The Price of Forced Attendance

Matthijs Oosterveen* Sacha Kapoor† Dinand Webbink‡

April 15, 2017

Abstract

Forced attendance policies assume attendance is valuable and that students are incapable of taking the right attendance decisions. To investigate whether this is the case, we estimate the causal impact of forced attendance on student performance. We draw on a natural experiment at a large European University, where second-year tutorial attendance was mandatory if the student failed to earn a GPA of 7 (on a 10-point scale) in their first year. Using the discontinuity at 7, we show forced attendance decreases grades by 0.35 of a standard deviation. We rule out stigma, teacher or peer quality, and externalities to other courses as explanations for the decrease in performance, implying that the policy operates via its influence on attendance only. Our IV estimates imply in turn that a 10 percentage point increase in attendance decreases grades by 0.1 of a standard deviation. We show further that forced attendance induces students to spend less time studying on their own, implying that it lowers grades because students treat tutorials as a substitute for self study. The increase in attendance was largest for younger students and students living far from campus who, in addition, are most likely to use classes as a substitute for self study. We conclude forced attendance is most harmful to the students it, in principle, purports to help.

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

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

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

§We would like to thank Robert Dur,... for helpful comments and suggestions. Seminar participants at EEA-ESEM 2016, Erasmus University Rotterdam, IZA Workshop on the Economics of Education and the Tinbergen Institute are also gratefully acknowledged for their feedback.

For many people their first real encounter with autonomy happens at college or university. Out from under the roofs of their parents and high school teachers, how they manage their lives is now largely up to them. Many students use their 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, watching March Madness, or chasing other young men and women. To combat the rampant absenteeism this newfound autonomy begets,[1] and because of the substantial returns to college performance and graduation [Cunha, Karahan, and Soares, 2011, Jones and Jackson, 1990], university administrators and instructors often mandate that students attend their classes.[2] These forced attendance policies presume that attendance is a valuable input into academic performance and, importantly, that young men and women are generally unable to make good decisions for themselves. If the presumption is correct then forced attendance will be good for academic performance. At the same time, however, the policies violate the most basic principles of economic analysis. Forced attendance constrains time allocations and, by doing so, precludes rational students from allocations that best serve their own self interest. If the basic principles of economic analysis are correct then forced attendance will be bad for academic performance.

We estimate the causal impact of forced attendance and show that it is bad for academic performance. We draw on a natural experiment at a large European University, where between 2009 and 2013 the student's first year grade determined whether they were required to attend tutorials in their second year. Students who average less than 7 (out of 10) in first year must attend 70 percent of tutorials in each of their second year courses. Students who fail to meet the attendance requirement face a severe penalty, as they are not allowed to write the final exam for their course, and must wait a full academic year before taking the course again. The policy imposes heavy time costs on students, some of whom can expect to spend 250 additional hours traveling and attending tutorials over a full academic year. Because students have imprecise control over their *average* grade in the first year, the experiment facilitates a regression discontinuity design [Lee, 2008, Lee and Lemieux, 2010] for identifying the various impacts of forced attendance.

What does it mean to be "forced"? In what way are students "forced"? Our working definition is that a person is forced if an authority figure takes away some of their potential choices. Or, a bit more formally, if the authority imposes a heavy, sometimes infinite (death), penalty on a particular choice.[3] The policy we study is well within confines of

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

[2] See, for example, the early discussion on mandatory attendance in the correspondence section of the Journal of Economic Perspectives in 1994 [Correspondence, 1994].

[3] 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

this economic definition.[4] The policy forces students to come to campus, a choice which is normally fully under the purview of the student, and imposes a heavy penalty when they fail to do so. On top of all this, the policy was communicated to students as something they had to do. Our data strongly supports the notion that students were forced, as less than one half of one percent of students were in violation of the policy. In other words, a more severe penalty, such as death, would have increased attendance by less than one half of one percent.

We complement the experiment with rich administrative data that allows us to go beyond a simple evaluation of the impact of forced attendance on student performance. The data includes information on the actual attendance of the student. Attendance data allows us to adjust the impact on grades by attendance and, in doing so, draw inferences about the impacts of marginal, and more moderate, increases in forced attendance. The data includes information on time use outside of tutorials, which lets us evaluate whether students treat tutorial attendance as a substitute for or complement to the time they spend studying on their own. The data also includes information on the personal characteristics of the students, such as their home address and their age. Personal data lets us identify the students who are most affected by the forced attendance policy, and evaluate how the elasticity of substitution between tutorial attendance and self study differs depending on their personal characteristics. Taken together, the natural experiment and data facilitate answers to whether, why, and how forced attendance affects student performance.

We find that the forced attendance policy decreases grades by 0.35 of a standard deviation. Rescaling the point estimate (using the policy as an instrument) by the student's attendance reveals that a 10 percentage point increase in attendance decreases grades by 0.1 of a standard deviation for the forced student. We show further that while tutorial attendance improves by 35 percentage points over a baseline of 55 percent, students spend less time studying on their own. These results imply forced attendance decreases performance because the average student treats tutorial attendance as a substitute for self study.

Our data lets us rule out mechanisms other than the substitutability of attendance for self study. Courses differed how closely they followed the university policy: some followed it to a tee; others layered their own more stringent policy on top of the university policy such that attendance does not jump near 7; the remainder were more lackadaisical

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.

[4]Our definition differs from the notion of labor coercion, which focuses on how physical force or the threat of physical force influences institutions and outcomes in labor markets [Acemoglu and Wolitzky, 2011].

regarding its enforcement. Because in our sample the same students had to take all these courses, we can plausibly test for direct effects of the policy on academic performance. Bartling, Fehr, and Herz [2014] show that individuals intrinsically value decision rights and autonomy, suggesting the policy might have directly caused a utility loss for students left of 7. In a similar vein, the policy might have affected personal well-being or stigmatization of the student, or general discontent with the policy that might caused the student to exert less effort on their studies. Our results are inconsistent with these direct channels. Indeed we find that university policy had no impact in courses that did not follow it exactly. For these courses both attendance and grades do not significantly differ near 7. Our data lets us rule out other mechanisms as well, including poor teaching quality, negative peer effects, and spillovers from one course to the next. We also find no effects of the policy on outcomes in the third year of the student: thesis, course grades, and major choice. The negative impact of the policy is not reflected in the long term performance of the student. In all our results support one mechanism, namely that the impact operated through tutorial attendance and that the policy decreases performance because attendance is a substitute for self study. As such, our findings are consistent with the exclusion restriction, generating additional interest in our IV estimates,

We find the attendance of younger students and students who live far from campus are most affected by the policy. We show further that younger students have a higher elasticity of substitution between tutorial attendance and self study. They are less likely to use self study to compensate for the additional time they spend on tutorials. The larger impact on younger students, who in principle are less able to make good decisions for themselves, implies that forced attendance is most harmful to the students it purports to help.

Our study is of practical and academic relevance. On the practical side, the results should prove useful to university administrators or instructors who put a lot of money and effort into personalized classroom instruction, but experience poor classroom attendance, and who are considering forced attendance policies.[5] To this end the study contributes to a small literature which has studied the impact of mandatory attendance on academic performance [Dobkin, Gil, and Marion, 2010]. Our study advances this literature in a couple of ways. First, (almost) all students comply with the policy because of the heavy penalty for noncompliance. As such, our study really speaks to the impact of *forcing* students to attend classes. Second, while students are well aware of the policy in the first year, the policy design is such that it limits a role for anticipation effects. The policy is based on the *average* grade in the first year. As a student accumulate grades, it becomes

[5]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.

more difficult for them to precisely control their average grade over all their courses. Their lack of control generates plausibly exogenous variation in who is assigned to the forced attendance policy.[6] Third, our study is large in scale, as it lasted 5 years and was applied to all students enrolled during this time.

The study also contributes to a literature that examines the impact of class attendance on academic performance. Most studies analyzing this relationship find that better attendance improves performance [Durden and Ellis, 1995, Kirby and McElroy, 2003, Latif and Miles, 2013, Lin and Chen, 2006, Marburger, 2006, Rodgers, 2002, Romer, 1993, Snyder, Lee-Partridge, Jarmoszko, Petkova, and DOnofrio, 2014, Stanca, 2006]. Because attendance is normally a choice, and because it is consequently used to measure noncognitive skills [Jackson, 2012], other studies have searched for credibly exogenous variation in attendance. There the results are more mixed. One has found that better attendance improves performance [Chen and Lin, 2008], while another has found on average no evidence of an effect on performance [Arulampalam, Naylor, and Smith, 2012]. By the same token, we find that better attendance actually decreases performance. Our findings differ from past studies most likely because our estimates identify the causal effect of attendance among forced students.

On the more academic side, our study contributes to two strands of the economics literature. Many recent studies have focused on the role of incentives in higher education, via higher quality instruction like academic support services or better tutors [Angrist, Lang, and Oreopoulos, 2009, Martins and Walker, 2006], or even explicit financial incentives [Leuven, Oosterbeek, and Klaauw, 2010, Visaria, Dehejia, Chao, and Mukhopadhyay, 2016]. While they have their appeal, incentive-driven interventions are expensive to implement, and the advantages are unclear [Fryer Jr, 2016]. We instead focus on a natural yet less expensive alternative, one that constrains rather than incentivizes students, and show it is, in fact, bad for academic performance. Note that compared to the existing literature on compulsory schooling, initiated by Angrist and Krueger [1991], our study targets university students and is about compulsory class attendance within an existing curriculum.

