The Price of Forced Attendance

Sacha Kapoor* Matthijs Oosterveen† Dinand Webbink‡

January 13, 2020

Abstract

We draw on a discontinuity at a large university, wherein second-year students with a low first-year GPA are allocated to a full year of forced, frequent, and regular attendance, to estimate the causal effect of additional structure on academic performance. We show that the policy increases student attendance but has no average effect on grades. The effects differ however depending on how course instructors handled unforced students, such that we observe significant grade decreases in courses where unforced students were given full discretion over their attendance. Our evidence suggests that grades decrease in these courses because the policy prevented forced students from picking their desired mix of study inputs.

JEL: I23, D12, J22

Keywords: Structure, Mandatory Attendance, Academic Performance, Higher Education

*Department of Economics, Erasmus School of Economics, Erasmus University Rotterdam. kapoor@ese.eur.nl

†Department of Economics, Lisbon School of Economics and Management, University of Lisbon. oosterveen@iseg.ulisboa.pt

‡Department of Economics, Erasmus School of Economics, Erasmus University Rotterdam, Tinbergen Institute, IZA Bonn. webbink@ese.eur.nl

§We thank Suzanne Bijkerk, Robert Dur, Julian Emami Namini, Johanna Posch, and Philip Oreopoulos for helpful comments and suggestions. The paper has also benefited from the comments and suggestions of participants at EEA-ESEM 2016, Erasmus University Rotterdam Seminar Series, IZA Summer School 2017, IZA Workshop on the Economics of Education, and the Tinbergen Institute Seminar Series. All omissions and errors are our own.

For many people their first real encounter with autonomy happens at college or university. Many students use this newfound autonomy to skip class, especially in the early years of their undergraduate education, choosing instead to focus on extracurricular activities such as student government or leisure with their friends. To combat the rampant absenteeism this newfound autonomy begets,[1] and because of the returns to college performance and graduation [Cunha, Karahan, and Soares, 2011, Jones and Jackson, 1990, Oreopoulos and Petronijevic, 2013], university administrators and instructors often mandate frequent and regular class attendance among their students.[2][3] These attendance policies provide students with structure, helping them to circumvent behavioral predispositions towards non-academic activities and ultimately avoid decisions that can be bad for their lifetime utility [Lavecchia, Liu, and Oreopoulos, 2014]. By this token, and as long as attendance is valuable, additional structure should be good for academic performance. At the same time, however, additional structure constrains choices (*e.g.* time on self study) which are important for grades and, by doing so, precludes sensible students from choices that best serve their own self interest. This can be bad for academic performance.

We draw on a natural experiment at a large European University to estimate the causal effects of a full year of forced, frequent, and regular attendance. The experiment requires students who average less than 7 (out of 10) in their first year to attend 70 percent of tutorials in each of their second-year courses. It imposes heavy time costs on students, as they can expect to spend 250 additional hours travelling and attending tutorials over a full academic year, amounting to approximately 7 additional hours per week. Students who fail to meet the attendance requirement face a stiff penalty, not being allowed to write the final exam for their course and having to wait a full academic year before they can take the course again. Because students have imprecise control over their

[1] Student absenteeism can be upwards of 60 percent of classes [Desalegn, Berhan, and Berhan, 2014, Kottasz et al., 2005, Romer, 1993].

[2] An early discussion of mandatory attendance can be found in the correspondence section of the Journal of Economic Perspectives in 1994 [Correspondence, 1994]. Motivated by Romer [1993], it consists of short letters by economics professors detailing their use of mandatory attendance.

[3] American universities spend 33 percent of their total budget on student instruction. This amounts to 56.7 billion dollars (for private nonprofit universities, years 2013-2014). Obtained via NCES: `https://nces.ed.gov/programs/digest/d15/tables/dt15_334.40.asp`, retrieved on 15-02-2017.

average grade in first year, the experiment facilitates a regression discontinuity design [Lee, 2008, Lee and Lemieux, 2010] for identifying the effects of forced attendance.

What does it mean to be forced? Our working definition is that a person is forced if a higher authority unilaterally takes away some of their potential choices. Or, more formally, if the authority imposes a heavy sometimes infinite penalty on a particular choice.[4] The policy we study is well within confines of this definition. The policy asks students to come to campus frequently and regularly, choices which are normally under the purview of the student, and imposes a heavy penalty when they fail to do so. In addition to fitting well with a natural definition for economists, students perceived the policy as one where their attendance was forced because this was how it was communicated to them by the university. Our data supports the notion that attendance was forced, as below-7 students collectively failed to meet the 70 percent criteria in less than one half of one percent of their courses. A more severe penalty, automatic expulsion *e.g.*, would have increased participation by less than half a percent, in other words.

Our estimates imply that the policy had no effect on second-year performance on average across all courses. The point estimate is negative, however, and allows us to rule out positive effects larger than 0.1 standard deviations with reasonable confidence. We document that this average effect hides substantial effect heterogeneity across courses. While the university required all students below 7 to attend 70 percent of tutorials in all their second-year courses, it had no policy on how students above 7 should be treated. Several courses overlaid their own attendance initiatives onto the university policy, each differing in the intensity of the attendance constraint they imposed on students who scored above 7 in first year. Some courses penalized absenteeism by any student, others strongly intimated and explained why all students should attend, while a third group of courses followed the university policy and left attendance decisions up to above-7 students. We observe the same students in all three scenarios because students have no discretion over

[4]Our paper is about more than just the role of sticks versus carrots in university education. A stick is typically defined as a penalty on performance, which itself is determined by choices and luck. Sticks constrain choices only implicitly, as the decision maker still has the freedom to make "bad" choices, and can simply hope that good luck helps them avoid penalties for poor performance.

course choice in second year.

The university policy had its largest effects in the third group of courses, which we will refer to as attendance-voluntary courses. For these courses the attendance of forced students increased by more than 50 percent, while their grades decreased by 0.16-0.26 standard deviations. We delve into mechanisms behind the decrease. We show first that the policy had its largest effects on the attendance of students who live far from campus and who had a greater propensity to miss tutorials in first year. We use course evaluations to show next that the policy generated an increase in lecture attendance similar to the increase in tutorial attendance, without having a measurable impact on total study time. The first result is consistent with students making calculated decisions about their attendance. The second result is consistent with the policy altering time spent on self study. The results together suggest that the policy prevented students from attaining their desired mix of study inputs, in line with existing evidence on the importance of time use for student performance [Stinebrickner and Stinebrickner, 2008]. We rule out a variety of other mechanisms, including the importance of an increase in exposure to other forced (and thus relatively low-achieving) peers in the tutorials, as well as the possibility of course heterogeneity in the usefulness of tutorials, or heterogeneity in course design more generally.

The university policy was abolished in the last year of our sample. The abolition came as a surprise, as students only learned of it after the start of their second year. We find no grade difference between above- and below-7 students in the abolition cohort, that the grades of above-7 students were the same in the abolition and treated cohorts, and that the grades of below-7 students in attendance-voluntary courses were higher in the abolition cohort compared to the treated cohorts. The abolition cohort evidence supports continuity of mean grades near 7 in the absence of treatment, a key identifying assumption in our regression discontinuity design. The evidence is also consistent with the grade decrease in attendance-voluntary courses being driven by grade decreases among forced students alone, and thus with the treatment having no (negative) spillover effects on the grades of unforced students.

Our study contributes to an expanding literature on incentives in education. A good deal of recent work analyzes the effects of interventions that reward students financially for "good" choices or better academic performance [Angrist, Oreopoulos, and Williams, 2014, Castleman, 2014, Cohodes and Goodman, 2014, De Paola, Scoppa, and Nistico, 2012, Dynarski, 2008, Leuven, Oosterbeek, and van der Klaauw, 2010].[5] We instead analyze the effect of an intervention which penalizes students heavily for "bad" choices, where the penalty is in terms of time rather than money.

Our findings contribute to debates over the merits of mandatory attendance in higher education [Romer, 1993].[6] The argument for mandatory attendance is based on a robust positive correlation between grades and attendance.[7] The argument has been reinforced by studies that use classroom or course-level evidence to show positive correlations between mandatory attendance and grades (see *e.g.* Marburger [2006], Dobkin, Gil, and Marion [2010], and Snyder et al. [2014]). We build on these studies by estimating the causal effect of a large-scale and year-long mandatory attendance policy.

There are several plausible explanations for why we find negligible to negative effects whereas positive effects have been reported in a wide range of contexts. One explanation relates to identification concerns that we are able to resolve, such as selection bias relating to cohort-specific unobservables or gaming for the purposes of avoiding mandatory attendance policies. A second explanation may simply be that our negative to negligible effects are not inconsistent with the positive effects researchers have found. It could be that the (average) treatment effect is positive and that our effects are specific to the types of students who would be at 7 in the context we study.

[5]For more comprehensive lists, at all levels of education, see Lavecchia, Liu, and Oreopoulos [2014] and Gneezy, Meier, and Rey-Biel [2011].

[6]Our study has an indirect link with the compulsory schooling literature [Angrist and Krueger, 1991, Oreopoulos, 2007]. We also examine the effect of a policy that penalizes people for specific choices. We differ in that our focus is on attendance at university, with steep and enforced penalties for absenteeism, and that we show that such policies can be costly for students.

[7]For some of the many examples, see Romer [1993], Durden and Ellis [1995], Kirby and McElroy [2003], Stanca [2006], Lin and Chen [2006], Marburger [2001], Martins and Walker [2006], Chen and Lin [2008], and Latif and Miles [2013].

This article contributes, more generally, to debates over the role of structure in higher education [Lavecchia, Liu, and Oreopoulos, 2014, Scott-Clayton, 2011]. Arguments for additional structure usually focus on student predispositions towards non-academic activities, emanating from behavioral biases such as impatience, or imperfect information about behaviors that engender success at university. Our findings imply that additional structure does not increase performance for students with a GPA of 7 (out of 10) at a large public university in the Netherlands.

1 Context

Our venue is the economics undergraduate program of a public Dutch university. The economics program itself is large; in the 2013-14 academic year alone, the program saw an influx of approximately 700 students. Students have no discretion over the courses they take in the first two years of the program, as all students follow the same ten courses per year, covering basic economics, business economics, and econometrics (See Table A.1 in the Appendix). Students have discretion over their courses in third year and, in line with this, declare a minor and major specialization (*e.g.* Accounting and Finance) which they can subsequently continue through to a Masters program.[8] The economics program is given in both Dutch and English. The only difference between the programs is that the Dutch program has approximately 2.5 times more students.

Academic years are divided into five blocks, of eight weeks each (seven weeks of teaching and one week of exams). First- and second-year students have one light and one heavy course in each block, where they get four credits for the light course, and eight for the heavy course.[9] Heavy courses have two to three large-scale lectures per week, while

[8] The Dutch and North American systems differ in two important ways. First, majors are defined more narrowly, as students decide to pursue economics, political science, sociology, and other social sciences before entering university. Second, they do three rather than four years of bachelors before a Masters.

[9] In Europe study credits are denoted by ECTS, which is an abbreviation for European Transfer Credit System. This is a common measure for student performance to accommodate the transfer of students and grades between European Universities. One ECTS is supposed to be equivalent to roughly 28 hours of studying. 60 ECTS account for one year of study.

light courses have one to two. Lecture attendance is always voluntary. Heavy courses have two small-scale tutorials (≈ 30 students) per week, while light courses have one. Lectures and tutorials both last for 1 hour and 45 minutes. Unlike lectures, but much like what may be found in structured college programs, tutorials require preparation and active participation of the student, via *e.g.* discussions of assignments and related materials.

Second-year courses each have several time slots for tutorials and students can choose the one they wish to attend. Students register for slots a few weeks before the block begins. At the time of registration, students are unaware of the teaching assistant (TA) that will teach each tutorial group, which are mostly senior-undergraduate and PhD students. Students cannot switch their tutorial group after the registration period ends. All students must register for a tutorial. We observe for which group and at which time the student registered and can evaluate whether there were systematic differences in registration patterns for forced and unforced students.

Grading is done on a scale that ranges from 1 to 10. Students fail a course if their grade is below 5.5. The grade point average (GPA) in first year is weighted by the credits obtained upon course completion. Note that a GPA of 8.25 or more at the end of first year is awarded *cum laude*.

1.1. University Policy. Second-year students that scored a GPA of less than 7 in first year were forced to attend 70 percent of tutorials for all of their second-year courses. Failure to fulfill the 70 percent attendance requirement precludes students from writing the final exam for the course. They must in turn wait a full year before they can take the course again in order to obtain these required credits.

Students that failed to complete their first year within year one, however, were forced to attend 70 percent of the second-year tutorials irrespective of their GPA. This implies there is only variation in the assignment to forced attendance for students near 7 that completed the first year on time. To complete the first year on time a student must score: (i) 5.5 or higher in each of the 10 first-year courses; (ii) or 5.5 or higher in most courses and use their high scores in these courses to compensate for grades of 4.5-5.4 in their remaining courses. The 10 first-year courses are assigned to one of three groups and students can only compensate one course within each group (see Appendix Table A.1). For example, a student who receives an 8 in microeconomics and 4.5 in macroeconomics can complete the first year by taking 1 point from their micro grade and use it towards their macro grade. The on-time completion rate for students near 7 is 92 percent. Half do this via criterion (i), while the other half do this via the compensation method described in criterion (ii).

