Economía through 2017 and renamed Introducción a la Microeconomía in 2018, with the course code unchanged.