Finally, our study is fruitfully juxtaposed against the influential study of Bloom et al. [2014]. That study investigates the productivity implications of letting workers work from home. Our study investigates the “productivity” implications of forcing students to come to campus. Indeed our results are symmetric. They find that working from home is good for productivity. We find that forcing students to come to campus is bad for productivity. To this end it is worth noting that it would be difficult, if not impossible, to study questions about forced attendance (or forced behavior of any sort)

[6] Our logic is a simple application of the law of large numbers.

via the randomization of students or workers into forced attendance policies.[7]

1 Context and Data

Our venue for examining the impact of forced attendance is the economics undergraduate program at a large university in the Netherlands. The economics program itself is large - in the 2013-2014 academic year alone, the program saw an influx of approximately 800 students. During the first two years of the program, students follow the same ten courses per year. The courses and their descriptions are found Table 1. They cover basic economics, business economics, and econometrics. In their third year, students can choose their own minor and major and subsequently can continue this specialization through to their graduate program. 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.[8] Heavy courses have three large-scale lectures per week, while light courses have two. Each lecture lasts 1 hour and 45 minutes. Lecture attendance is always voluntary. Importantly for our purposes, heavy courses have two small-scale tutorials per week, while light courses have one. Tutorials also last for 1 hour and 45 minutes. Each tutorial has about 30 students. Unlike lectures, students actively participate in tutorials, via *e.g.* discussions of assignments and related materials.

Second year courses each have several tutorials, at different times, and students have flexibility in the tutorial they attend. Students register for tutorials 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- and PhD students. Students cannot switch their tutorial group after the registration period ends. Importantly, all students, including the ones whose first year grade is above 7, must register for a tutorial. It is worth noting that we observe for which group and at which time the student registered. With this information, we can evaluate whether there were systematic differences in registration patterns for forced and free students.

Grading is done a scale that ranges from 1 to 10. Students fail a course if their grade

[7] Most ethics boards would rightfully hesitate to approve such a proposal.

[8] 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.

is below 5.5. The average grade in the first year is weighted by the amount of credits the student gets for completing the course. In our main sample, the first year grade has a mean and standard deviation of 6.99 and 0.70.

1.1. The Policy. In their second year students must attend 70 percent of the tutorials unless they had:

1. an average grade (weighted by course credits) of 7 or more in the first year;
2. completed the whole first year within the first year (passed all ten courses).[9]

The following table summarizes the students who had to comply with the policy.

Completed first year	GPA< 7	GPA ≥ 7
✓	Forced	Free
✗	Forced	Forced

Students were not allowed to write the final exam and had to wait a full year before retaking the course if they failed to fulfill the 70 percent attendance requirement. The policy lasted five years, starting in 2009-2010 and lasting until 2013-2014.

The forced attendance policy was implemented as part of a university initiative to personalize education via small-scale tutorials. The initiative came about for four reasons: (i). the university had grown to a scale that made education impersonal; (ii). tutorials encourage active participation; (iii). tutorials create dialogs between students and the teaching staff; (iv). the tutorials facilitate student involvement in the university community. The forced attendance policy was implemented because the initiative was costly, and because the university wanted to ensure that they were getting a return on the sizeable investment in small-scale tutorials. Importantly, its introduction had nothing to do with the grade distribution of first year students. The forced attendance policy was abolished in 2014-2015 at the request of the student body and faculty, motivations thereof were unclear.

The policy imposes nonneglible time costs on students. Students who score just below 7 in their first year must spend 26 hours per block (3.5 hours per week) in tutorials.[10] Once we account for the travel time of the average student, about 45 minutes each way, forced students must spend 50 hours per block traveling to and attending tutorials.[11]

[9]Courses are grouped, so that a student can compensate a failing grade of between 4.5 and 5.4 from one course in the group with a passing grade from another. Table 1 provides more details on this.

[10]The 3.5 hours per week 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. In this calculation we assume that the forced student would not otherwise attend.

[11]The average student lives 22.9 kilometers from the university campus. To get an idea of the travel

1.2. Data. Our main information source is the administrative data of the university. Our sample ranges from the 2008-2009 academic year until 2014-2015. We observe grades at the level of the student for all three of their undergraduate years, tutorial attendance for the first two years, course evaluations, and various demographic characteristics, including nationality, date of birth, gender, and living address. For Dutch students we also observe their performance in secondary school. After restricting the sample to be with 0.5 grade points of 7, our preferred bandwidth, we have a sample of about 5000 course-student observations, based on more than 700 students.

The university uses attendance lists to track the attendance of students. 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.

Our attendance variable is measured at the student-course level, and is simply the percentage of tutorials attended per course. It measures actual attendance with error if students sign for their peers, something which in principle could be especially problematic among students who are forced to attend. It would also have measurement error if students who attend voluntarily have insufficient incentives to sign in. Both reasons would lead us to underestimate the effects of attendance on performance for the forced student.

We view the measurement error as a minor concern for several reasons. First, our baseline results all come from reduced form estimates of the policy impact on student performance. Actual attendance has no role in these estimates. Second, teaching assistants were tasked with preventing fraudulent sign-ins, as instructors required them to count the number of students present at the tutorial, to make sure that the number coincides with the number of signatures on the attendance list. Third, attendance rates among voluntary students is fairly high. On average these students attend 55 percent of the time.[12]

Our data also includes information from course evaluations. One week before the exam, students are invited by email to anonymously evaluate the course online at the university portal. They are reminded of the evaluations shortly after the exam. All evaluations contain the same set of 21 core questions, which are grouped into the general

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.

[12]While matching attendance with the administrative data, we experienced a match rate of 93 percent (in our main 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 reduced-form effect 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.1 in the Appendix). Therefore we work with this 93 percent of the sample throughout the paper.

opinion of the course, structure, fairness, quality of lecturer and tutor, and usefulness of the lectures. Importantly, for our purposes, students are asked about their attendance at lectures, as well as the time they spend studying overall. Together with the data on tutorial attendance, this information lets us construct a measure of the time students spend studying on their own. It thus lets us evaluate whether and to what extent the forced attendance policy influences trade offs between attendance and self study.[13] The caveat here of course is that the evaluations are only filled out by 20 percent of the students. Later we will show that this poses little to no problems for our empirical analysis.

Our analysis focuses on the sample of students who completed the first year on time. In principle, one could estimate a local difference-in-difference, comparing changes in the grades of these students, around the cutoff, with changes in the grades of students who did not complete the first year. We did not do this because first year completion rates for students around the cutoff is 92 percent.

1.3. Preview of Baseline Results. Table 2 provides a basic summary of the data. The table compares students with an average first-year grade between 6.5 and 7 to students whose average grade was between 7 and 7.5. The unit of observation in the upper panel is the student-course combination. The unit of observation in the bottom panel is the student. Second-year grades are measured in standard deviations.

The top panel shows that on average forced students score 0.4 of a standard deviation worse than their peers. This is despite the fact that they attend tutorials 15 percentage points more of the time. The bottom panel implies students on one side of the cutoff are roughly similar to students on the other. The main difference being that poor performing students are likely to be overrepresented to the left of 7 as visualized by their GPA in secondary school. Accordingly, we account for this in our more flexible regression specifications.

Figure 1 visually examines how attendance and grades change around the cutoff over the entire treatment period for eight of the 10 second year courses,[14] including the ones that did not follow the policy perfectly. Forced students attend approximately 15 percentage points more tutorials than their peers who score just above 7. Their grades are lower by 0.15 of a standard deviation, though the estimate is marginally insignificant at conventional significance levels.[15]

[13] For comprehensive details of the course evaluations see Table A.2 in the Appendix.

[14] Two courses cannot be used as they do not contain tutorials, but instead (mostly) consist out of writing a research report in groups. See Table 3 for a detailed overview on the characteristics of all the ten second-year courses.

[15] The figures contain both the preferred local linear regression and the third order polynomial, estimated on the optimal bandwidths of 0.2 and 0.5 respectively. As expected, second-year performance and first-year GPA exhibit a positive correlation, where moreover the third order polynomial closely follows

We will, in what follows, make use of differences in how the eight courses implemented the forced attendance policy. Two of the eight courses followed the university policy to a tee. Their instructors allow absences from students whose first year grade is at least seven. Three courses followed more elaborate policies where tutorial attendance was also incentivized, as students had to complete assignments that made up five to thirty percent of their final grade.[16] The remaining three courses followed policies that were more relaxed than the university policy. Table 3 provides a detailed overview on the characteristics and group-classification of all the ten courses.

The left panel of Figure 2 examines the impact on attendance for the three types of courses. Courses which followed the university policy strictly saw an increase in attendance of more than 30 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.[17] Courses with a more elaborate policy saw no difference in the attendance of students above and below 7. All students show up because their attendance has a direct impact on their final grade. Attendance in courses with a more relaxed policy only jumped up 12 percentage points, where the effect of the policy is considered weak with F-statistics below 10.

The right panel of Figure 2 examines the impact on grades. In compliant courses grades decreased by 0.35 of a standard deviation. In the other courses the impact on grades is statistically indistinguishable from zero.

Table 4 provides a more comprehensive breakdown of the differences across courses. We use a narrow bandwidth of 0.1 because it helps us to emphasize the stark differences in compliance. Table 4 has two notable features. First, lecture attendance strongly coincides with tutorial attendance. For courses that followed the policy the treatment increased the probability to attend lectures by 20 percentage points over a baseline of 64 percent, where for the courses with a more elaborate policy the probability of attending the lectures is 90 percent, independent of being below or above the 7. This is consistent with the idea that campus visits have large sunk costs. Once there for the tutorials, lecture attendance is relatively cheap. Second, TA quality is ruled out as an additional reason for the difference in compliance between courses, as we observe that the average TA quality is fairly high (scale between 1 and 5) and does not differ between the groups of courses.[18]

the flexible Lowess.