Note that the on-time completion rule has no bearing on causal identification. The rules for completing the first year apply to both above- and below-7 students such that there is no sample selection. Consistent with this, we observe no statistical imbalance in the first-year completion rate nor in the use of the compensation method near 7. Throughout the paper we thus restrict the sample to students who completed the first year on time, which contains 92 percent of all students near 7.[10] In this sample, the mean and standard deviation of first-year GPA are 6.99 and 0.70. The analogues in the unrestricted sample are 6.65 and 0.79.

The policy imposes sizeable time costs on students. Forced students must spend 26 hours per block (3.5 hours per week) in tutorials.[11] Once we account for travel time, about 45 minutes each way on average,[12] forced students must spend 50 hours per block travelling to and attending tutorials. All costs are in terms of time rather than money because student travel is fully subsidized in the Netherlands.

The introduction of the policy had nothing to do with the historical grade distribution of first-year students. The policy was introduced as part of a university-wide initiative to personalize education via small-scale tutorials. The initiative came about for three rea-

[10]Note that the high first-year completion rates prevent us from estimating a local difference-in-difference, which would compare changes in the grades of students near the cutoff who completed first year, with the changes in the grades of students near the cutoff who did not complete the first year.

[11]This is based on the fact that there are 3 tutorials of 1.75 hour per week, 7 non-exam weeks in a block, and that students must attend 70 percent of tutorials.

[12]The average student lives 22.9 kilometers from campus. From the Dutch student survey "Studenten Monitor" we observe that more than 70 percent of university students travel by public transport (`http://www.studentenmonitor.nl/`). To get an idea of the travel time, we used the Dutch public transport website (`http://9292.nl/`) to check travel times between the university and the few larger cities within a radius of 20 and 30 kilometers of the university.

sons: first, the university had grown to a scale that made education impersonal; second, the tutorials encourage active participation; third, the tutorials facilitate student involvement in the university community. Forced attendance was made part of the initiative to ensure a return on the university's sizeable investment in small-scale tutorials.

1.2. Course Policies. While the university forced the attendance of below-7 students in all their second-year courses, courses differed in how they dealt with above-7 students. Appendix Table A.2 provides a detailed overview on the courses and, in particular, on how they dealt with these students. Attendance was voluntary in two of the courses; we will refer to them as *7⁺vol* courses (*i.e.* the attendance-voluntary courses). Three courses strongly encouraged the above-7 students to attend; *7⁺enc* courses. Three courses penalized them, and in fact also the below-7 students, for not attending; *7⁺for* courses. In these last set of courses, students had to complete assignments at the tutorials that made up five to thirty percent of their final grade. By not attending, students received a zero on this part of the course, meaning that at most they could obtain a 7 to 9.5 (rather than 10). The remaining two courses had no tutorials, and the final grade (mostly) consist of writing a research report in groups. Accordingly, these two courses are excluded from our analysis.[13] Ultimately, the course policies provide us with three counterfactuals: the grades of above-7 students whose attendance is voluntary, strongly-encouraged, and forced. The three counterfactuals help us sort through mechanisms which can generate and foster a relationship between forced attendance and academic performance.[14]

1.3. Abolition. The policy lasted five years, starting in 2009-10 and ending 2013-14. Thus, the 2008-09 cohort was the first to be subjected to the policy in their second year, while the 2012-13 cohort was the last. The policy was abolished in 2014-15 because the student body and faculty, rightfully, as this paper shows, lobbied against it. The abolition came as a surprise to the 2013-14 cohort, as they were only made aware of

[13]There is no difference in grades near 7 for these two courses. Note that they do not provide credible placebo tests as final grades are largely determined via group work.

[14]Appendix Table A.2 also shows multiple choice questions are used on the exams of all but one course. This precludes TAs from having a direct effect on grades.

it *after* their second year had started, in the first block of the academic year 2014-15. They had the same incentive to score above 7 in first year as earlier cohorts, even though below-7 students were ultimately given discretion over their attendance in second year.

2 Data

Our main information source is the administrative data of the university. Our sample ranges from the 2008-09 academic year until 2014-15. We observe grades at the level of the student for all three undergraduate years, tutorial attendance for the first two years, course evaluations, and various personal characteristics. As discussed in Section 1.1 we will restrict the analysis to students who completed their first year on time. After further restricting the sample to students that score a first-year GPA within 0.365 grade points of 7, our baseline estimation sample, we have 524 students and 3585 (second-year) course-student observations.

The university uses attendance lists to track student attendance at tutorials. Students must sign in and teaching assistants must upload the attendance data to the university's online portal. The uploaded data is then used by the exam administration to verify that the attendance requirement is met.[15]

We observe the attendance of each student at each tutorial session. We expect little measurement error because instructors required teaching assistants to prevent fraudulent sign-ins via counts of the number of students present. The attendance statistics for above-7 students reinforces this point. These students attend 55-60 percent of their tutorial sessions. We show later that they also attend roughly 55-60 percent of their lectures. The similarity between tutorial and lecture attendance, together with the idea that students incur sunk costs of visiting campus, suggests tutorial attendance is measured accurately.

[15]While matching attendance with the administrative data (*e.g.* grades and demographics), we experienced a match rate of 93 percent (in our baseline sample). We compared the matched observations with the non-matched observations and find that: (i) grades do not differ between the two groups; (ii) the treatment effect on grades is not different between the two groups; (iii) scoring below a seven in the first year could not explain whether or not a record is matched (See Table A.3 in the Appendix). Therefore we work with this 93-percent sample throughout the paper.

Our data includes information from course evaluations. One week before the exam, students are invited by email to evaluate the course anonymously. They are reminded of the evaluations shortly after the exam. All evaluations have the same 16 core questions, grouped into the general opinion of the course, structure, fairness, quality of lecturer and TA, and usefulness of lectures. Importantly, students are asked about their lecture attendance, as well as time spent on their studies in total.[16] Note that the evaluations are filled out by 20 percent of students. Later we will show that the response rate is the same just left and right of 7.

Our personal characteristics data includes gender, age, distance from their residence to the university (in kilometers), and whether they are from the European Economic Area (EEA).[17] For Dutch students, which form roughly 80 percent of our baseline estimation sample, we also have information on high school performance. Their grade for each of their high school courses is a 50-50 weighted average of the grade they earned in the course and the grade they earned on a nationwide exam for that course.

2.1. Basic Descriptives. Table 1 provides a basic summary of the data. The table compares students (who completed their first year within year one) with a first-year GPA between 6.635 and 7 to students whose GPA was between 7 and 7.365. The top panel restricts the sample to second-year courses, where the unit of observation is the student-course combination. The student is the unit of observation otherwise.

The top panel shows forced students score 0.48 standard deviations worse despite being 13 percentage points more likely to attend tutorials. The bottom panel implies students left and right of 7 are roughly similar. The lone statistical difference is for high school GPA, wherein poor performing students appear to be over-represented to the left of 7. Note, however, that the difference is statistically insignificant according to our main balancing tests presented later.

[16]For comprehensive details of the course evaluations see Table A.4 in the Appendix.

[17]Tuition fees are based on the student's EEA classification. Students who enroll in 2017-18, for example, pay €2,006 if they are from inside the EEA and €8,900 if not.

2.2. Preview of Baseline Results. The left column of Figure 1 examines the attendance effect for the three course types, where attendance is simply the percentage of tutorials attended (per course). In particular, the figures plot the second-year attendance rate against first-year GPA, where the difference (or jump) at a GPA of 7 measures the impact of the policy. In *7⁺vol* courses (panel a) this difference in attendance was between 30-35 percentage points. This translates into five extra tutorials for an eight credits course (three for a four credit course), or about 13 hours of extra schooling per block. In $7^{+}enc$ courses (panel b) the difference at a GPA of 7 was approximately 13 percentage points. There was no attendance difference in $7^{+}for$ courses (panel c).

Figure 1 suggests that the $7^{+}vol$ and $7^{+}enc$ attendance rates for forced students are higher than necessary. Forced students attend roughly 90 percent of their tutorials for both of these courses, whereas the requirement is 70 percent. What explains the discrepancy? One explanation relates to the discrete number of tutorials. Light (4 credit) courses have 7 tutorials. Attending 5 of 7 tutorials would give students a 71 percent attendance rate. Going from 5 to 6 tutorials, however, increases the completion rate to 86 percent. If there is some uncertainty about the completion rate, relating for example to how it is recorded, then risk averse students may wish to attend an additional tutorial just to make sure. In Section 5.2 we further document heterogeneous effects of the policy on attendance that support this interpretation.

The right column of Figure 1 examines the unconditional effect on grades. Grades in $7^{+}vol$ courses decrease by roughly 0.2 standard deviations. For the other courses there seems to be no effect on grades. The attendance and grade effects suggest that grades might only decrease if the additional constraint on choices is especially severe.

3 Empirical Specification

Let $G_j(D)$ denote the student's second-year grade in course j under regime D, where D indicates whether first-year GPA is less than 7. We are interested in the parameter

$$\tau = \mathbb{E}[G_j(1) - G_j(0) \mid GPA = 7], \tag{1}$$

the effect of forced attendance at 7. The adoption and use of the forced attendance policy suggests $\tau > 0$. The constraining effects of the policy on choices suggests $\tau < 0$.

We assume the conditional expectations $\mathbb{E}[G_j(1)|GPA = 7]$ and $\mathbb{E}[G_j(0)|GPA = 7]$ are continuous at 7 [Hahn, Todd, and Van der Klaauw, 2001]. Under this assumption τ is identified by

$$\lim_{x \to^- 7} \mathbb{E}[G_j \mid GPA = x] - \lim_{x \to^+ 7} \mathbb{E}[G_j \mid GPA = x], \tag{2}$$

where x is a realization of GPA, G_j is the observed grade, and $-$ and $+$ indicate whether GPA approaches 7 from below or above. The continuity assumption can fail if students have precise control over their first-year GPA [Cattaneo, Idrobo, and Titiunik, 2019b, Lee, 2008]. Because students were made aware of the policy early in their first year, they could try to avoid forced attendance in the second year. Our identification strategy works as long as first-year grades are somewhat outside of the student's control.

The above is generally a weak identifying assumption [Lee, 2008] and is reasonable in our setting. The assignment to forced attendance is based on the student's *average* grade. As students accumulate grades they lose control over the average. Importantly, first-year adjustments to the threat of second-year forced attendance, such as the practice of asking professors for grade increases,[18] have less of an effect on first-year GPA than on the grade of any one course.[19] Limited control over the average favors the continuity of

[18] Asking professors for grade increases, or any other such practice, can effect treatment assignment only when cumulative GPA is very close to 7.

[19] We are developing a companion article that studies adjustments to the threat of forced attendance. Our evidence shows that the threat does elicit a response but that, as expected, the response is almost never enough to get out of forced attendance. This claim is supported by various balancing and density

conditional expectations (for potential outcomes) at 7.

We use weighted least squares to estimate the limits in Equation (2) via the locally linear regression specification [Cattaneo et al., 2019b, Imbens and Lemieux, 2008]:

$$G_{ij} = \beta_0 + \beta_1 D_i + f_+(GPA_i - 7) + f_-(GPA_i - 7)D_i + \varepsilon_{ij}, \tag{3}$$

where i denotes the student. Second year grades G_{ij} are measured in standard deviations ($1\sigma = 1.45$), $f_+(\cdot)$ and $f_-(\cdot)$ are normalized linear polynomials in first-year GPA_i, and ε_{ij} is a random variable reflecting unobserved differences in second-year grades. We allow the polynomial to differ across 7 (see the discussion by Lee and Lemieux [2010]) and weight observations by a triangular kernel, which (linearly) assigns less weight to observations further from the cutoff. Our main estimates are based on the sample of students with first-year GPA between 6.635 and 7.365, *i.e.* within a bandwidth of 0.365 of 7. This is the optimal bandwidth for student grades relative to a MSE criterion [Calonico, Cattaneo, Farrell, and Titiunik, 2017] for the full sample of student-course observations (*i.e.* when including all three course types). Our decision to use a common bandwidth for the main estimates stems from the panel data structure. A common bandwidth ensures that the sample of students is the same across all our specifications.

Inference is based on standard errors clustered at the level of the student.[20] We rely mostly on conventional (clustered) standard errors because of our preference for a consistent sample across specifications. Since conventional standard errors are invalid for inference by construction [Calonico, Cattaneo, and Farrell, 2019], we report results based on robust bias-corrected standard errors and MSE-optimal bandwidths unique to each specification in the Appendix. A comparison will show that the estimates and statistical significance that are reported in the main text are on the conservative side.

3.1. Continuity Near the Cutoff. We examine the validity of the continuity assumption. We test for discontinuities in predetermined personal characteristics as well as the

discontinuity tests, as well as the null effects for abolition year.

[20] We do not cluster on the tutorial group because peer composition differs from course to course.

density of students near 7.

Table 2 presents estimates of equation (3) on the student level, where instead of grades the dependent variables are personal characteristics. Students to the left and right of the cutoff are similar in whether they come from the European Economic Area, age, distance from the university (in kilometers), and high school GPA. This conclusion holds if we select the bandwidth MSE-optimally for each background characteristic (Appendix Table A.5). It also holds if we consider grade differences for various high school courses separately (Appendix Figure A.1).

We next draw specifically on the test developed in [Cattaneo et al., 2018, 2019] to test for a discontinuity in the probability density for GPA at 7 [McCrary, 2008]. If students can manipulate their GPA, then we could observe bunching just above 7. The results of the test are summarized in Figure 2. The figure shows no evidence of bunching above the threshold. The bias corrected discontinuity test statistic is 0.25 with a p-value of 0.80, implying that we cannot reject the null hypothesis of no discontinuity at 7. This supports the absence of manipulation around the cutoff.

Although much of the evidence favors continuity, Column (3) of Table 2 indicates that women are underrepresented just to the right of the cutoff. The gender imbalance is problematic if it reflects men having a general tendency to ask for and obtain higher grades and, importantly, if this tendency generates a discontinuity in the conditional expectations for potential outcomes at 7. To test for this more specifically, Appendix Figure A.2 breaks down the density manipulation test by gender. The results imply that the probability density for both males and females around 7 is continuous, though graphically the support is strongest for males. Later we will show that our results are robust to controls for gender.

3.2. Abolition. We use the abolition cohort to further test the continuity assumption. For this cohort, the treatment regime D equals 0 across all realizations of GPA, such that there is no treatment effect at 7 for this cohort. To analyze whether this is the case, we plot second-year grades against first-year GPA as in Figure 1, but this time for the 2013-

14 cohort only. Appendix Figure A.3 documents the results and provides strong support for no difference in second-year grades at the GPA of 7 across all three course types. We confirm these zeros more formally in Appendix Table A.6, which reports estimates of equation (3) using only the abolition cohort.[21]

3.3. Sample Attrition. The policy may have incentivized students to drop courses if and once they fail the 70 percent attendance requirement. Attrition of this sort could threaten identification because dropouts are not graded. Accordingly, we test for a policy effect on the number of second-year courses for which a student obtained a valid grade. The results in Appendix Table A.7 (Column (1)) imply the policy has no effect on the number of completed courses, consistent with the fact that students near the cutoff tend to complete most of their second-year courses (more than 9 out of 10 on average).

Students near 7 may differ in their propensity to complete course evaluations and thus compromise the use of course evaluations in our analysis. Columns (2) to (4) of Appendix Table A.7 report estimates of the policy effect on an indicator for whether students completed the course evaluation for all three course types. We find no statistical differences in the propensity to complete the evaluation near 7. As with course completion, our evidence suggests no differential selection into course evaluations.

3.4. Mass Points. One remaining concern relates to whether first-year GPA has enough mass points to warrant a continuity-based RD design, which allows for the possibility that average potential outcomes vary with the running variable. To this end, note that there are 168 unique GPA values for the 524 students in our estimation sample of 6.635 to 7.365, amounting to approximately one GPA value for every 3 students. This amount of coverage of the support for GPA is usually sufficient for a continuity-based design.[22]

[21] Note that we cannot analyze the implications for attendance because the university stopped registering attendance in the abolition year.

[22] Cattaneo, Idrobo, and Titiunik [2019a] analyze an example where for every 110 observations one unique value for the running variable is observed. They conclude that continuity-based analysis might be possible in this context.

4 Baseline Results

Table 3 reports estimates for student grades based on pooled data from the 8 affected courses. Pooling is advantageous because it lets us account for across-course error correlation within students. Average effect estimates are found in Columns (1) to (3). Columns (2) and (3) show the estimates do not change when controlling for fixed effects for the course-cohort combination and for personal characteristics. The point estimates are negative, but not statistically different from zero, and imply that the university-wide policy had little to no average effect on student performance.

4.1. Course-Level Attendance Policies. Table 4 evaluates the policy effect for the three course types separately. Moving left to right, the table reports estimates for $\mathcal{T}^+ vol$ courses, $\mathcal{T}^+ enc$ courses, and $\mathcal{T}^+ for$ courses. The table starts with the effect on tutorial attendance in the top panel. The estimates show that the policy increased the attendance of forced students in $\mathcal{T}^+ vol$ courses by 31 percentage points ($p < 0.01$), increased attendance by 13 percentage points in $\mathcal{T}^+ enc$ courses ($p < 0.01$), and had no statistical effect on attendance in $\mathcal{T}^+ for$ courses. The estimates in the top panel show that the policy had a first-order effect on student choices.

The middle panel of Table 4 evaluates the effect on grades. It shows that the policy decreased grades by 0.18 standard deviations in $\mathcal{T}^+ vol$ courses ($p < 0.1$). On the Dutch grading scale this amounts to approximately 0.3 grade points ($\approx 0.18 \times 1.45$). The grades of forced students were 0.04 standard deviations higher in $\mathcal{T}^+ enc$ courses and 0.03 standard deviations lower in $\mathcal{T}^+ for$ courses. The latter two estimates are statistically insignificant. Note that Columns (4) to (6) of Table 3 show similar conclusions are reached with pooled data and interactions between the treatment and course type.

Whereas forced students obtain lower grades in $\mathcal{T}^+ vol$ courses, this does not necessarily mean that they also obtain lower passing rates. We explore whether passing rates are affected in the bottom panel of Table 4, which is equivalent to checking whether the grade decreases happen at 5.5; the threshold for passing a course. The results show that the probability of passing is 7 percentage points lower in *7⁺vol* courses. The estimate is insignificant at conventional significance levels, however ($p \approx 0.12$). Columns (2) and (3) show there is effectively no difference in passing rates for *7⁺enc* and *7⁺for* courses. We conclude that the impact on passing rates are small to negligible.

4.2. Robustness. We analyzed the robustness of the heterogeneous policy effects across the three course types. Appendix Table A.8 tests whether the effects are robust to the inclusion of course-cohort fixed effects and personal characteristics. Appendix Table A.9 further includes high school GPA, which we only observe for the Dutch students in our sample. Both tables show that the baseline results in Table 4 are robust. This is reassuring, especially with respect to the possible gender imbalance at the cutoff.

Appendix Table A.10 reports estimates of specifications that use bandwidths which are optimal for each course type. The bandwidths are MSE optimal when estimating the discontinuity at 7. They are CER (coverage error) optimal for the purpose of robust bias-corrected inference, as recommended in [Cattaneo, Idrobo, and Titiunik, 2019b]. The results across all outcomes are very similar, where the estimate in Column (1) of the middle panel implies grades of forced students decrease by 0.26 standard deviations in *7⁺vol* courses, which is statistically significant at the 5 percent level.

Appendix Figure A.4 explores whether the estimate for *7⁺vol* courses is robust to the choice of the bandwidth. It shows that the estimate on student grades hovers between -0.15 and -0.30 while using bandwidths between 0.10 (first-year GPA of 6.9 to 7.1) and 0.50 (first-year GPA of 6.5 to 7.5). Unsurprisingly the confidence intervals are too wide to reject a null estimate with very small bandwidths. The baseline estimate, however, is statistically significant at bandwidths between 0.15 and 0.40, where the p-value reaches values slightly above 10 percent while using the largest bandwidths. We also tested for discontinuities at the fake cutoffs of 6, 8, 8.25 (cum laude), and 9 in all our main outcomes for *7⁺vol* courses. Appendix Table A.11 documents an absence of significant discontinuities across all student outcomes and all fake cutoffs.

4.3. Abolition Cohort. Table 5 reports mean unstandardized *7⁺vol* grades for just below and just above 7 students in the treated and abolition cohorts. The top row shows a grade difference of 0.37 (on a 10-point scale) across below- and above-7 students in treated cohorts. The bottom row shows a grade difference of 0.13 for the abolition cohort. The grade difference for the abolition cohort is statistically insignificant. It is approximately one third of its analog for treated cohorts. The evidence is consistent with no grade difference in the abolition cohort or with a grade difference that is abnormally small.

The left column shows below-7 students from the abolition cohort have grades that are 0.35 points higher than the grades of below-7 students from earlier treated cohorts. The across cohort difference in the left column is similar to the within cohort difference of 0.37 in the top row. The grade decrease we observe therefore reflects behavioral changes by forced students rather than behavioral changes by unforced students.

The right column of Table 5 supports this conclusion, showing the grades of above-7 students from the abolition cohort are 0.11 points higher than the grades of above-7 students from earlier treated cohorts. The difference is statistically insignificant and small relative to other differences in the table. If cohort-specific differences are negligible, then no difference in the grades of above-7 students would be consistent with no spillovers from forced to unforced students [Dong and Lewbel, 2015]. This in turn would suggest it is the behavior of forced students themselves that drives the grade decrease in *7⁺vol* courses.

5 Baseline Mechanisms

5.1. Tutorial Quality. The grade decrease in *7⁺vol* courses may be attributable to especially poor tutorial quality in these courses. This may also explain why there is a grade decrease in *7⁺vol* courses and no grade difference in *7⁺enc* courses. We investigate this possibility using TA evaluations as proxy for tutorial quality, regressing these evaluations on indicators for the three different course types. If *7⁺vol* tutorials are indeed poor or ineffective, then we expect lower TA evaluations in these courses. Estimates are found in Appendix Table A.12. Note that we use TA evaluations from the abolition year to circumvent concerns about whether the evaluations are contaminated by participation in forced attendance.

Column (1) shows that *7⁺vol* TAs score 0.21 points higher than *7⁺enc* TAs ($p < 0.1$) on the question "TA gives good tutorials". They score about the same relative to TAs in *7⁺for* courses. The results in Column (2) imply that the TAs across all three course types provide similar levels of assistance. We conclude that TA quality is in fact moderate to relatively high in *7⁺vol* courses.

The grade decrease in *7⁺vol* courses may be attributable more broadly to how these courses are designed. Course instructors may let above-7 students keep their discretion over tutorial attendance and consequently ensure that all students could obtain everything they needed to know via the plenary lectures alone. In this scenario, the TAs for *7⁺vol* courses can be excellent yet contribute little to student performance. Two pieces of evidence contradict this possibility. Section 5.4 will show first that the university policy generated parallel increases in lecture and tutorial attendance. If the lectures for *7⁺vol* courses were exceptionally useful then grades should have been higher, rather than lower, for forced students. Second, we regress student perceptions of lecturer quality on indicators for the three different course types again using data from the abolition year (columns (3) and (4) of Appendix Table A.12). If lecturers made their courses lecture-heavy, then we expect higher perceived lecturer quality in these courses. Yet we find that the perceived lecturer quality is the same across the three course types.

5.2. Attendance Price and Propensity. We investigate whether our treatment effects differ depending on the distance of a student's residence to the university and on the propensity to attend first-year tutorials. We first estimate:

$$A_{ij} = \gamma_0 + \gamma_{1i} D_i + f_+(GPA_i - 7) + f_-(GPA_i - 7) D_i + \varepsilon_{ij}, \tag{4}$$

where A_{ij} is the percentage of tutorials attended in second year. If γ_{1i} is large then the student's desired attendance is low, such that they would have attended far fewer tutorials in the absence of forced attendance. Alternatively, a small γ_{1i} implies attendance is desirable, such that the student attends the same number of tutorials with or without forced attendance. In the parlance of the treatment effects literature [Angrist and Pischke, 2008], students who otherwise prefer not to attend (large γ_{1i}) are compliers. Students who would attend anyways (small γ_{1i}) are always takers. There are no never takers or defiers by the very definition of the policy, as it leaves students with no choice but to attend tutorials when their first-year GPA is below 7. Indeed, of the courses from students with a first-year GPA below 7, only 0.44 percent have an attendance rate below 70 percent.[23]

We interpret distance to the university as a proxy for the price of attendance and average tutorial attendance in first year as a proxy for the additional utility from attendance. Distant students pay a higher price for attendance because they have to spend more time travelling to campus. Students with a high propensity to attend in first year presumably derive additional utility from attendance in second year.[24] We thus operationalize γ_{1i} via treatment interactions with our proxies for the price of and additional utility from attendance. Estimates are found in the first three columns of Table 6, where Column (1) focuses on *7⁺vol* courses. Note that distance and first-year attendance are standardized, where the standard deviations are 30.3 kilometers for distance and 0.07 for attendance (on a scale from 0 to 1).

Three patterns stand out. First, the direct effect of the proxy is always opposite, but similar in magnitude, to its interaction effect. This suggests the interactions pick up the student's counterfactual attendance had the policy not been in place. Second, the policy had a larger effect on students who live far from campus. The attendance effect increases by 6 percentage points for students that live one standard deviation further

[23] One might argue that the grade for never takers are never observed, as they cannot write the exam. However, in Section 3.3 we showed students generally participate in every second-year course, and that their near-perfect course participation is unaffected by the treatment (leaving no room for never takers).