[16]For these three courses, the course guides stated: if an exempted student does not show up at the tutorials, he or she obtains the grade zero for the tutorial part and thereby can only obtain 70 to 95 percent of their final grade by writing the exam.

[17]As there are 21 tutorials per block, this corresponds to ≈ 7.5 extra tutorials and ≈ 13 extra hours of schooling per block.

[18]Note that the panels differ in the number of observations primarily because two of the courses started

We use the abolition of the policy in 2014-2015 to provide strong supplementary evidence that lower grades are a consequence of the forced attendance policy. What is particularly useful in this regard is that the abolition was a surprise, as students only became aware of it *after* their second-year had already started, in the first block of the academic year. Their prior ignorance increases the value of comparisons of grades before and after the abolition of the policy. The left panel of Table 5 compares students who were just below the cutoff, before and after the abolition. The right panel makes a similar comparison, but for students who were just above the cutoff.

The left panel of Table 5 shows that (unstandardized) student grades just below the cutoff decreased by 0.36, which roughly equals -0.31 standard deviations. This estimate of the across-cohort difference is very similar to the within-cohort estimate of -0.36 standard deviations. It too is statistically significant at conventional levels. In addition to providing compelling evidence that it is the forced attendance that decreases grades, Table 5 shows that our estimates are being generated by lower performance of forced students, rather than by better performance of unforced students. This reinforces our confidence in the conclusion that it is the being forced that worsens performance.

2 Empirical Specification

We assume the second-year grade $G_{ijc}^{(2)}$ of student i in course j and cohort c is generated in accordance with

$$G_{ijc}^{(2)} = \beta_0 + \beta_1 D_{ic} + f(\bar{G}_{ic}^{(1)} - 7) + f(\bar{G}_{ic}^{(1)} - 7)D_{ic} + C_{jc}^{(2)} + \mathbf{X}_i\mathbf{\Gamma} + \varepsilon_{ijc}^{(2)} \tag{1}$$

where D_{ic} is a binary variable that equals 1 if the student's first-year GPA is below seven, $\bar{G}_{ic}^{(1)}$ is their average grade in the first year, $C_{jc}^{(2)}$ are course-year fixed effects, and $f(\cdot)$ is some higher order polynomial expansion of $\bar{G}_{ic}^{(1)}$. Our primary interest is the parameter β_1, which measures the impact of forced attendance near the GPA of a 7. The adoption and use of the forced attendance policy suggests $\beta_1 > 0$. Basic decision theory tells us that this constraint on attendance generates a $\beta_1 < 0$.

We can interpret estimates of β_1 causally if [Lee, 2008]:

> **Identifying Assumption:** Students have imprecise control over their average grade in the first year, meaning that conditional on their characteristics, the distribution for average grades is continuous around 7.

While this is generally a weak identifying assumption [Lee, 2008], it seems particularly

in 2010, rather than 2009 (see Table 3 for details).

reasonable in our setting. To see why, let

$$G_{ijc}^{(1)} = e_{ijc}^{(1)} + a_{ijc} + \delta_{jc}^{(1)} + \eta_{ijc}^{(1)}$$

denote the student's grade in first-year course j, where $e_{ijc}^{(1)}$ is their effort, a_{ijc} is their ability, $\delta_{jc}^{(1)}$ is something particular about the course-cohort combination (such as the professor or teaching assistant), and $\eta_{ijc}^{(1)}$ is the idiosyncratic component of the first-year grade. Attendance in second-year tutorials is mandatory for the student if:

$$\bar{e}_{ic}^{(1)} + \bar{a}_{ic} + \bar{\delta}_{c}^{(1)} + \bar{\eta}_{ic}^{(1)} < 7$$

where the bars indicate that the variable is averaged over all first-year courses j.

A student clearly has some control over their first year grade through their ability and effort. What the equation does make concrete is the notion that this control is imprecise. The identifying assumption holds as long as there are some idiosyncratic factors driving the first-year performance of students. Put another way, it will hold if some students experience a death in the family, surprises to their income, or just bad luck across all the exams they wrote that year. Random shocks like these ensure that two students, with similar ability and effort end up on either side of the cutoff. As a result, the (conditional) distribution of first year GPA is continuous and the variation in treatment status will be random in a neighborhood of 7. Note that $\bar{\delta}_{c}^{(1)}$ reflects that randomization near the cutoff takes place for every cohort separately. Our regressions control for this via the inclusion of fixed effects for the course-cohort combination $C_{jc}^{(2)}$.

2.1. Opportunities for Manipulation. Our identifying assumption hinges on the ability for students to precisely manipulate their position around the cutoff. We use attendance and performance data from the first year to look for evidence of such opportunities. We consider naive regressions of the student's grade in each of their first-year courses on their tutorial attendance for these courses. Our goal is to examine whether attendance and a host of other controls explain a lot of the variation in first-year grades. If attendance and controls do explain a lot of variation it would suggest that students can control their first year grade, and that there is little room for $\eta_{ijc}^{(1)}$ to randomly assign students around the cutoff. If the controls explain little of the variation in first-year grades, then we infer that students have imprecise control over the grade they get in the first year.

Estimates are found in Table 6. Moving left to right shows how the correlation changes as we: (1) restrict our sample to regular students who attended more than 70 percent of the classes; (2) and (3) include exceptional cases for which attendance was below the 70 percent requirement;[19] (4) until (8) include various controls. Our main interest is in the estimated R^2 for these regressions. These estimates support the idea that students have imprecise control over the first-year grades. They imply that more than 65 percent of the variation in first-year grades is unaccounted for even after we control for course-year fixed effects, for date of birth, distance to university, gender, nationality, the average grade in the secondary school (for the Dutch sample of students), or student fixed effects. Note that the naive estimates found for attendance are in line with typical estimates of the positive relationship between attendance and student performance found in previous research.

2.2. Continuity Near the Cutoff. Local randomization of the treatment near the cutoff gives us two testable implications: (i). the distributions for observed and unobserved characteristics are identical from one side to the next; (ii). the probability density for GPA is continuous. We evaluate the implications one by one.

Table 7 examines the distributions for observed characteristics, presenting estimates of our main empirical specification Equation (1), where instead of grades the dependent variables are student characteristics. The table presents results for local linear regressions (panel A) and the third order polynomial, both with the optimally chosen bandwidths (panel B) and the full sample (panel C).

Students are very similar to the left and right of the cutoff. They are similar in their nationality, age, distance from the university (in kilometers), and in high school characteristics. Level of secondary education refers to the two different levels that a Dutch high school student might have followed before enrolling at university (easy=0, difficult=1), where track refers to the four different tracks (within a level) available to the Dutch high school student.[20] In our specification with secondary school grades (column (7)) we control for track choice of the student by including a dummy per track. Note that because women are a bit over-represented to the left of the cutoff, we will control for gender when estimating many of our specifications. In addition, Table A.3 in the Appendix tests for differences in grades for various secondary school courses. It shows no significant differences near the cutoff.

We examine whether the probability density for GPA is continuous around 7. Doing

[19] In their first year students were required to attend 70 percent of their tutorials. Similar to the policy we study, failure to meet the 70 percent criteria excluded the student from writing the exam. Some students might have been exempted from the attendance rule for various reasons. These students account for less than one percent of our sample. Moreover, we exclude students with missing attendance or who did not attend at all, as for the latter group their attendance would not have been registered.

[20] We follow Buser et al. [2014] in ordering the four Dutch tracks according to their academic prestigiousness (0=least prestigious, 4=most prestigious). As such, column (6) is a regression with an ordered outcome variable while controlling for the level of secondary education. Estimating an ordered probit does not change our results.

so lets us examine whether students can manipulate their grade in this part of the distribution [McCrary, 2008]. If they can then we would most likely observe bunching just above 7. To check we estimated Equation (1) using normalized counts of the number of students as the dependent variable.[21] The main results are summarized in Figure 3, which shows no evidence of bunching above the threshold. Table A.4 in the Appendix verifies this, showing formally that none of our specifications lead us to reject the null of continuity for average first-year grades near the cutoff.

To be sure we examine data from the 2014-2015 academic year, the year that the forced attendance policy was abolished. This is particularly relevant as the abolition of the policy was only announced at the start of the second year for this cohort of students. That is, during their first year they had an incentive to end up above the 7. Figure 4 depicts the grade distribution for all eight courses (left figure) and for the two courses that previously followed the university policy (right figure). They show that there is no difference in student grades around the cutoff after the policy was abolished.[22]

2.3. Sample Attrition. We investigate whether the policy induced students to drop courses where they failed to meet the 70 percent attendance requirement. Attrition of this sort would compromise a causal interpretation because there are no grades for dropouts. 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 Columns (1) to (3) of Table 8 imply the policy has no effect on the number of completed courses. Our estimates of the intercepts lend further support to this, as they show that near the cutoff students complete almost every course (nine out of ten). Moreover, it suggests every forced student actually complies with the policy, something which we investigate in more depth later in the paper.

Because our main analysis makes use of data from course evaluations, we evaluate whether students above and below threshold differ in their propensity to complete the evaluations. Selection of this sort could arise mechanically, simply because forced students are present more often. Columns (4) to (7) of Table 8 present estimates of the impact of the policy on a binary variable that indicates whether students completed the course evaluation. Columns (4) and (5) show results for all eight courses, while (6) and (7) show results for the two courses that strictly adhered to the university policy. Our estimates of the intercept show, unsurprisingly, that response rates are low on average. They are roughly 20 percent. There is, however, no statistical difference in the propensity to complete the evaluation for students who are just above or below the threshold. As with

[21] To count he number of students we select bin sizes in accordance with the proposed strategy of McCrary [2008]. Either way, our results are robust to the bin size.