[24] This proxy is implied by the assumption that preferences over tutorial attendance are stable from first to second year. Our results are consistent with this assumption.

from campus. This suggests distant students have a greater propensity to attend less in the absence of forced attendance. Third, the policy had a smaller effect on students who have a higher attendance propensity. The attendance effect decreases by 13 percentage points for students who attended one standard deviation more tutorials in first year. The last two patterns are consistent with students making calculated decisions about their tutorial attendance.

5.3. Differential Grade Effects. The differential effects on tutorial attendance are consistent with the university policy constraining calculated decisions by forced students. We check for similar differential effects on academic performance. Our idea is that, if the additional constraint on attendance drives the grade decreases in *7⁺vol* courses, then grades should decrease by more for students who live far from campus and who have a low propensity for tutorial attendance in first year. Column (4) to (6) of Table 6 show the heterogeneity results for grades and Column (7) to (9) do so for passing rates. Column (4) and (7) focus on the sample of *7⁺vol* courses.

The results imply that the interaction effects for distance and attendance propensity on academic performance are both statistically and substantively small. While the estimates fail to support a mechanism where grades decrease because the policy constrains student behavior, it is not necessarily inconsistent with this mechanism. Students may be compensating for the lost time and energy in a variety of unobserved ways. For example, distant students may use their additional travel time to contemplate and study the material.

5.4. Other Input Choices. To better understand the impact of the policy on student input choices, Table 7 uses course evaluations to investigate the effect on lecture attendance and total study time. The top panel reports the effect on an indicator for whether the student attended lectures. The bottom panel reports the effect on total study time (lectures+tutorials+self study).

The top panel shows forced students are 25 percentage points more likely to attend lectures in *7⁺vol* courses ($p \approx 0.11$), 8 percentage points more likely in *7⁺enc* courses

($p > 0.10$), and 5 percentage points less likely in *7⁺for* courses ($p > 0.10$). The slope estimates, while insignificant, align well with how tutorial attendance changed across the three course types (top panel of Table 4). The slope estimates for lecture and tutorial attendance are both largest in 7^+*vol* courses and smallest in 7^+*for* courses, and have similar orders of magnitudes. The intercept estimates of Table 7 also align well with the intercepts for tutorial attendance (left panel of Figure 1, right of 7). The similarities between the slopes and intercepts suggest that the policy forces students to pay a time cost that becomes sunk after they arrive at campus, such that lecture attendance is relatively cheap when the student is already there.

The bottom panel of Table 7 also shows that the policy increased total study time by about 2 hours in 7^+*vol* courses, 4.5 hours in 7^+*enc* courses, and 2 hours in 7^+*for* courses. The estimates are all statistically insignificant, implying we cannot rule out no effect of the university policy on total study time. A null or small positive effect on total study time, together with large attendance increases for tutorials and lectures, would imply that the university policy decreased time spent on self study. Less time on self study would further suggest that input choices other than tutorial attendance were affected by the policy.[25] This explanation for the grade decrease fits well with the careful time use study of Stinebrickner and Stinebrickner [2008]. They use the random assignment to a roommate with a video game to show that a one hour reduction in self study (in the first semester) causes GPA to decrease by 0.36 points.

5.5. Peer Effects. By forcing tutorial attendance, the policy increases the exposure of forced students to other forced and therefore relatively low-achieving students. Additional exposure to low achievers provides an alternative explanation for the grade decrease in 7^+*vol* courses. As a first step towards understanding the importance of this mechanism,

[25]Though our estimates suggest a decline in self study, we do not make a precise calculation because magnitudes relating to lecture attendance and total study time should be interpreted with caution. First, while not selective with respect to the policy, the course evaluations are filled out by 20 percent of the students. Second, total study time is measured in 10 categories (1=0 hours, 2=1 to 5 hours, and 10=more than 40 hours). We used the maximum for the interval to convert the categories into hours, where the category 10 is assigned 45 hours. Only the intercepts change if we use the minimum or the mean.

we evaluate whether there are indeed differences in exposure to forced students. We use our rich attendance data to construct the following exposure measure for student i in course j:

$$Exposure_{ij} = \frac{1}{S_j} \sum_{s=1}^{S_j} \mathbb{1}[A_{isj} = 1] \Big(\frac{N_{-isj}}{N_{sj} - 1} \Big),$$

where S_j is the total number of tutorial sessions in course j, A_{isj} is the attendance of i in session s, and $N_{-isj}/(N_{sj} - 1)$ is the leave-out proportion of forced students that attended that specific tutorial session. We then use the treatment effect on $Exposure_{ij}$ to quantify the additional exposure of forced students.

Estimates are found in the odd numbered columns of the top panel of Table 8. The policy increased exposure by 24 percentage points in $7^{+}vol$ courses ($p < 0.01$), by 12 percentage points in $7^{+}enc$ courses ($p < 0.01$), and had no effect on exposure in $7^{+}for$ courses.

Our exposure measure stresses two channels for the increased exposure in $7^{+}vol$ and $7^{+}enc$ courses. One channel relates to the simple fact that forced students are more likely to attend tutorials ($\mathbb{1}[A_{isj} = 1] = 1$). The other channel relates to the possibility that forced students may be more likely to attend tutorials with other forced students even after conditioning on attendance probabilities. This can be the case if forced students deliberately register for the same tutorial group or deliberately attend the same tutorial sessions within that group. These sorts of coordination can foster bad peer influence among forced students.

We assess these channels separately by estimating specifications that control for the course-specific attendance rate of the student in the even numbered columns of the top panel of Table 8. We show that the exposure differences are much smaller (close to 0 in fact) once we control for attendance rates. This is consistent with the unconditional treatment effect reflecting a mechanical increase in attendance rates rather than increased and deliberate coordination with other low-achieving peers. The evidence suggests in turn that if peer effects are present, they are not operating through the coordination decisions of forced students.

We also evaluated the potential importance of peers while using the most common peer effects specification in the literature [Booij, Leuven, and Oosterbeek, 2017]. More specifically, we re-estimated our baseline equation while including treatment interactions with measures of peer quality, the average first-year GPA of the peer group and the average peer registration time for tutorials. Both are leave-out means, where the average first-year GPA is standardized and the tutorial registration time is measured in differences in days from the course mean registration time and subsequently averaged across one's peers. The latter measure reflects the idea that weak students might leave tutorial registration to the last minute.

The bottom panel of Table 8 shows the results, where all the effects of treatment interactions with peer quality are modest. All the estimates are statistically insignificant at conventional significance levels, while the main treatment effect estimate is unchanged compared to our baseline specifications. Negligible peer effects are unsurprising given recent discussions and results in the literature [Sacerdote, 2014].[26] Altogether the evidence suggests that relatively heavy exposure to forced peers is not an important mechanism for the grade decrease in *7^{+}vol* courses.

6 Conclusion

We draw on a discontinuity at a large public Dutch university, wherein second-year students with a first-year GPA below 7 were allocated to a full year of forced, frequent, and regular attendance, to estimate the causal effect of additional structure on academic performance. Our estimates imply that forced students, with a first-year GPA at 7, can not expect a positive effect on their GPA in second year. The average null estimate masks differential effects that are attributable to how course instructors dealt with above-7 students. We show that the policy had its largest effects in courses where above-7 students

[26] See Feld and Zölitz [2017] and Booij, Leuven, and Oosterbeek [2017] for two recent examples from the Dutch context.

were allowed to decide their attendance, as the attendance of forced students increased by more than 50 percent, and their grades decreased by about 0.16 to 0.26 standard deviations in these courses.

We find some evidence that the grade decreases are explained by the constraining effects of the policy, *i.e.* that the policy prevented students from attaining their desired mix of study inputs. We rule out a variety of other mechanisms, including the importance of an increase in exposure to other forced (and thus relatively low achieving) peers, as well as the possibility of course heterogeneity in the usefulness of tutorials, or heterogeneity in course design more generally.

References

Angrist, J. D. and A. B. Krueger (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics 106*(4), 979–1014.

Angrist, J. D., P. Oreopoulos, and T. Williams (2014). When opportunity knocks, who answers? new evidence on college achievement awards. *Journal of Human Resources 1*(1), 1–29.

Angrist, J. D. and J.-S. Pischke (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Booij, A. S., E. Leuven, and H. Oosterbeek (2017). Ability peer effects in university: Evidence from a randomized experiment. *The Review of Economic Studies 84*(2), 547–578.

Calonico, S., M. D. Cattaneo, and M. Farrell (2019). Optimal bandwidth choice for robust bias corrected inference in regression discontinuity designs. *Working Paper*.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2017). rdrobust: Software for regression-discontinuity designs. *The Stata Journal 17*(2), 372–404.

Castleman, B. L. (2014). Prompts, personalization, and pay-offs: Strategies to improve the design of college and financial aid information. *The George Washington University Graduate School of Education and Human Development*.

Cattaneo, M., M. Jansson, and X. Ma (2018). Manipulation testing based on density discontinuity. *Stata Journal 18*(1), 234–261.

Cattaneo, M., M. Jansson, and X. Ma (2019). Simple local polynomial density estimators. *Journal of the American Statistical Association forthcoming*.

Cattaneo, M. D., N. Idrobo, and R. Titiunik (2019a). A practical introduction to regression discontinuity designs: Extensions. *Cambridge Elements: Quantitative and Computational Methods for Social Science*.

Cattaneo, M. D., N. Idrobo, and R. Titiunik (2019b). A practical introduction to regression discontinuity designs: Foundations. *Cambridge Elements: Quantitative and Computational Methods for Social Science*.

Chen, J. and T.-F. Lin (2008). Class attendance and exam performance: A randomized experiment. *The Journal of Economic Education 39*(3), 213–227.

Cohodes, S. and J. Goodman (2014). Merit aid, college quality and college completion: Massachusetts' adams scholarship as an in-kind subsidy. *American Economic Journal: Applied Economics 6*(4), 251–285.

Correspondence (1994). Correspondence: Should class attendance be mandatory. *Journal of Economic Perspectives 8*(3), 205–216.

Cunha, F., F. Karahan, and I. Soares (2011). Returns to skills and the college premium. *Journal of Money, Credit and Banking 43*(s1), 39–86.

De Paola, M., V. Scoppa, and R. Nistico (2012). Monetary incentives and student achievement in a depressed labor market: Results from a randomized experiment. *Journal of Human Capital 6*(1), 56–85.

Desalegn, A. A., A. Berhan, and Y. Berhan (2014). Absenteeism among medical and health science undergraduate students at hawassa university, ethiopia. *BMC medical education 14*(1), 81.

Dobkin, C., R. Gil, and J. Marion (2010). Skipping class in college and exam performance: Evidence from a regression discontinuity classroom experiment. *Economics of Education Review 29*(4), 566–575.

Dong, Y. and A. Lewbel (2015). Identifying the effect of changing the policy threshold in regression discontinuity models. *Review of Economics and Statistics 97*(5), 1081–1092.

Durden, G. C. and L. V. Ellis (1995). The effects of attendance on student learning in principles of economics. *The American Economic Review 85*(2), 343–346.

Dynarski, S. (2008). Building the stock of college-educated labor. *Journal of Human Resources 43*(3), 924–937.

Feld, J. and U. Zölitz (2017). Understanding peer effects-on the nature, estimation and channels of peer effects. *Journal of Labor Economics 35*(2).

Gneezy, U., S. Meier, and P. Rey-Biel (2011). When and why incentives (don't) work to modify behavior. *The Journal of Economic Perspectives 25*(4), 191–209.

Hahn, J., P. Todd, and W. Van der Klaauw (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica 69*(1), 201–209.

Imbens, G. W. and T. Lemieux (2008). Regression discontinuity designs: A guide to practice. *Journal of econometrics 142*(2), 615–635.

Jones, E. B. and J. D. Jackson (1990). College grades and labor market rewards. *The Journal of Human Resources 25*(2), 253–266.

Kirby, A. and B. McElroy (2003). The effect of attendance on grade for first year economics students in university college cork. *The Economic and Social Review 34*(3), 311–326.

Kottasz, R. et al. (2005). Reasons for student non-attendance at lectures and tutorials: An analysis. *Investigations in university teaching and learning 2*(2), 5–16.

Latif, E. and S. Miles (2013). Class attendance and academic performance: a panel data analysis. *Economic Papers: A journal of applied economics and policy 32*(4), 470–476.

Lavecchia, A. M., H. Liu, and P. Oreopoulos (2014). Behavioral economics of education: Progress and possibilities. Technical report, National Bureau of Economic Research.

Lee, D. S. (2008). Randomized experiments from non-random selection in us house elections. *Journal of Econometrics 142*(2), 675–697.

Lee, D. S. and T. Lemieux (2010). Regression discontinuity designs in economics. *Journal of economic literature 48*(2), 281–355.

Leuven, E., H. Oosterbeek, and B. van der Klaauw (2010). The effect of financial rewards on students' achievement: Evidence from a randomized experiment. *Journal of the European Economic Association 8*(6), 1243–1265.

Lin, T.-F. and J. Chen (2006). Cumulative class attendance and exam performance. *Applied Economics Letters 13*(14), 937–942.

Marburger, D. R. (2001). Absenteeism and undergraduate exam performance. *The Journal of Economic Education 32*(2), 99–109.

Marburger, D. R. (2006). Does mandatory attendance improve student performance? *The Journal of Economic Education 37*(2), 148–155.

Martins, P. S. and I. Walker (2006). Student achievement and university classes: Effects of attendance, size, peers, and teachers. Technical report.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of econometrics 142*(2), 698–714.

Oreopoulos, P. (2007). Do dropouts drop out too soon? wealth, health and happiness from compulsory schooling. *Journal of Public Economics 91*(11), 2213–2229.

Oreopoulos, P. and U. Petronijevic (2013). Making college worth it: A review of research on the returns to higher education. *NBER Working Paper No. 19053*.

Romer, D. (1993, September). Do students go to class? should they? *Journal of Economic Perspectives 7*(3), 167–174.

Sacerdote, B. (2014). Experimental and quasi-experimental analysis of peer effects: two steps forward? *Annu. Rev. Econ. 6*(1), 253–272.

Scott-Clayton, J. (2011). The shapeless river: Does a lack of structure inhibit students' progress at community colleges. *CCRC Working Paper No. 25*.

Snyder, J. L., J. E. Lee-Partridge, A. T. Jarmoszko, O. Petkova, and M. J. D'Onofrio (2014). What is the influence of a compulsory attendance policy on absenteeism and performance? *Journal of Education for Business 89*(8), 433–440.

Stanca, L. (2006). The effects of attendance on academic performance: Panel data evidence for introductory microeconomics. *The Journal of Economic Education 37*(3), 251–266.

Stinebrickner, R. and T. R. Stinebrickner (2008). The causal effect of studying on academic performance. *B.E. Journal of Economic Analysis and Policy 8*(1).

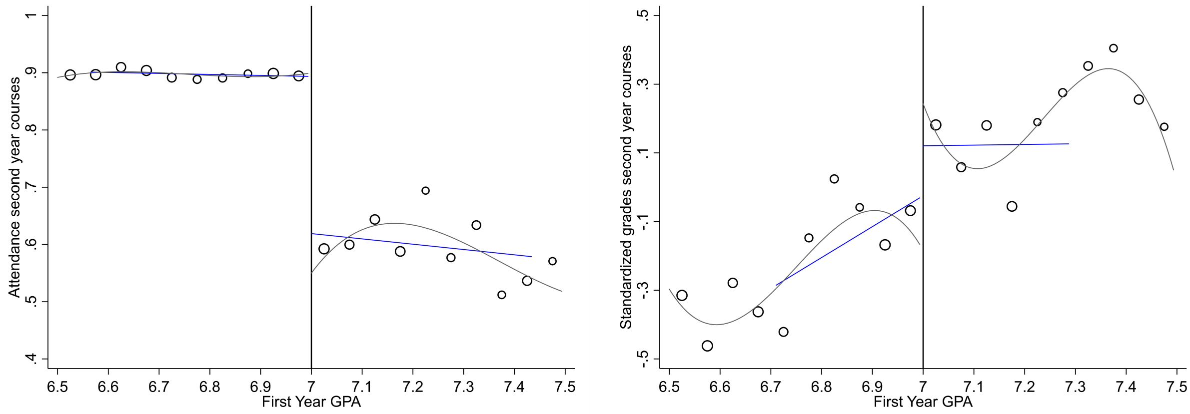
Table 1: **Basic Descriptives.**

Variable	First-year GPA [6.635,7)	[7,7.365]	Diff
Course Level (Second Year)			
Grade	6.33	6.81	0.481***
	(1.33)	(1.19)	(0.059)
Tutorial Attendance	0.90	0.77	-0.130***
	(0.12)	(0.29)	(0.011)
Observations	1827	1758	3585
Student Level (All Students)			
Distance to University (km)	23.18	22.12	-1.061
	(31.66)	(28.81)	(2.649)
Age	20.28	20.16	-0.126
	(1.07)	(1.20)	(0.099)
Gender (Female=1)	0.30	0.31	0.008
	(0.46)	(0.46)	(0.040)
European Economic Area	0.94	0.92	-0.015
	(0.24)	(0.27)	(0.022)
Observations	269	255	524
Student Level (Dutch Students)			
High School GPA	6.68	6.92	0.237*
	(1.33)	(1.34)	(0.128)
Observations	225	206	431

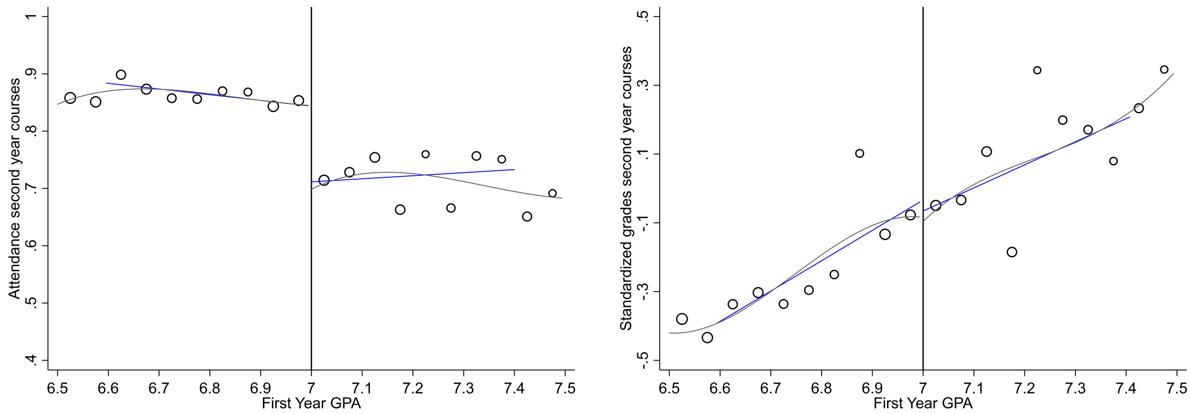
Notes:
1. Sample is from all 8 eligible courses.
2. Grades and high school GPA range from 1 to 10.
3. Each high school grade is a 50-50 weighted average of the grade the high school assigned and the grade the student received on a national exam for the course.
4. First two columns have standard deviations in parentheses. Last column has standard errors in parentheses.
5. Stars denote statistical significance for difference in means, standard errors clustered on student level.
6. Significance levels: $^{*} < 10\%$ $^{**} < 5\%$ $^{***} < 1\%$.

Figure 1: **Second Year Attendance and Grades, by Course Type.**

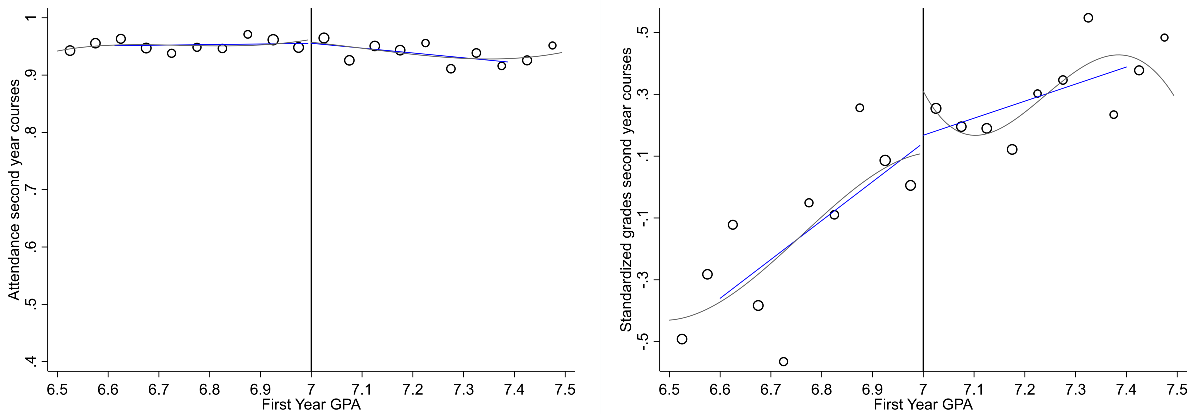
(a) Attendance is Forced Left of 7, Voluntary to Right



(b) Attendance is Forced Left of 7, Strongly Encouraged to Right



(c) Attendance is Forced to Left and Right of 7



Notes:

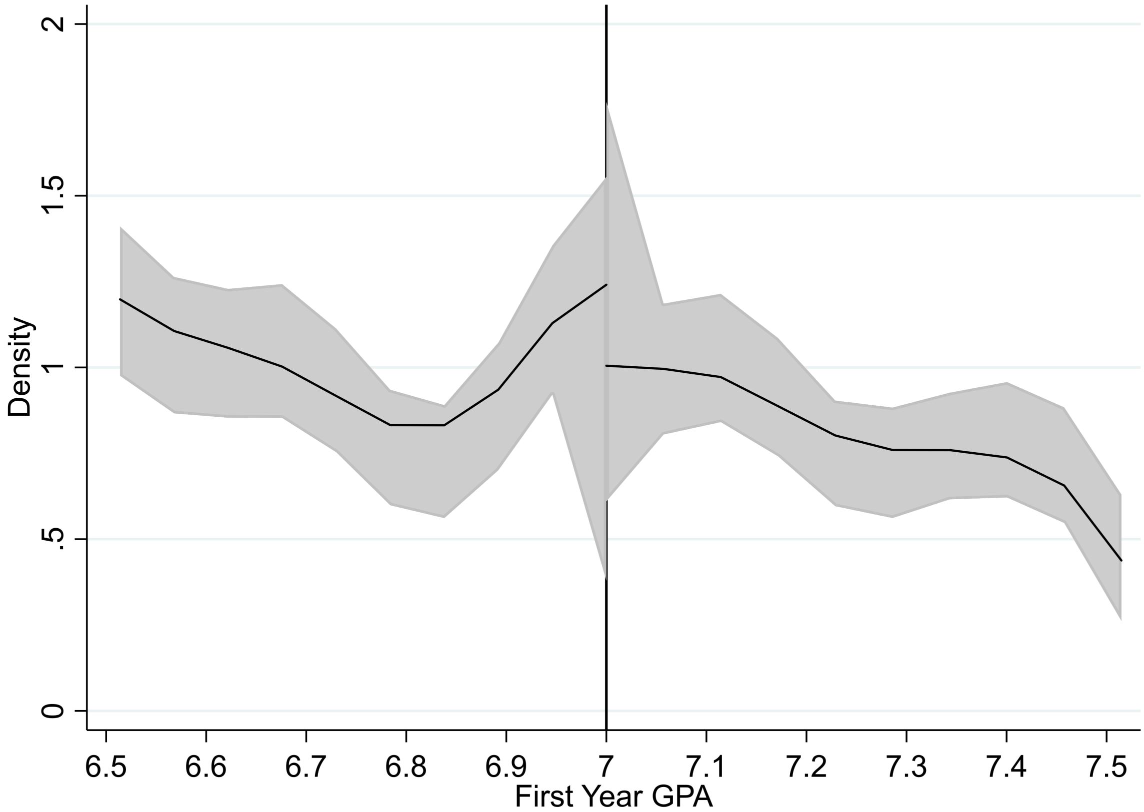
1. Locally linear and cubic scatterplots for attendance or second-year grades against first-year GPA.
2. The local linear polynomial is estimated upon the optimal bandwidth for each outcome relative to a MSE criterion [Calonico, Cattaneo, Farrell, and Titiunik, 2017]. The cubic polynomial is estimated upon a bandwidth of 0.5, which is the same across all figures.
3. Dots are based on local averages for a binsize of 0.05. Dot sizes reflect the number of observations used to calculate the average.
4. Binsizes for local averages are selected via F-tests from regressions of second-year grades on K bin dummies and $2K$ bin dummies for the first-year GPA.

Table 2: **Balancing Tests around the Cutoff.**

	Distance to Uni. (km)	Age	Gender	European Economic Area	High School GPA
	(1)	(2)	(3)	(4)	(5)
1^{st}-year GPA is Below 7	3.213	0.247	0.173*	-0.024	-0.428
	(6.134)	(0.178)	(0.083)	(0.051)	(0.310)
Mean Dep. Var.	22.979	20.289	0.290	0.938	6.882
Observations	524	524	524	524	431

Notes:
1. Unit of observation is the student. The outcome variable is displayed at the top of each column. The outcome variables are not standardized.
2. The regressions include a first-order polynomial which is interacted with the treatment. The bandwidth is 0.365 and the kernel is triangular.
3. Standard errors are clustered on the student and in parentheses.
4. Significance levels: $^{*} < 10\%$ $^{**} < 5\%$ $^{***} < 1\%$.

Figure 2: **Discontinuity in Density Test.**



Notes:

1. Density is for the number of students.
2. Figure plots unrestricted manipulation test, where under the null hypothesis the limiting densities of the number of students to the left and right of 7 are the same. Test is unrestricted in that the estimates of densities to left and right of 7 are unrelated.
3. Figure uses a second-order polynomial for density estimation and a third-order polynomial for the bias-correction estimate (see [Cattaneo, Jansson, and Ma, 2018, 2019]). Kernel is triangular. Confidence intervals use jackknifed standard errors.
4. The bias corrected discontinuity test statistic and p-value are 0.25 and 0.80 respectively, implying that we cannot reject the null hypothesis of no discontinuity around the cutoff, and suggesting that there is no manipulation around the cutoff.