[22] We do not have attendance data for this cohort (because the policy was no longer place).

course completion, the results suggest sample selection is not a major problem for course evaluations.

2.4. Estimation and Inference. Our evidence points to the randomization of students around the cutoff of 7, implying that we can estimate the specifications locally. Our two preferred specifications include a linear and third order polynomial for $f(\cdot)$ and a bandwidth of, respectively, 0.2 and 0.5 for $\bar{G}_{ic}^{(1)}$. For all specifications we cluster standard errors at the level of the student.[23]

We elaborate on how $f(\cdot)$ and the bandwidth were chosen. Local randomization implies that we can estimate local linear specifications for $f(\cdot)$, as proposed by Imbens and Lemieux [2008], while selecting the optimal bandwidth (0.2) via the cross-validation method. However, if we are willing to assume a functional form for the relationship between GPA in the first year and second-year grades ($f(\cdot)$), we can use more observations and extrapolate from above and below the cutoff to what a locally-randomized experiment would have shown [Van der Klaauw, 2002]. To select $f(\cdot)$, we estimate Equation (1) while adding equal-sized bin dummies of GPA and including higher-order polynomials until the bin dummies are jointly insignificant.[24] Our preferred specification includes a third order polynomial, as for multiple bandwidth choices we reject joint significance of the bin dummies at conventional levels. The evidence implies the polynomial is a good fit and that other discontinuities are absent. We allow the polynomial to differ across both sides of the cutoff (see the discussion in Lee and Lemieux [2010]). For the choice of our preferred bandwidth (0.5) we make use of the various bandwidth selectors provided by Calonico et al. [2016].[25]

3 Baseline Results

Table 9 presents estimates of the impact of the policy on student grades. The specifications in the top and bottom panel are based on local linear regressions and a third order polynomial, with the optimally chosen bandwidths. Columns (1) and (2) report the impacts for courses that followed the university policy, (3) and (4) the impacts for courses that followed a more elaborate policy, (5) and (6) the impacts for courses that

[23]We do not cluster on the tutorial group because peer composition differs from course to course. However, we show that our results are robust to including tutorial fixed-effects.

[24]We ran various regressions while changing the number of bins, but our preferred specification includes the number of bins (8) for which we first stopped rejecting the small versus the big model while choosing the binsize for the local averages for the RD graphs (see Figure 1).

[25]See Figure A.1 and Table A.5 in the Appendix for more details on the methods and results concerning the optimal bandwidth selection. Note that we use student grades as outcome for both selecting the bandwidth and the order of the polynomials (the reduced form). This seems reasonable as the relationship between attendance and GPA is relatively flat (first stage). We would therefore expect the polynomial to be linear and the optimal bandwidth to be wide.

were more relaxed.[26] Even-numbered columns report regressions with controls for background characteristics, excluding secondary school performance as this limits our sample to Dutch students only.

In courses that followed the university policy, forced attendance decreases grades by 0.34 to 0.43 of a standard deviation. The estimates are all statistically significant at a 1 percent significance level. In all other courses there are no statistical differences between students above and below the cutoff.

Table 10 presents estimates of the policy impact on attendance. Students with a grade of less than 7 in the first year increased their attendance by 30 to 35 percentage points (columns (1) and (2)). As with grades, the impact is much more pronounced in compliant courses. For courses that followed a more elaborate policy there is no impact at all, where for courses that were more relaxed the impact is weak (with F-statistics below 10).

Moderate to negligible impacts in other courses imply that grades are lower because of the increase in attendance, rather than because of other (direct) channels. For instance, recent evidence shows that individuals value decision rights beyond their instrumental value [Bartling, Fehr, and Herz, 2014]. The loss of independence and autonomy for the forced student might have caused a loss in (procedural) utility and happiness. In a similar vein, a stigma attached to, or general discontent with, the policy could have upset students.[27] These channels might have caused students to exert less effort on their studies, and thus worsen their performance overall. The moderate to negligible impacts in other courses casts doubt on this being the case.

Another concern relates to whether forced attendance in compliant courses had negative, but more mechanical, spillover effects onto grades in other courses. Specifically, it is possible that being forced to spend extra time on one course comes at the cost of performance in another course. Mechanical spillovers are suggested by the point estimates in columns (3) and (4) of Table 9, which show that students just below 7 performed a bit worse than their close peers on courses that followed a more elaborate policy. We have good reasons for doubting the importance of such spillovers. First, the point estimates in columns (3) and (4) are statistically insignificant. Moreover, most of the point estimates in columns (5) and (6) are mainly positive, in addition to being statistically insignificant. Second, mechanical spillovers can only really present themselves in blocks with one compliant course and one course with an elaborate policy. There is only one block where compliant and non compliant courses overlap. We looked for negative spillovers in this block and found no evidence of this being the case. Moreover, note that spillovers due

[26] Table A.6 in the Appendix presents estimates for the pooled data (all 8 courses).

[27] This relates to the results on prosocial behavior described by Gneezy, Meier, and Rey-Biel [2011]. The policy might signal distrust and affect social norms, which decreases cooperation of the student.

to knowledge on material are likely to be small as all second year courses have strongly different content.

Our next piece of evidence suggests that forced attendance decreases the chances of passing and that, moreover, the decrease is substantial. It could be that while students earn lower grades, they are more likely to pass because the tutorials give them a better overview of the minimum amount they need to know. The top panel of Table 11 shows that this is not the case, the probability of passing decreases by about 10 percentage points because of forced attendance. Note that the p-values hover around 10 percent, depending on our choice for the polynomial in the first-year grade. The bottom panel uses probit to show that forced attendance decreases the passing probability by the same amount for all bandwidth choices.

3.1. Robustness. We analyze the robustness of the result that forced attendance lowers grades. We estimate the reduced form with the third order polynomial while varying the size of the bandwidth from 0.3 until 1.0. Figure 5 shows the estimate hovers around -0.4 and -0.3 and is significant across the whole range of optimal bandwidths using the various bandwidth selectors of Calonico et al. [2016]. The estimate becomes marginally insignificant with bandwidths larger than 0.9. As our identification relies mostly on variation near the cutoff, we interpret Figure 5 as strong evidence for the robustness of the negative impact of the policy on student grades.

We test for significance at fake cutoffs. We estimate our main specification using the third order polynomial with the optimal bandwidth of 0.5, while implementing fake cutoffs at every 0.005 points for GPA between 6.5 and 7.5, where the true cutoff is at 7.[28] Figure 6 presents a histogram and probability density of the estimates for compliant courses. The distribution has a mean of zero. Our point estimate at the true cutoff is extreme relative to the mean, having an empirical p-value that ranges between 3 and 6 percent (depending on whether we assume normality). Finally, we also tested whether our results change if we restrict the linear polynomial $f(\cdot)$ to be the same on both sides of the cutoff. Table A.7 in the Appendix shows our results, where the estimates are virtually unchanged for all groups of courses.

3.2. Decline in Self Study. Table 12 uses data from course evaluations to investigate why forced attendance is bad for student performance. The table reports the policy impact on attendance at lectures, total time spent studying (lectures + tutorials + self study), as well as perceptions of the value of lectures and the quality of lecturers for the compliant courses. Lecture attendance is measured through a binary variable (0=no, 1=yes), total study time is measured in 10 categories (1=0 hours, 2=1 to 5 hours, and

[28] For example, at the fake cutoff of 6.9, we estimate Equation (1) for the sample with a first-year GPA of between 6.4 and 7.4.

10=more than 40 hours), and the value of lectures and quality of lecturer reflects a Likert scale (1=strongly disagree and 5 strongly agree).[29] The table allows for inferences concerning the impact of the policy on self study. This is because we have information on tutorial and lecture attendance as well as the total time spent studying.

Forced attendance induces students to spend less time studying own their own. Column (1) shows forced students attend 35 percentage points more lectures (over a baseline of 55 percent). Note that the estimates for the intercept and slope coefficient in the lecture attendance regressions are identical to their counterparts in the tutorial attendance. This lends support to the idea that the policy forces students to pay a cost that becomes sunk after they arrive at campus, so that going to lectures is relatively cheap when the student is already there. Column (2) shows no statistical difference in the total study time of students above and below the threshold of 7.[30] The results in Columns (1) and (2), together with the increase in tutorial attendance, imply that students sacrifice self study time when forced to attend classes. More specifically, forced students on average replace 35 hours of self study with going to class per block.[31] The results ultimately support the idea that students treat attendance as a substitute for self study.[32]

3.3. Peer Effects and TA Quality. Our results imply performance declines because students treat attendance as a substitute for self study. We consider two alternative mechanisms that could also be set in motion by tutorial attendance. The first relates to whether student performance is worse because of more regular interaction with lower quality peers. The second relates to whether performance is worse because TAs are simply not that good. Our analysis is based on the estimates in Table 13.

Column (1) considers treatment effect estimates for grades which are conditional on fixed effects for the tutorial. The fixed effects pick up the influences of peers and the TA because both are set and fixed before the start of the course. The estimates in Column (1) resemble our baseline results. The resemblance casts doubt on the role of peer effects and TA quality towards the decrease in student performance.

The specifications of columns (2) and (3) in Table 13 investigate whether the impact of forced attendance differs depending on the peer, resembling the most commonly estimated

[29]See Table A.2 for a more detailed description of the course evaluations.