Table 3: **Student Performance for All 8 Eligible Courses.**

	Grade (Standardized)					
	(1)	(2)	(3)	(4)	(5)	(6)
1^{st}-year GPA	-0.04	-0.02	-0.01	0.04	0.07	0.07
is Below 7	(0.08)	(0.07)	(0.07)	(0.10)	(0.10)	(0.10)
Attendance is Voluntary				-0.22*	-0.23*	-0.23*
× Treatment				(0.13)	(0.12)	(0.12)
Absence is Penalized				-0.07	-0.08	-0.08
× Treatment				(0.12)	(0.11)	(0.10)
Course-Cohort FE	No	Yes	Yes	No	Yes	Yes
Personal Characteristics	No	No	Yes	No	No	Yes
Observations	3585	3585	3585	3585	3585	3585

Notes:
1. Grades are standardized, where one standard deviation equals 1.45 grade points on the Dutch grading scale.
2. Controls for personal characteristics include distance to the university, age, gender, and European economic area.
3. The regressions include a first-order polynomial which is interacted with the treatment. The bandwidth is 0.365, the (MSE) optimal bandwidth for the baseline RD specification with all 8 eligible courses and no controls. The kernel is triangular. In Columns (4) to (6) the treatment effect and the polynomials are allowed to differ by course type.
4. Standard errors are clustered on the student and in parentheses.
5. Significance levels: $^{*} < 10\%$ $^{**} < 5\%$ $^{***} < 1\%$.

Table 4: **Student Attendance and Performance by Course Type.**

	(1)	(2)	(3)
		Attendance Rate	
1^{st}-year GPA is Below 7	0.31***	0.13***	0.00
	(0.04)	(0.03)	(0.01)
		Grade (Standardized)	
1^{st}-year GPA is Below 7	-0.18*	0.04	-0.03
	(0.11)	(0.10)	(0.11)
		Passes Course	
1^{st}-year GPA is Below 7	-0.07	0.01	-0.03
	(0.05)	(0.05)	(0.04)
Course Type	*7+ vol*	*7+ enc*	*7+ for*
Observations	927	1424	1234

Notes:
1. The outcome variable is displayed at the top of each panel. Attendance Rate is the percentage of tutorials attended. Passes Course is a binary variable where pass=1 and fail=0.
2. Course type refers to how individual courses dealt with above-7 students. *7+ vol* means above-7 students had full discretion over their attendance. *7+ enc* means above-7 students were strongly encouraged to attend. *7+ for* means that above and below-7 students were penalized for being absent, effectively forcing both groups to attend.
3. The regressions include a first-order polynomial which is interacted with the treatment. The bandwidth is 0.365 and the kernel is triangular.
4. Standard errors are clustered on the student and in parentheses.
5. Significance levels: $^{*} < 10\%$ $^{**} < 5\%$ $^{***} < 1\%$.

Table 5: **Unstandardized Grades Above and Below 7, both Before and After the Abolition.**

	First-year GPA		
Cohort	[6.9 – 7.0)		[7.0 – 7.1]
2009 - 2013	6.40	$p = 0.004^{***}$	6.77
	$(N = 161)$		$(N = 146)$
	$p = 0.126$		$p = 0.487$
2014	6.75	$p = 0.651$	6.88
	$(N = 38)$		$(N = 61)$

Notes:
1. Local averages of unstandardized grades for a bandwidth of 0.1. The number of observations used to calculate the averages are displayed in parentheses.
2. Averages are for the *7⁺vol* courses only, which are the courses where above-7 students had full discretion over their attendance during the policy.
3. The p-values refer to two-sided significance tests for the difference means.
4. Significance levels: $^{*} < 10\%$ $^{**} < 5\%$ $^{***} < 1\%$.

Table 6: **Heterogeneous Effects by Distance and First-Year Attendance.**

	Attendance Rate			Grade (Standardized)			Passes Course		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1^{st}-year GPA is Below 7	0.34***	0.14***	0.00	-0.19*	0.03	-0.02	-0.08*	0.01	-0.03
	(0.04)	(0.03)	(0.01)	(0.10)	(0.10)	(0.11)	(0.04)	(0.05)	(0.04)
Distance to University (Standardized)	-0.05**	-0.05***	-0.01	0.04	-0.06	-0.00	0.02	-0.01	-0.00
	(0.02)	(0.02)	(0.01)	(0.04)	(0.07)	(0.04)	(0.02)	(0.03)	(0.01)
Distance×Treatment	0.06**	0.05***	-0.01	-0.02	0.09	-0.02	0.00	0.02	-0.00
	(0.02)	(0.02)	(0.02)	(0.06)	(0.08)	(0.07)	(0.02)	(0.04)	(0.02)
Attendance in First Year (Standardized)	0.15***	0.07***	0.02***	-0.05	-0.05	0.09	-0.02	-0.01	0.03*
	(0.02)	(0.02)	(0.01)	(0.06)	(0.04)	(0.05)	(0.02)	(0.02)	(0.02)
Attendance in First Year × Treatment	-0.13***	-0.04**	-0.00	-0.01	0.02	-0.01	0.01	0.02	-0.02
	(0.02)	(0.02)	(0.01)	(0.08)	(0.06)	(0.07)	(0.03)	(0.02)	(0.03)
Course Type	*7+vol*	*7+enc*	*7+for*	*7+vol*	*7+enc*	*7+for*	*7+vol*	*7+enc*	*7+for*
Observations	927	1424	1234	927	1424	1234	927	1424	1234

Notes:
1. Attendance rate is the percentage of tutorials attended.
2. Course type refers to how individual courses dealt with above-7 students. *7+vol* means above-7 students had full discretion over their attendance. *7+enc* means above-7 students were strongly encouraged to attend. *7+for* means that above and below-7 students were penalized for being absent, effectively forcing both groups to attend.
3. The regressions include a first-order polynomial which is interacted with the treatment. The bandwidth is 0.365 and the kernel is triangular.
4. Distance and attendance in first year are standardized, where the standard deviations are 30.3 kilometers for distance and 0.07 for attendance (on a scale from 0 to 1).
5. Standard errors are clustered on the student and in parentheses.
6. Significance levels: $^{*} < 10\%$ $^{**} < 5\%$ $^{***} < 1\%$.

Table 7: **Lecture Attendance and Total Study Time.**

	(1)	(2)	(3)
		Attended Lectures	
1^{st}-year GPA is Below 7	0.25	0.08	-0.05
	(0.17)	(0.09)	(0.07)
Intercept	0.59***	0.87***	0.95***
	(0.13)	(0.07)	(0.04)
Observations	170	292	272
		Total Study Time	
1^{st}-year GPA is Below 7	1.98	4.54	2.12
	(3.53)	(3.71)	(3.40)
Intercept	11.00***	15.13***	13.44***
	(2.54)	(1.97)	(2.09)
Observations	170	292	272
Course Type	*7⁺ vol*	*7⁺ enc*	*7⁺ for*

Notes:
1. Attended Lectures is a binary variable based on the answer to "Have you attended lectures?". Total Study Time is an ordinal variable based on the answer to "Average study time (hours) for this course per week (lectures+tutorials+self study)?" The maximum for the interval was used to convert the categories into hours.
2. The regressions include a first-order polynomial which is interacted with the treatment. The bandwidth is 0.365 and the kernel is triangular.
3. Intercepts approximate the outcome mean near the threshold of students right of seven.
4. Standard errors are clustered on the student and are in parentheses.
5. Significance levels: $^{*} < 10\%$ $^{**} < 5\%$ $^{***} < 1\%$.

Table 8: **Peer Exposure and Peer Effects.**

	(1)	(2)	(3)	(4)	(5)	(6)
			Exposure to Forced Peers			
1^{st}-year GPA is Below 7	0.24***	0.02	0.12***	0.03**	-0.01	-0.01
	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)	(0.03)
Attendance Rate		0.70***		0.68***		0.46***
		(0.02)		(0.01)		(0.05)
Mean Dep. Var.	0.56	0.56	0.56	0.56	0.61	0.61
Observations	926	926	1421	1421	1231	1231
			Grades (Standardized)			
1^{st}-year GPA is Below 7	-0.18*	-0.17	0.03	0.03	-0.03	-0.04
	(0.11)	(0.11)	(0.11)	(0.10)	(0.11)	(0.11)
Peer Average 1^{st}-year GPA	0.01		-0.03		0.06*	
	(0.04)		(0.04)		(0.03)	
Peer Avg. GPA × Treatment	-0.04		0.06		-0.00	
	(0.06)		(0.06)		(0.05)	
Peer Average Registration Time		-0.01		0.01		0.01
		(0.01)		(0.01)		(0.01)
Peer Avg. Registration Time × Treatment		0.00		-0.01		-0.01
		(0.01)		(0.01)		(0.01)
Observations	927	927	1424	1424	1234	1234
Course Type	*7⁺vol*	*7⁺vol*	*7⁺enc*	*7⁺enc*	*7⁺for*	*7⁺for*

Notes:
1. The exposure variable is missing if nobody within a tutorial group attended any of the sessions. This explains the slightly fewer number of observations compared to the baseline regressions by course type (compared to *e.g.* the bottom panel).
2. The regressions include a first-order polynomial which is interacted with the treatment. The bandwidth is 0.365 and the kernel is triangular.
3. Attendance Rate refers to the percentage of tutorials attended (top panel).
4. Peer group averages are leave-out means (bottom panel). Peer average first-year GPA is standardized with mean 0 and standard deviation 1 and average peer tutorial registration time is measured in differences in days from the course mean registration time.
5. Standard errors are clustered on the student and in parentheses.
6. Significance levels: $^{*} < 10\%$ $^{**} < 5\%$ $^{***} < 1\%$.

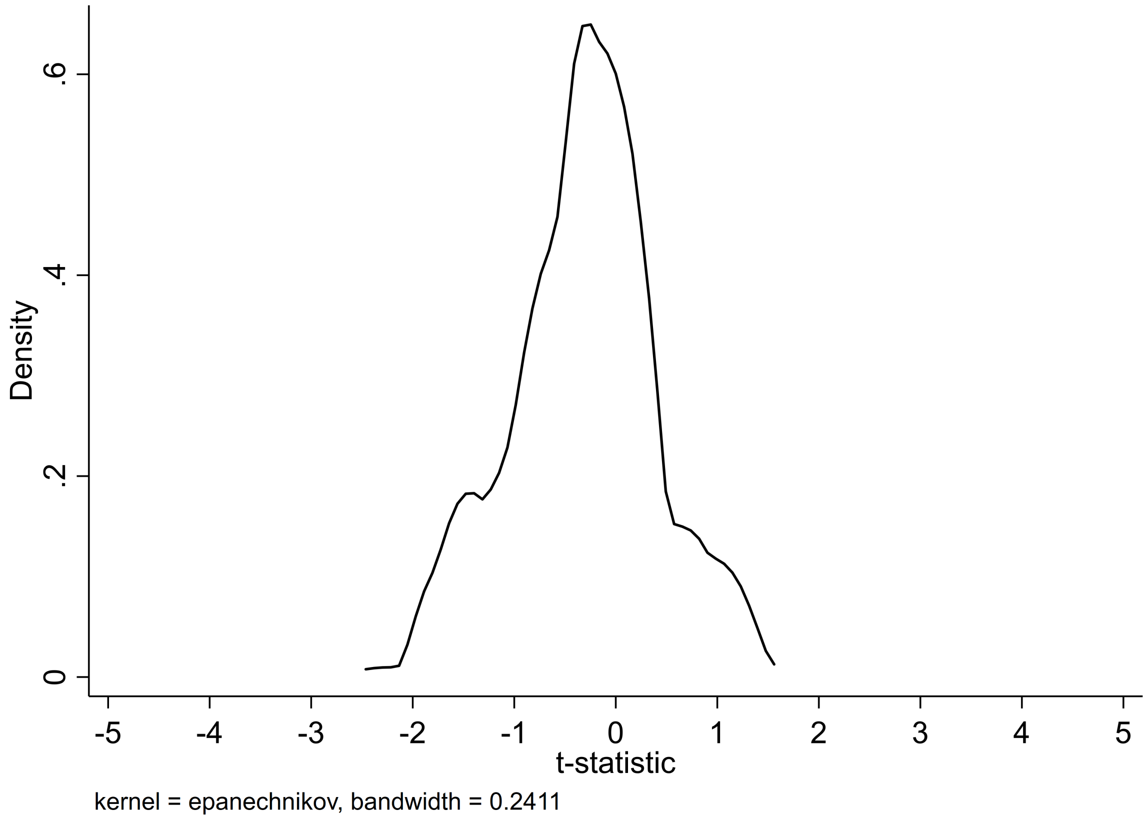
The Price of Forced Attendance

Online Appendix

Sacha Kapoor Matthijs Oosterveen Dinand Webbink

January 13, 2020

Figure A.1: **Additional Balancing Tests with High School Grades.**

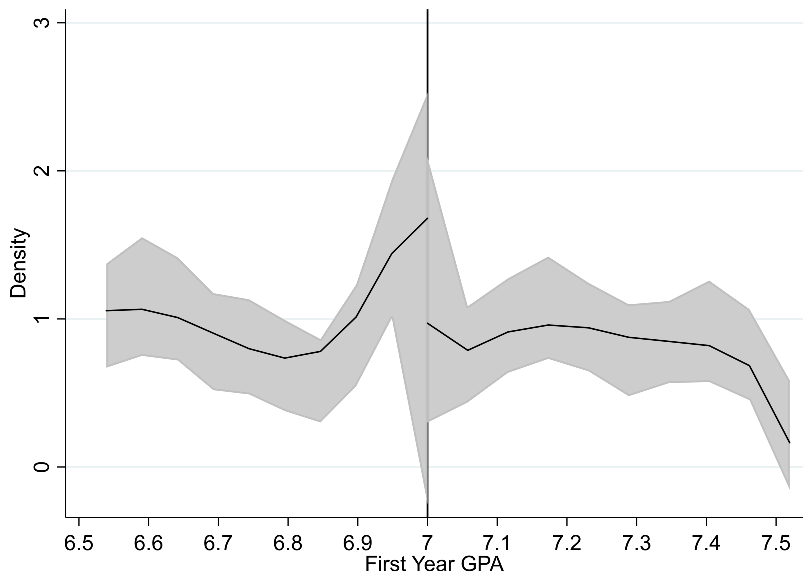


Notes:

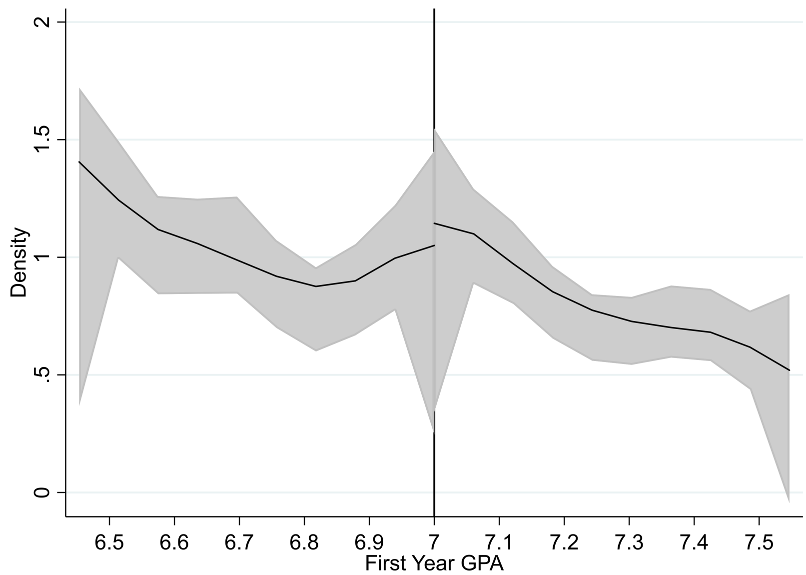
1. Each t-statistic is from a balancing test where the dependent variable is the high school grade for a particular subject.
2. The balancing-test regressions include a first-order polynomial which is interacted with the treatment. The bandwidth is 0.365 and the kernel is triangular.
3. The data includes 133 subjects. Our regressions use grades from 44 of these subjects because for several subjects the number of observations was insufficient.
4. The density estimates are weighted by the number of students taking the subject in high school. The average number of students with a grade in these subjects is 111.

Figure A.2: **Discontinuity in Density Test for Females and Males.**

(a) Females



(b) Males

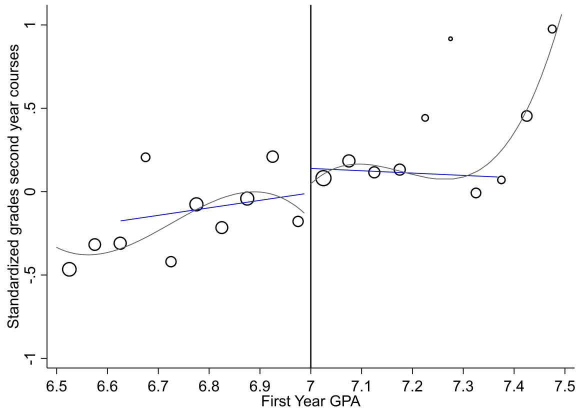


Notes:

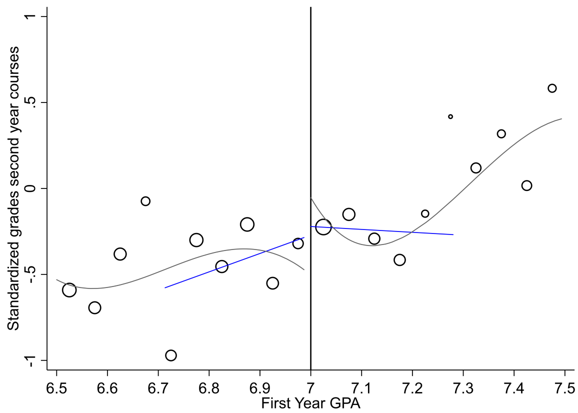
1. Density is for the number of students.
2. Figures plot unrestricted manipulation test, where under the null hypothesis the limiting densities of the number of students to the left and right of 7 are the same. Test is unrestricted in that the estimates of densities to left and right of 7 are unrelated.
3. Figures use a second-order polynomial for density estimation and a third-order polynomial for the bias-correction estimate (see [Cattaneo, Jansson, and Ma, 2018, 2019]). Kernel is triangular. Confidence intervals use jackknifed standard errors.
4. The bias corrected discontinuity test statistic and p-value for females are -1.28 and 0.20. The analogs for males are 0.03 and 0.98. The statistics imply that in both cases we cannot reject the null hypothesis of no discontinuity around the cutoff.

Figure A.3: **Second Year Grades, by Course Type, in the Abolition Year.**

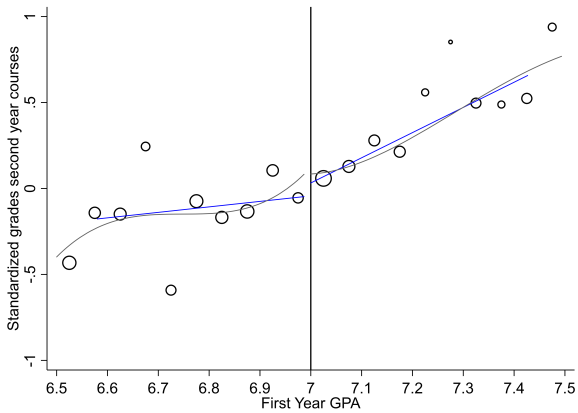
(a) Attendance is Forced Left of 7, Voluntary to Right



(b) Attendance is Forced Left of 7, Strongly Encouraged to Right



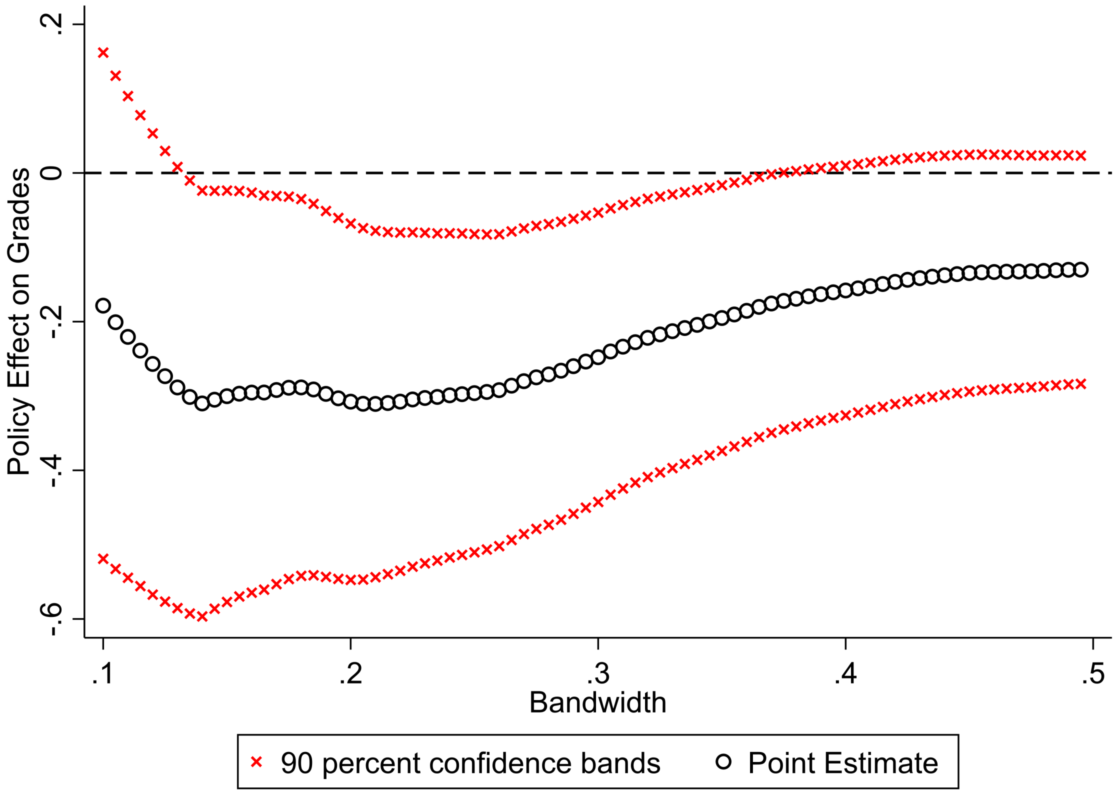
(c) Attendance is Forced Left and Right of 7



Notes:

1. Grades after the university policy was abolished (2014-15).
2. Locally linear and cubic scatterplots for second-year grades against first-year GPA.
3. The local linear polynomial is estimated upon the optimal bandwidth for each outcome relative to a MSE criterion [Calonico, Cattaneo, Farrell, and Titiunik, 2017]. The cubic polynomial is estimated upon a bandwidth of 0.5, which is the same across all figures.
4. Dots are based on local averages for a binsize of 0.05. Dot sizes reflect the number of observations used to calculate the average.
5. Binsizes for local averages are selected via F-tests from regressions of second-year grades on K bin dummies and $2K$ bin dummies for the first-year GPA.

Figure A.4: **Sensitivity to Bandwidth for Attendance-Voluntary Courses.**



Notes:

1. This figure shows the policy estimate and its confidence interval against different bandwidths for *7⁺vol* courses only; the courses where above-7 students had full discretion over their attendance.
2. The regressions include a first-order polynomial which is interacted with the treatment. The kernel is triangular.
3. Bandwidth ranges from 0.1 until 0.5.

Table A.1: **Overview of Program.**

Group	First Year Courses	Second Year Courses
A	Microeconomics Macroeconomics Organisation and Strategy	Applied Microeconomics International Economics History of Economic Thought
B	Financial Information Systems Marketing Financial Accounting	Intermediate Accounting Behavioral Economics Finance I
C	Mathematics I Mathematics II Applied Statistics I ICT	Methods & Techniques Research Project Applied Statistics II Economics of Ageing (Eng) or Fiscal Economics (Dutch)

Notes:
1. Economics of Ageing is taught in the English program and is replaced by Fiscal Economics in the Dutch program.
2. Students can compensate one insufficient grade (between a 4.5 and 5.4) with grades from other courses in the same group if: the other grades are sufficient (above 5.5) and the (weighted) average within the group is above 5.5. This applies to all students, whether they are above or below the threshold of the forced attendance policy.

Table A.2: **Attendance Policies of Second-Year Courses.**

Course	ECTS	Tutorials	Policy	Years	Tutorial Description	Exam Qs.	Block
International Economics	8	Yes	7^+*enc*	2009/13	Students explicitly told to attend 10 of 13 tutorials. Discussion of exercises that are hand in before tutorial. No direct influence on final grade.	MC	1
Ageing or Fiscal Economics	4	Yes	7^+*for*	2010/13	Economics of Ageing: Exercises + Presentations, Accounts for (roughly) 30 percent of their final grade. Fiscal Economics: Exercises, Accounts for 25 percent of final grade. Absence implies a 0 out of respectively 30 and 25.	MC	1
Finance I	8	Yes	7^+*enc*	2009/13	Exercises, Outside tutorials there are weekly quizzes that account for 20 percent of final grade.	MC	2
Applied Statistics II	4	Yes	7^+*for*	2009/13	Exercises, Accounts for 15 percent of final grade. Absence implies a 0 out of 15.	Open	2
Applied Microeconomics	8	Yes	7^+*enc*	2009/13	Draws on tutorial exercises for two interim tests which account for 20 percent of the final grade.	MC	3
History of Economic Thought	4	No		2009/13	Group and individual research projects.		3
Methods & Techniques	8	Yes	7^+*for*	2009/13	Exercises in Computer Lab, Accounts for 5 percent of final grade. Absence implies a 0 out of 5.	MC	4
Behavioral Economics	4	Yes	7^+*vol*	2010/13	Exercises, Actual Experiments, No direct influence on final grade.	MC	4
Intermediate Accounting	8	Yes	7^+*vol*	2009/13	Exercises, No direct influence on final grade.	MC	5
Research Project	4	No		2009/13	Group research projects.		5

Notes:
1. The Tutorial Description is extracted from course guides.
2. Following the Tutorial Description, the Policy column summarizes how the course treated the above-7 students. The abbreviations are discussed in more detail in the main text: 7^+*vol* indicates attendance was voluntary for above-7 students, 7^+*enc* indicates that attendance was strongly encouraged for above-7 students, and 7^+*for* indicates absence is penalized for below and above-7 students, effectively forcing both groups to attend.