[30]We get similar results if we use probit for column (1) and ordered probit for column (2).

[31]Note, we assume that students do not take into account travel time for attending tutorials and lectures while filling in the question for total study time.

[32]An alternative to our sunk cost interpretation would be that the tutorials enhance student capacity to learn from lectures. Columns (3) and (4) of Table 12 present evidence against this alternative interpretation. The columns show that forced students do not differ in their views on the usefulness of lectures or the quality of the lecturer. Note also that our sunk cost interpretation is further reinforced by Table 4, which shows that for courses with a more elaborate policy, lecture and tutorial attendance also coincide. The average student, forced or otherwise, will attend 90 percent of lectures and 90 percent of tutorials.

peer effects regression in the literature [Booij, Leuven, and Oosterbeek, 2016]. Column (2) reports the effects of interactions of the treatment dummy and the average 1^{st}-year grade for the peer group. Column (3) reports interaction effects for the average peer registration time for tutorials, measured in differences in days from the course mean registration time. This accounts for the possibility that students coordinate their tutorial times with their most preferred peers, which for forced students might very well include other poor performing peers. It also helps with the notion that the worst students end up in the same tutorial simply because they leave registration to the last minute. Finally, column (4) reports interaction effects for TA quality, as measured by the average response to the statement "The TA gives good tutorials". This accounts for overexposure of forced students to bad TAs.[33]

The effects of treatment interactions with peer and TA quality are modest at best. All the estimates are statistically insignificant at conventional significance levels, while the main treatment estimate is unchanged compared to our baseline specifications. Negligible peer effects are unsurprising given recent discussions and results in the literature [Sacerdote, 2014].[34] The same is true for TA quality, as instructors are typically careful to select a fairly homogenous group of high performing TAs, and on average TAs receive a rating of four out of five.

3.4. Longer-Run Impacts. We analyze the impact of the policy on student performance in the third year, where all students were free (from the university policy) to choose their attendance. The longer-run impacts on performance are ambiguous, at least in principle. Forced attendance in the second year may have led to less learning, or self learning, and thus provided students with worse preparation for the third year. Alternatively, forced students may have acquired skills that allow them to outperform students who scored just above 7 in their first year.

Table 14 examines the impact of second-year forced attendance on various measures of third-year performance. Estimates of the impact on the thesis grade are found in columns (1) and (2).[35] The impact on days until the graduation date are found in columns (3) and (4). Columns (5) and (6) estimate the policy impact on a binary variable that indicates whether the student is graduated for their bachelor, thereby also investigating selection issues with respect to the two previous outcome variables. Columns (7) and (8) examine

[33] At the time of registration students are unaware of the identity of TAs for the various tutorial groups. Their ignorance precludes a role for TA quality in registration decisions.

[34] For our study, Feld and Zölitz [2017] is especially relevant. They estimate peer effects of tutorial education for economics students at a Dutch university (Maastricht University). Similar to our context, for every course students are faced with a different set of peers. They find that higher-achieving peers increase student's grades by a statistically significant but small amount: a one standard deviation increase in the average peer GPA causes an increase of 1.26 percent of a standard deviation in student grades.

[35] All students are required to write a thesis to complete their bachelors.

the impact on course grades in the third year.

While the regression coefficients are for the most part imprecisely estimated and statistically insignificant, they are consistent with the policy having adverse impacts on performance in the longer run. The first two columns indicate that the policy decreased the thesis grade by 0.3 to 0.5 standard deviations. Columns (3) and (4) suggest forced students took longer to graduate, 30 to 50 days more than their counterparts. Columns (5) and (6) suggest graduation probability is not affected and therefore sample selection is not a major issue.[36] Columns (7) and (8) show a minor reduction in grades (at the course level). For the latter analysis we only include courses that are taken in the third year of the student. As such, consistent with our baseline results for the second year, with course-year fixed effects we are comparing students near 7 that belong to the same cohort.

We also analyze the impact of forced attendance on the choice of major, which is declared in the third year of the program.[37] We estimate a multinomial logit model where the choice probabilities for each major are given by

$$P[u_{ig} + \varepsilon_{ig} > u_{ih} + \varepsilon_{ih} \text{ for all } g \neq h] = \frac{e^{u_{ig}}}{1 + \sum\limits_{g=2}^{12} e^{u_{ig}}},$$

and $u_{ig} = \beta_{0g} + \beta_{1g} D_{ic} + f_g(\bar{G}_{ic}^{(1)} - 7) + f_g(\bar{G}_{ic}^{(1)} - 7) D_{ic} + C_{cg}$. β_{1g} estimates the effect of the policy upon the probability of choosing major g. The results are found in Table 15, for the linear specification in $f(\cdot)$ and third order polynomial respectively.

Where the model does not assume any ordering in the twelve different majors, we freely order the estimates for every major g in Table 15 from what is perceived to be the least technical to the most technical: fiscal economics, management accounting, financial accounting, entrepreneurship, marketing, urban and transport economics, applied economics, financial economics, behavioral economics, microeconomics, macroeconomics, and econometrics. Like the estimates for third-year performance, our estimates for major choice are imprecise. However, they do suggest that forced attendance in the second year induces students to choose somewhat less technical courses in the third year. In all, the analysis of this subsection implies that while there is some evidence of adverse longer-term impacts, most of the impact seems to be coming from grades in the year and courses where attendance is forced. To this end second-year performance alone provide a reasonable barometer for the impact of the policy on student welfare.

[36]The results in columns (5) and (6) are borne out in estimates of probit regressions.

[37]We categorize courses in their majors and indicate the major choice g of a student if he followed most of his courses within the major g. The majors in this European system are more likely to be interpreted as minors in the North American system.

4 Student Welfare

We use a simple economic model of student time use to guide the remainder of our empirical analysis. Taking this model seriously, it yields an attendance-adjusted welfare measure that can be implemented via instrumental variables. We will in turn present and discuss the IV estimates. In the next section we will use the model to guide our analysis of who is most affected by the policy, as well as to obtain an estimating equation for the relationship between the personal characteristics of students and the elasticity of substitution between attendance and self study.

4.1. A Model of Student Time Use. The problem for student i in the absence of forced attendance is

$$\begin{aligned} \underset{\tau,s}{\text{maximize}} \quad & g(\tau, s; \theta_i) \\ \text{subject to} \quad & p_i\tau + s = T \\ & \tau \geq 0, s \geq 0. \end{aligned}$$

$g_i = g(\tau, s; \theta_i)$ is the grade (utility) they derive from spending τ hours in tutorials and s hours studying on their own. θ_i indexes the production or utility function for grades by, for example, the ability of the student. g_i is twice continuously differentiable, increasing, and concave: $\partial g_i/\partial\tau > 0$, $\partial g_i/\partial s > 0$, $\partial^2 g_i/\partial\tau^2 \leq 0$, and $\partial^2 g_i/\partial s^2 \leq 0$. We assume $\partial g_i/\partial\tau$ and $\partial g_i/\partial s$ are infinite at $\tau = 0$ and $s = 0$ to avoid interior solutions at (0,0). T is the total time available.[38] p_i is the effective price of tutorial attendance, encapsulating the time and money cost of tutorial attendance. Let $\tau_i^* = \tau^*(p_i, T; \theta_i)$ denote the optimal attendance for student i and assume that there is well-behaved measure of the distribution for (θ_i, p_i) in the student population. In the next section we will make explicit how g_i and p_i vary per student i.

Forced attendance requires students to spend at least $\tau \geq \kappa > 0$ time at tutorials. The additional constraint has no impact on the student if $\tau_i^* > \kappa$, that is if the student would attend more than κ tutorials with or without the constraint. It has an impact on the student if $\tau = \kappa$, as they would prefer attending $\tau_i^* < \kappa$ of the time. The attendance we observe for any given student is thus:

$$\tau_i = D_i\Big[\tau_i^*\mathbf{1}(\tau_i^* > \kappa) + \kappa\mathbf{1}(\tau_i^* \leq \kappa)\Big] + (1 - D_i)\Big[\tau_i^*\mathbf{1}(\tau_i^* > \kappa) + \tau_i^*\mathbf{1}(\tau_i^* \leq \kappa)\Big]$$

where D_i equals 1 for forced students, and 0 otherwise. Note that τ_i^* is a latent variable

[38] We assume T is the same for everyone because there is strong norm towards treating university like work in the Netherlands, and because our results show that total study was unaffected by the forced attendance policy. A more general model would endogenize T by, for instance, including leisure as a choice variable for the student.

when the attendance constraint binds.

Taking expectations over (θ_i, p_i) gives

$$E[\tau_i|D_i] = \bar{\tau}^* + \underbrace{E[(\kappa - \tau_i^*)\mathbf{1}(\tau_i^* \leq \kappa)]}_{\substack{>0,\text{ Average Time Cost} \\ \text{Among ``Forced'' Students}}} D_i. \tag{2}$$

with $\bar{\tau}^* = E[\tau_i^*\mathbf{1}(\tau_i^* > \kappa) + \tau_i^*\mathbf{1}(\tau_i^* \leq \kappa)]$. The counterpart for self study is

$$E[s_i|D_i] = \bar{s}^* + \underbrace{E[p_i(\tau_i^* - \kappa)\mathbf{1}(\tau_i^* \leq \kappa)]}_{\substack{<0,\text{ Value of Lost Time} \\ \text{Among ``Forced'' Students}}} D_i.$$

The coefficient on D_i in $E[\tau_i|D_i]$ is such that the attendance requirement unsurprisingly, and always, increases attendance. The impact grows with the distance from the attendance requirement κ to the preferred attendance of a forced student τ_i^*. The students who always attend shape the expected attendance decision through their impact on the intercept. The coefficient on D_i in $E[s_i|D_i]$ has the opposite sign, as it decreases the time students spend studying on their own. It becomes more negative when there are increases in the distance from the attendance requirement κ to the preferred attendance of a forced student.