Table A.3: **No Sample Selection when Matching Grades with Attendance.**

	Grade (Standardized)		Matched
	(1)	(2)	(3)
Matched	0.02	-0.02	
	(0.05)	(0.07)	
1^{st}-year GPA is Below 7		-0.07	0.01
		(0.13)	(0.01)
Matched×Treatment		0.02	
		(0.11)	
Observations	3873	3873	3873

Notes:
1. Matched is a variable which equals 1 if the grade record found a match with the attendance data and 0 otherwise.
2. Column (1) regresses second-year grades on the matched-variable. The column shows grades are similar for matched and nonmatched records.
3. Column (2) shows no difference in the policy effect between matched and nonmatched records.
4. Column (3) regresses the match-variable on a treatment indicator, showing the policy is unable to explain whether or not a record is matched.
5. The regressions in Column (2) and (3) include a first-order polynomial which is interacted with the treatment. The bandwidth is 0.365 and the kernel is triangular.
6. Standard errors are clustered on the student and in parentheses.
7. Significance levels: $^{*} < 10\%$ $^{**} < 5\%$ $^{***} < 1\%$.

Table A.4: **Overview of Categories and Questions in Course Evaluations.**

Question	Measurement scale	Category
Objectives of course are clear	1-5	General
Course is relevant for my studies	1-5	General
Course is interesting	1-5	General
Course is well organized	1-5	Structure
Course material is understandable	1-5	Structure
Can be completed within allocated study points	1-5	Fairness
Time needed to complete exam is enough	1-5	Fairness
Exam reflects course content	1-5	Fairness
Exam questions are clearly defined	1-5	Fairness
Total study time (lectures+tutorials+self study)	1-10	Total study time
Have you attended lectures?	0-1	Lecture attendance
Lectures are useful	1-5	Lectures useful
Lecturer is competent	1-5	Quality lecturer(s)
Lecturer makes you enthusiastic	1-5	Quality lecturer(s)
TA gives good tutorials	1-5	Quality TA
TA provides sufficient assistance	1-5	Quality TA

Notes:
1. Most questions are measured on a 5-point scale, where 1 equals strongly disagree and 5 equals strongly agree.
2. Total study time is measured on a 10-point scale, where 1 is 0 hours, 2 is $[1-5]$ hours, 3 is $[6-10]$ hours, and 10 is ≥ 40 hours.
3. Lecture attendance equals 1 if yes and 0 if no.

Table A.5: **Balancing Tests around the Cutoff with MSE optimal Bandwith for each Outcome.**

	Distance to Uni. (km)	Age	Gender	European Economic Area	High School GPA
	(1)	(2)	(3)	(4)	(5)
1^{st}-year GPA is Below 7	3.062	0.224	0.211**	-0.026	-0.383
	(5.799)	(0.167)	(0.098)	(0.049)	(0.270)
Bandwidth	0.41	0.44	0.27	0.40	0.46
Observations	585	643	381	564	554

Notes:
1. Unit of observation is the student. The outcome variable is displayed at the top of each column. The outcome variables are not standardized, their means can be found in Table 2.
2. The regressions include a first-order polynomial which is interacted with the treatment. The bandwidth is MSE optimal for each outcome variable. The kernel is triangular.
3. Standard errors are clustered on the student and in parentheses.
4. Significance levels: $^{*} < 10\%$ $^{**} < 5\%$ $^{***} < 1\%$.

Table A.6: **Policy Effects when Forced Attendance was Abolished.**

	Grade (Standardized)		
	(1)	(2)	(3)
1^{st}-year GPA is Below 7	-0.11	-0.18	-0.00
	(0.27)	(0.21)	(0.18)
Course Type	*7⁺ vol*	*7⁺ enc*	*7⁺ for*
Observations	279	430	425

Notes:
1. Sample is restricted to the cohort for which forced attendance was abolished.
2. The regressions include a first-order polynomial which is interacted with the treatment. The bandwidth is 0.365 and the kernel is triangular.
3. Standard errors are clustered on the student and in parentheses.
4. Significance levels: $^{*} < 10\%$ $^{**} < 5\%$ $^{***} < 1\%$.

Table A.7: **Sample Attrition.**

	Number of Courses	Course Evaluation Completed (1=yes, 0=no)		
	(1)	(2)	(3)	(4)
1^{st}-year GPA is Below 7	0.13	-0.03	-0.07	-0.05
	(0.21)	(0.05)	(0.06)	(0.06)
Intercept	9.17***	0.15***	0.19***	0.20***
	(0.15)	(0.04)	(0.04)	(0.04)
Course Type	-	7^{+} *vol*	7^{+} *enc*	7^{+} *for*
Observations	524	927	1424	1234

Notes:
1. Unit of observation is the student in Column (1). It is the student-course combination in Columns (2) to (4).
2. The regressions include a first-order polynomial which is interacted with the treatment. The bandwidth is 0.365 and the kernel is triangular.
3. Intercepts approximate the outcome mean near the threshold of students right of seven. For Column (1) this shows that students, forced or otherwise, complete (more than) 9 out of 10 courses.
4. Standard errors are robust or clustered on the student and in parentheses.
5. Significance levels: $^{*} < 10\%$ $^{**} < 5\%$ $^{***} < 1\%$.

Table A.8: **Student Outcomes by Course Type, with Main Control Variables.**

	(1)	(2)	(3)
		Attendance Rate	
1^{st}-year GPA is Below 7	0.30***	0.12***	0.00
	(0.04)	(0.03)	(0.01)
		Grade (Standardized)	
1^{st}-year GPA is Below 7	-0.16*	0.08	-0.02
	(0.10)	(0.10)	(0.10)
		Passes Course	
1^{st}-year GPA is Below 7	-0.07	0.02	-0.03
	(0.04)	(0.05)	(0.03)
Course Type	*7+ vol*	*7+ enc*	*7+ for*
Observations	927	1424	1234

Notes:
1. Main control variables are course-cohort fixed effects, distance to the university, age, gender, and European economic area.
2. Attendance Rate is the percentage of tutorials attended. Passes Courses is a binary variable where pass=1 and fail=0.
3. The regressions include a first-order polynomial which is interacted with the treatment. The bandwidth is 0.365 and the kernel is triangular.
4. Standard errors are clustered on the student and in parentheses.
5. Significance levels: $^{*} < 10\%$ $^{**} < 5\%$ $^{***} < 1\%$.

Table A.9: **Student Outcomes by Course Type, with Main Control Variables and High School GPA.**

	(1)	(2)	(3)
		Attendance Rate	
1^{st}-year GPA	0.28***	0.13***	-0.00
is Below 7	(0.04)	(0.04)	(0.01)
		Grade (Standardized)	
1^{st}-year GPA	-0.17*	0.05	-0.02
is Below 7	(0.10)	(0.12)	(0.12)
		Passes Course	
1^{st}-year GPA	-0.07	0.03	-0.00
	(0.05)	(0.06)	(0.04)
Course Type	7^{+} *vol*	7^{+} *enc*	7^{+} *for*
Observations	762	1166	990

Notes:
1. High school GPA is observed for Dutch students only, which explains the fewer number of observations compared to the baseline regressions by course type.
2. Main control variables are course-cohort fixed effects, distance to the university, age, gender, and European economic area. These regressions additionally control for high school GPA.
3. Attendance Rate is the percentage of tutorials attended. Passes Courses is a binary variable where pass=1 and fail=0.
4. The regressions include a first-order polynomial which is interacted with the treatment. The bandwidth is 0.365 and the kernel is triangular.
5. Standard errors are clustered on the student and in parentheses.
6. Significance levels: $^{*} < 10\%$ $^{**} < 5\%$ $^{***} < 1\%$.

Table A.10: **Student outcomes by Course Type, with MSE Optimal Bandwidth and Robust Bias-Corrected Inference.**

	(1)	(2)	(3)
		Attendance Rate	
1^{st}-year GPA is Below 7	0.30***	0.13***	0.00
	[0.00]	[0.00]	[0.91]
MSE RD Bandwidth	0.43	0.41	0.39
CER RD Bandwidth	0.30	0.28	0.27
Observations	1125	1569	1310
		Grade (Standardized)	
1^{st}-year GPA is Below 7	-0.26**	0.03	-0.02
	[0.03]	[0.84]	[0.79]
MSE RD Bandwidth	0.29	0.41	0.40
CER RD Bandwidth	0.20	0.28	0.28
Observations	724	1598	1350
		Passes Course	
1^{st}-year GPA is Below 7	-0.07	0.01	-0.07*
	[0.16]	[0.92]	[0.07]
MSE RD Bandwidth	0.40	0.50	0.26
CER RD Bandwidth	0.28	0.35	0.18
Observations	1020	1965	906
Course Type	7^+ *vol*	7^+ *enc*	7^+ *for*

Notes:
1. Attendance Rate is the percentage of tutorials attended. Passes Courses is a binary variable where pass=1 and fail=0.
2. The regressions include a first-order polynomial which is interacted with the treatment. The bandwidth is MSE optimal for each regression, *i.e.*, for each course-type and outcome-variable combination. The kernel is triangular.
4. Standard errors are robust, bias-corrected, and clustered on the student. p-values in squared parentheses.
5. Significance levels: * < 10% ** < 5% *** < 1%.

Table A.11: **Fake Cutoffs for Attendance-Voluntary Courses.**

	Fake Cutoff at			
	6	8	8.25	9
	Attendance Rate			
1^{st}-year GPA	0.01	0.08	-0.01	-0.15
Below Fake Cutoff	(0.02)	(0.09)	(0.10)	(0.44)
	Grade (Standardized)			
1^{st}-year GPA	0.13	-0.24	-0.00	0.18
Below Fake Cutoff	(0.23)	(0.17)	(0.14)	(0.35)
	Passes Course			
1^{st}-year GPA	0.13	-0.01	0.01	0.00
Below Fake Cutoff	(0.14)	(0.03)	(0.01)	(0.00)
Observations	463	339	273	54

Notes:
1. The top of each column indicates at which first-year GPA we set the fake cutoff. Subsequently each column estimates the "policy effect" at that cutoff, for each outcome separately.
2. Sample is restricted to *7⁺vol* courses only; the courses where above-7 students had full discretion over their attendance.
3. Attendance Rate is the percentage of tutorials attended. Passes Courses is a binary variable where pass=1 and fail=0.
4. Main control variables are included: course-cohort fixed effects, distance to the university, age, gender, and European economic area.
5. The regressions include a first-order polynomial which is interacted with the treatment. The bandwidth is 0.365 and the kernel is triangular.
6. Standard errors are clustered on the student and in parentheses.
7. Significance levels: $^{*} < 10\%$ $^{**} < 5\%$ $^{***} < 1\%$.

Table A.12: **TA and Lecturer Quality by Course Type.**

	Teaching Assistant		Lecturer	
	Gives Good Tutorials	Provides Sufficient Assistance	Competent	Makes You Enthusiastic
	(1)	(2)	(3)	(4)
7⁺vol Course	0.21*	0.11	0.00	-0.02
	(0.12)	(0.14)	(0.09)	(0.10)
7⁺for Course	0.26***	0.31***	0.09	-0.08
	(0.09)	(0.10)	(0.06)	(0.07)
Intercept	3.95***	3.96***	3.95***	3.65***
	(0.07)	(0.07)	(0.05)	(0.06)
Observations	503	458	470	469
p-value: $7^+vol = 7^+for$	0.72	0.15	0.21	0.53

Notes:
1. Sample is from year when forced attendance was abolished.
2. Course type refers to how individual courses dealt with above-7 students during the years of the policy. 7^+vol indicates that above-7 students had full discretion over their attendance. 7^+enc indicates that above-7 students were strongly encouraged to attend. 7^+for indicates that above and below-7 students were penalized for being absent, effectively both groups were forced to attend in these courses.
3. The outcome variable is displayed at the top of each column. The questions are measured on a 5-point scale, where 1 is strongly disagree and 5 is strongly agree. See Appendix Table A.4 for more detailed definitions on the dependent variables.
4. Standard errors are clustered on the student and in parentheses.
5. Significance levels: $^* < 10\%$ $^{**} < 5\%$ $^{***} < 1\%$.