We can do the same for student grades, where $V_i = g(\tau_i^*, s_i^*; \theta_i)$, and

$$E[V_i|D_i] = \bar{V}^* + \underbrace{E[(V(\kappa) - V_i^*)\mathbf{1}(\tau_i^* \leq \kappa)]}_{\substack{<0,\text{ Utility Loss} \\ \text{Among ``Forced'' Students}}} D_i. \tag{3}$$

Note how the coefficients on D_i are in line with the treatment effect estimates in our baseline results. The estimates line up even though we have imposed no substantive assumptions on the problem for the student. In this way the problem for the student fits well with basic economic analysis. Taking this framework seriously, Equation (3) provides us with a measure for the student's welfare loss of forced attendance in terms of (standardized) grades.[39]

4.2. A Measure of Welfare Loss at the Margin. Dividing the coefficients on D_i in $E[V_i|D_i]$ and $E[\tau_i|D_i]$ gives the welfare loss for a marginal increase in forced attendance

$$\frac{E[(V(\kappa) - V_i^*)\mathbf{1}(\tau_i^* \leq \kappa)]}{E[(\kappa - \tau_i^*)\mathbf{1}(\tau_i^* \leq \kappa)]},$$

[39] Of course this is not a comprehensive measure of welfare. It excludes, for example, the cost to providing the tutorials and the potential benefits of student attendance for the university.

which is essentially the derivative of V_i^* for forced students, around the attendance requirement κ. Note that the sample counterpart is a Wald estimator, so that the derivative can be obtained via a simple IV strategy where D_i is used as an instrument in a regression of grades on tutorial attendance. This marginal measure should provide useful information to university administrators and instructors considering marginal or more moderate forced attendance policies.

Accordingly, we estimate

$$G_{ijc}^{(2)} = \beta_0 + \beta_1 A_{ijc}^{(2)} + f(\bar{G}_{ic}^{(1)} - 7)A_{ijc}^{(2)} + f(\bar{G}_{ic}^{(1)} - 7) + C_{jc}^{(2)} + \mathbf{X}_i\mathbf{\Gamma} + \varepsilon_{ijc}^{(2)} \tag{4}$$

where

$$A_{ijc}^{(2)} = \pi_0 + \pi_1 D_{ic} + f(\bar{G}_{ic}^{(1)} - 7) + f(\bar{G}_{ic}^{(1)} - 7)D_{ic} + \pi_{jc}^{(2)} + \mathbf{X}_i\mathbf{\Pi} + \nu_{ijc}^{(2)} \tag{5}$$

$A_{ijc}^{(2)}$ is the percentage of tutorials attended by student i in cohort c and course j. The instrument set includes the interactions between the treatment dummy and $f(\cdot)$, $\{D_{ic}, f(\cdot)D_{ic}\}$, to match the number of endogenous variables, $\{A_{ijc}^{(2)}, f(\cdot)A_{ijc}^{(2)}\}$. Note that the policy is a locally valid instrumental variable near the cutoff. The parameter of interest is β_1. Our reduced form and first stage estimates imply that $\beta_1 < 0$. The only question that remains relates to whether their ratio is meaningful statistically.

Our results all support the credibility of the IV estimation. Our instrument is exogenous, powerful, and they imply strongly the validity of the exclusion restriction as it seems to operate on grades only via its influence on attendance: (i). we did not find an effect of the policy on grades for courses where attendance was fixed; (ii). our identified channel of substitution between attendance and self study can only be set in motion by the policy impacting tutorial attendance; (iii). we find other channels that might not go through attendance are absent (e.g. general opinion about the course), see Table A.8 in the Appendix.

4.3. IV Estimates of the Marginal Loss. Table 16 display our results for the IV estimates. Columns (1) and (2) imply that a 10 percentage point increase in attendance decreases student performance by 0.1 of a standard deviation for the forced student. The IV estimate is statistically significant throughout all of our specifications. For the remaining courses (columns (3) to (6)) the IV estimates are statistical zeros, as expected.

It is worthwhile to reiterate how our estimates compare with other estimates in the literature that examines the impact of class attendance on academic performance. While most studies find a positive relationship between class attendance and academic performance, they have difficulty with the endogeneity of attendance. Because attendance is a choice, and better students tend to attend more often, the relationship is likely over-estimated. Our IV estimates, combined with the positive naive estimates for first-year attendance in Table 6, imply that this is the case.

As a robustness check for the exclusion restriction, we use a method proposed by Conley, Hansen, and Rossi [2012]. We use the fact that in courses with a more elaborate policy, the estimated impact on grades is not exactly zero. Acting conservatively, we can interpret this estimate, call it ψ_0, as a direct negative effect of the policy upon student grades that does not go through attendance. Conley, Hansen, and Rossi [2012] suggest to use the estimate as a correction, estimating an IV regression where grades are explicitly corrected for the direct effect of the policy $(G^{(2)}_{ijc} - \psi_0 D_{ic})$. We find that the sign and significance of our IV estimate is robust to this proposed correction.

5 Who Loses from Forced Attendance

The coefficients on D_i in $E[\tau_i|D_i]$ and $E[V_i|D_i]$ in Equations (2) to (3), $E[(\kappa - \tau_i^*)\mathbf{1}(\tau_i^* \leq \kappa)]$ and $E[(V(\kappa) - V_i^*)\mathbf{1}(\tau_i^* \leq \kappa)]$, make clear that our reduced form and IV estimates are being generated by students who would not attend in the absence of the forced attendance policy. In this section, we first use our data on personal characteristics and our simple model to draw inferences about the identities of these students. We start by estimating Equation (2)

$$A^{(2)}_{ijc} = \gamma_0 + \gamma_{1ic} D_{ic} + \varepsilon^{(2)}_{ijc}$$

where in effect $\gamma_{1ic} = (\kappa - \tau_{ic}^*)\mathbf{1}(\tau_{ic}^* \leq \kappa)$ and it measures the impact of the forced attendance policy on attendance.[40] If γ_{1ic} is large, then student i (in cohort c) has a low τ_{ic}^*, and would have attended far fewer tutorials in the absence of the forced attendance policy. Alternatively, a small γ_{1ic} implies τ_{ic}^* is high, and that the student would attend the same number of tutorials with or without a forced attendance policy. In the parlance of the treatment effects literature, students with a large γ_{1ic} are compliers. Students with a small γ_{1ic} are classified as always takers. There are no never takers or defiers by definition, as the policy gives them no choice but to attend the tutorials when their average first-year grade is below 7. Indeed, of the students with a first-year GPA below 7, only 0.44 percent has an attendance rate below 70 percent.[41]

[40] To be precise, γ_{1ic} would equal $(\kappa - \tau_{ic}^*)\mathbf{1}(\tau_{ic}^* \leq \kappa)$ if τ was measured in proportions.

[41] One might argue that we do not observe a grade for a never taker in our dataset, as he or she cannot write the exam. However, in Section 2.3 we showed that in general students participate in every course in the second year, as counted by the number of (valid) grades per student, and that this near-perfect course participation is unaffected by the treatment (leaving no room for never takers). As such, one might want to interpret our IV as an average treatment effect on the nontreated. The nontreated group contains compliers and never takers, where the latter do not exist in our setup. Note, however, in our specifications we treat compliance as going to the tutorials in general and do not interpret it as an adherence to the 70 percent rule.

We operationalize γ_{1ic} by re-estimating Equation (5) while interacting the treatment dummy with background characteristics of the student. Estimates are found in Table 17. From left to right the table reports interaction effects for distance to the university, whether the student pays the low tuition fee (measured through whether a student comes from within the European Economic Area),[42] their age, gender, and high school grade. All specifications include controls.

Three patterns stand out. First, the direct effect of the characteristic is always opposite to the effect of its interaction with the treatment. This implies that the interactions indeed pick up the counterfactual difference between the impact of the policy and the attendance the student would have chosen had the policy not been in place. Second, the policy had a larger impact on students who live far from campus. A forced student would experience a 4.4 percentage points larger impact on attendance with every standard deviation increase in the distance to campus. Second, the policy had a larger impact on younger students. The attendance rate for the forced student increases by about 3 percentage points if he or she is younger by one standard deviation. The impacts on younger students and ones that live far from campus coincides with our intuition, as these students are more prone to bad decisions, from the universities perspective, and face higher costs to traveling to campus. The remaining estimates for the low tuition fee, gender, and high school grade have the expected signs, but are statistically insignificant.

5.1. Age and Distance in Time Use We integrate personal characteristics, and in particular age and distance, into our simple model for the purposes of evaluating their influence on the marginal rate of substitution between class attendance and self study. We assume g $= (\tau^\rho + s^\rho)^{\frac{1}{\rho}}$, where the elasticity of substitution is $\sigma = 1 - 1/\rho$. We assume further g_i and p_i differ per student i in the following way: the elasticity of substitution is a function of age $\sigma = \sigma(a)$ and distance from campus d shape the price of tutorial attendance $p = p(d)$. This latter assumption is guided by our results of the complier analysis, which might also be interpreted as students that live far from campus paying a higher price (travel cost) for going to the tutorials.

The first order condition to this problem states

$$(\frac{\tau}{s})^{\rho-1} = p.$$

[42] Low tuition fee equals 1 if a student comes from within the European Economic Area (EEA) and 0 otherwise. In 2017-2018 for an EEA member the tuition fee was equal to €2.006, whereas for a non-EEA member it was equal to €8.900.

or equivalently

$$ln(\frac{\tau}{s}) = -\sigma ln(p).$$

Estimation requires that we make functional form assumptions for $\sigma(a)$ and $p(d)$. We assume $\sigma(a) = \delta_1 + \delta_2 a$ and $p(d) = \exp(d)$, where the latter assumption is motivated by the distribution of distance in our data.[43] Plugging the functional forms into the first order condition gives

$$ln(\frac{\tau}{s}) = -\delta_1 d - \delta_2 a \times d.$$

With this equation we can recover the elasticity of substitution via a simple regression of the ratio of the inputs upon distance and its interaction with age. The estimate δ_1 delivers the elasticity when $a = 0$, whereas δ_2 measures how the elasticity varies with age (standardized with mean 0 and standard deviation 1).

5.2. Age and the Elasticity of Substitution We estimate the natural empirical counterpart to the above equation

$$ln(\frac{\tau_{ijc}^{(2)}}{s_{ijc}^{(2)}}) = \theta_0 + \theta_1 d_i + \theta_2 d_i \times a_i + \theta_3 a_i + C_{jc}^{(2)} + \mathbf{X}_i\mathbf{\Gamma} + \varepsilon_{ijc}^{(2)} \tag{6}$$

where $-\theta_1$ measures the elasticity of substitution and $-\theta_2$ measures how it varies with age. For calculating $\tau_{ijc}^{(2)}$ (and $s_{ijc}^{(2)}$) we also take into account our data on lecture attendance, which, as previously discussed, shows identical attendance patterns to the tutorials. Results are found in Table 18. Column (1) estimates a specification directly related to the first order condition, putting θ_0 and θ_3 to zero and excluding course-year fixed effects and further controls. The next three columns gradually increase the specification, including a constant and course-year fixed effects (column (2)), age (column (3)), and the remaining control variables (column (4)).

Two patterns stand out across all specifications. First, class attendance and self study are complementary for the average (aged) student. Second, the complementarity becomes weaker and weaker for younger students. Eventually, for the youngest students, they become substitutes. We conclude the forced attendance policy is most harmful to the students it supposedly helps.

[43] These functional form assumptions are a bit *ad hoc*. They, however, deliver estimates that are consistent with reduced forms and, importantly, allows us to pin, more precisely, the heterogeneous impacts of the policy.

6 Conclusion

Studying the impacts of policies that force people do things is difficult, if not impossible, via randomization. We draw on a natural experiment at a large European University to estimate the causal impact of forced attendance on student performance. We exploit the introduction and abolition of a policy that require students to travel and attend up to 250 more hours of tutorials in their second year if they failed to earn a GPA of 7 (on a 10-point scale) in their first year. Using the discontinuity at 7, we estimate that forced attendance decreases the final grade of the student by 0.35 standard deviations. We argue and show that the policy impact operates via attendance, ruling out other channels that may have lowered student grades. We then use the policy as an instrument for estimating the marginal impact of attendance on grades, and show that a 10 percentage point increase attendance decreases grades by 0.1 of a standard deviation (among forced students). Forced attendance induces students to spend less time studying on their own, implying that it lowers grades because students treat tutorials as a substitute for self study. Accordingly, we show that attendance of younger students and students who live far from campus are most affected by the policy. We conclude that forced attendance is especially detrimental for the students it, in principle, aims to help.

Our findings should be of use to universities and university instructors around the world, especially ones who are considering forced or mandatory attendance policies. To this end, we discuss aspects of our context that might be worth thinking about ahead of time. First, because the university we study is one of the more reputable universities in the Netherlands, it attracts many of the best students. Second, and relatedly, we have studied students from a (large) school of economics. One could have anticipated our main results, namely that students respond in the way that economics predicts they would, as these students are trained in the reasoning of rational agents. Third, Dutch students have a propensity to fix the number of hours they spend studying. They treat university or school like a job in the sense that they have a cap on how many hours they study per week. Forced attendance cuts into those hours, as students are, as this study shows, reluctant to adjust their total number of study hours. In other contexts students may be more willing to adjust the total number of hours they spend studying. In the end, the decision to adopt forced or mandatory attendance policies should weigh the benefits to students who are prone to errors against the costs to students who generally make good decisions. To this end it might make sense to continue to use compulsory schooling laws at primary and secondary schools (see *e.g.* Oreopoulos [2007]), or even to adopt forced attendance policies in colleges and universities that tend to attract students who are prone to bad decisions.

Our focus has mainly been on the instrumental value of giving, or taking away, the decision rights of students. We have said relatively little about how the policy affected the intrinsic value of decision rights. Was there, in other words, a utility loss from simply not having the full right to decide how you spend your time?[44] On this front, it would be interesting to measure the intrinsic value students derive from having decision rights, in the spirit of the recent work by Bartling, Fehr, and Herz [2014], and to evaluate whether and to what extent the intrinsic value is altered by policies that force students to come to campus.

[44]Note that the finding of the absence of an effect on grades in non compliant courses does not rule this out. It could also be that this utility loss did not impact grades.

References

Acemoglu, D. and A. Wolitzky (2011). The economics of labor coercion. *Econometrica 79*(2), 555–600.

Angrist, J., D. Lang, and P. Oreopoulos (2009). Incentives and services for college achievement: Evidence from a randomized trial. *American Economic Journal: Applied Economics 1*(1), 136–163.

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.

Arulampalam, W., R. A. Naylor, and J. Smith (2012). Am i missing something? the effects of absence from class on student performance. *Economics of Education Review 31*(4), 363–375.

Bartling, B., E. Fehr, and H. Herz (2014). The intrinsic value of decision rights. *Econometrica*, 2005–2039.

Bloom, N., J. Liang, J. Roberts, and Z. J. Ying (2014). Does working from home work? evidence from a chinese experiment. *The Quarterly Journal of Economics*, qju032.

Booij, A. S., E. Leuven, and H. Oosterbeek (2016). Ability peer effects in university: Evidence from a randomized experiment. *The Review of Economic Studies*, rdw045.

Buser, T., M. Niederle, and H. Oosterbeek (2014). Gender, competitiveness, and career choices. *The Quarterly Journal of Economics 129*(3), 1409–1447.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2016). rdrobust: Software for regression discontinuity designs. *Unpublished manuscript available at: http://faculty. chicagobooth. edu/max. farrell/research/Calonico-Cattaneo-Farrell-Titiunik_2016_Stata. pdf*.

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

Conley, T. G., C. B. Hansen, and P. E. Rossi (2012). Plausibly exogenous. *Review of Economics and Statistics 94*(1), 260–272.

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.

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.

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.

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

Fryer Jr, R. G. (2016). The production of human capital in developed countries: Evidence from 196 randomized field experiments. Technical report, National Bureau of Economic Research.

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.

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

Jackson, C. K. (2012). Non-cognitive ability, test scores, and teacher quality: Evidence from 9th grade teachers in north carolina. Technical report, National Bureau of Economic Research.

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.

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. 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. (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.

Rodgers, J. R. (2002). Encouraging tutorial attendance at university did not improve performance. *Australian Economic Papers 41*(3), 255–266.

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.

Snyder, J. L., J. E. Lee-Partridge, A. T. Jarmoszko, O. Petkova, and M. J. DOnofrio (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.

Van der Klaauw, W. (2002). Estimating the effect of financial aid offers on college enrollment: A regression–discontinuity approach. *International Economic Review 43*(4), 1249–1287.

Visaria, S., R. Dehejia, M. M. Chao, and A. Mukhopadhyay (2016). Unintended consequences of rewards for student attendance: Results from a field experiment in indian classrooms. *Economics of Education Review 54*, 173–184.

Table 1: **Overview of Program.**

Group	First Year Courses	ECTS	Block	Second Year Courses	ECTS	Block
	Microeconomics	8	2	Applied Microeconomics	8	3
A	Macroeconomics	8	3	International Economics	8	1
	Organisation and Strategy	8	5	History of Economic Thought	4	3
	Financial Information Systems	4	1	Intermediate Accounting	8	5
B	Marketing	8	4	Behavioral Economics	4	4
	Financial Accounting	8	5	Finance I	8	2
	Mathematics I	4	1	Methods & Techniques	8	4
	Mathematics II	4	3	Research Project	4	5
C	Applied Statistics I	4	4	Applied Statistics II	4	2
	ICT	4	2	Economics of Ageing (Eng) or	4	1
				Fiscal Economics (Dutch)	4	1

Notes: The Economics of Ageing is taught in the English program. The Dutch program substitutes this for Fiscal Economics. Students can compensate an 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 cluster is above 5.5. This applies to all students, whether they are above or below the threshold for the forced attendance policy.

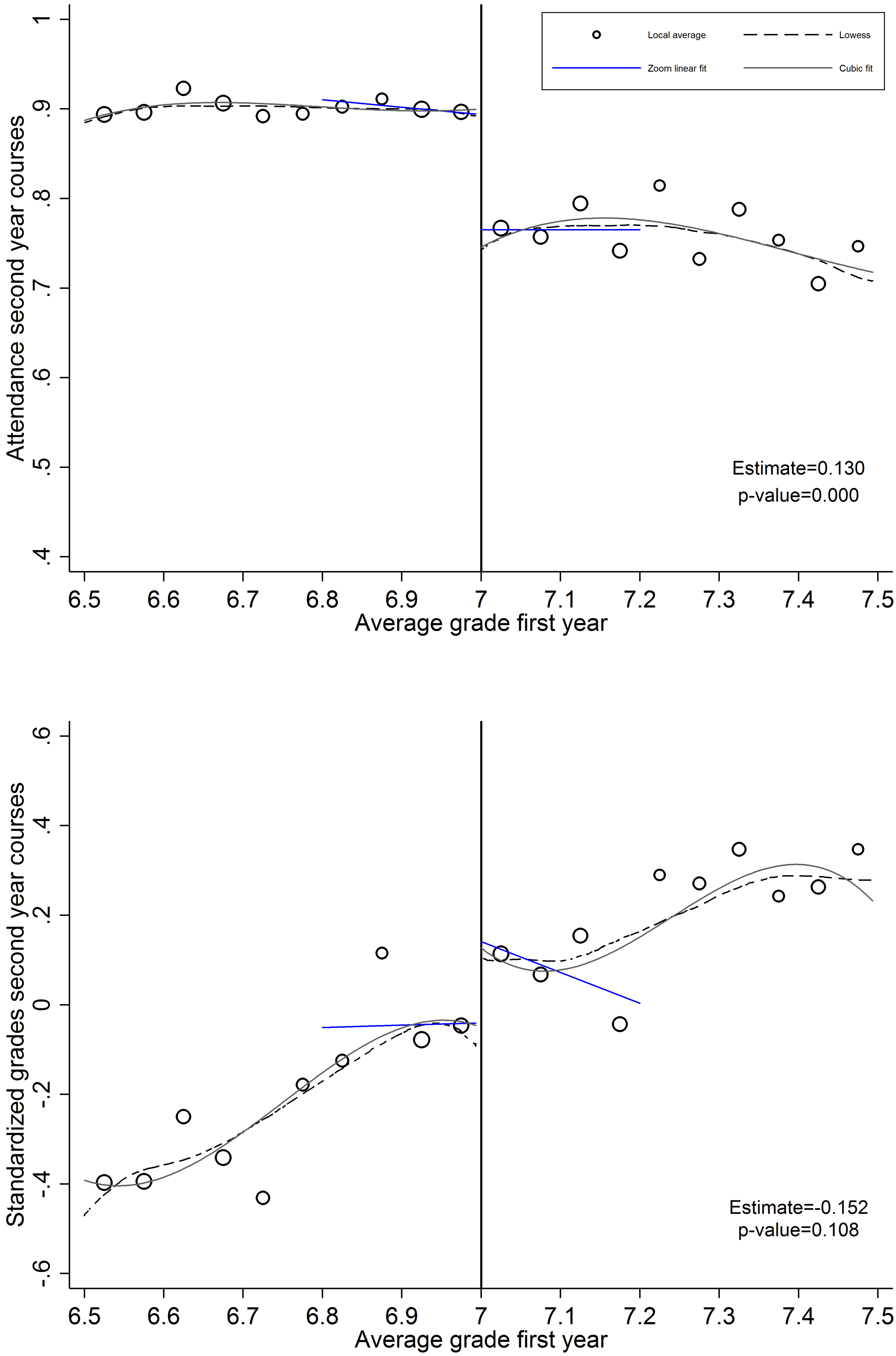
Table 2: **Basic Descriptives (All 8 Eligible Courses).**

	Grade Range		
	[6.5-7)		[7-7.5]
Course level (second year)			
Observations	2626		2323
Grades (std)	-0.226	***	0.185
Attendance tutorials	0.901	***	0.759
Student level			
Observations	388		335
Gender (1=female)	0.303		0.313
Age	20.16		20.22
Distance to uni. (in km)	19.56		21.79
Dutch (1=no)	0.121		0.143
GPA sec. school (std)	-0.098	***	0.117
Sec. education type	6.496		6.407

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Stars denote the statistical significance for the difference in means.

Figure 1: **Second Year Attendance and Grades (All 8 Eligible Courses).**



Notes: Locally linear, cubic and weighted scatterplots (lowess) for attendance or 2^{nd}-year grade against average 1^{st}-year grade. Dots are based on local averages for a binsize of 0.05. Dot sizes reflect the number of observations used to calculate the average. Linear and cubic fits are chosen according to our preferred specifications (see Section 2). Lowess makes no assumption on functional form (estimated with a bandwidth of $0.8N$). Binsizes for local averages are selected via F-tests for regressions of 2^{nd}-year grades on K bin dummies and $2K$ bin dummies for the average 1^{st} year grade.

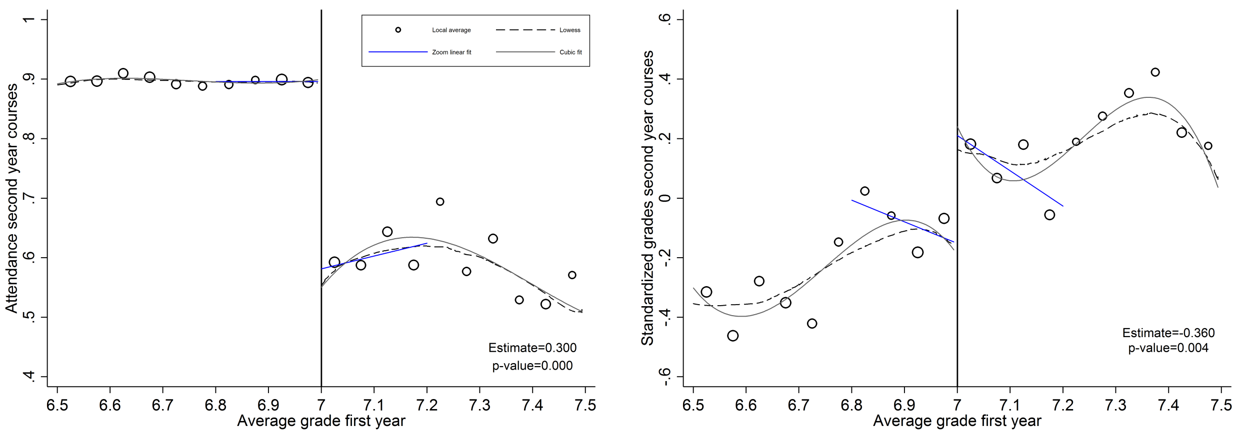
Table 3: **Compliance of Courses in the Second Year.**

Course	Tutorials	Compliance	Years	Tutorial Description
Intermediate Accounting	✓	✓	2009-2013	Exercises, No direct influence on final grade.
Behavioral Economics	✓	✓	2010-2013	Exercises, Actual Experiments, No direct influence on final grade.
Methods & Techniques	✓	✗	2009-2013	Exercises in Computer Lab, Accounts for 5 percent of final grade. Absence implies a 0 out of 5.
Applied Statistics II	✓	✗	2009-2013	Exercises, Accounts for 15 percent of final grade. Absence implies a 0 out of 15.
Ageing or Fiscal Economics	✓	✗	2010-2013	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.
International Economics	✓	✓✗	2009-2013	Students explicitly told to attend 10 of 13 tutorials. Discussion of exercises that are hand in before tutorial. No direct influence on final grade.
Finance I	✓	✓✗	2009-2013	Exercises, Outside tutorials there are weekly quizzes that account for 20 percent of final grade.
Applied Microeconomics	✓	✓✗	2009-2013	Exercises, No direct influence on final grade.
Research Project	✗		2009-2013	Group research projects.
History of Economic Thought	✗		2009-2013	Group and individual research projects.

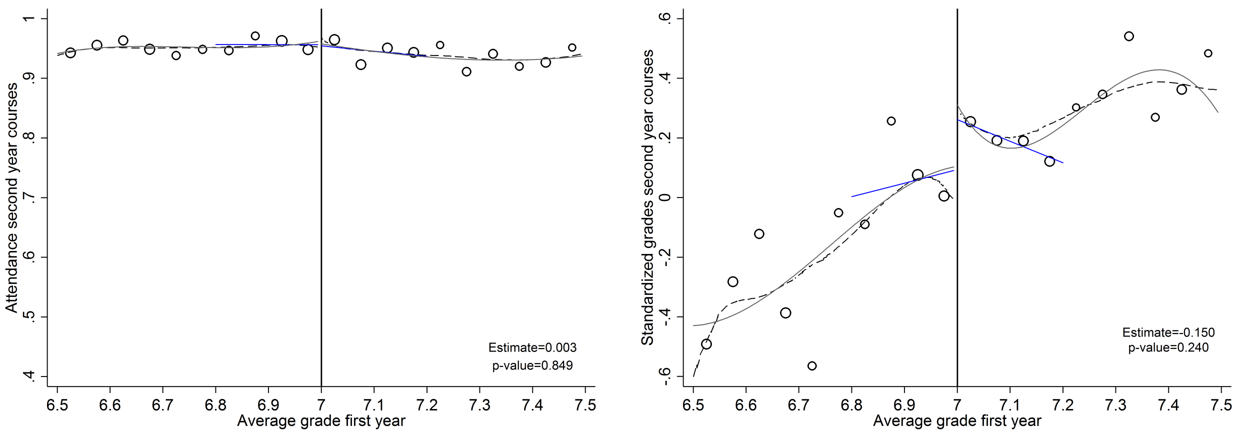
Notes: The description is extracted from course guides. ✓✗ indicates partial compliance by the course instructor.

Figure 2: **Breakdown by the Course's Adherence to the Policy.**

(a) Courses that follow the university policy



(b) Courses with a more elaborate attendance policy



(c) Courses with a more relaxed attendance policy

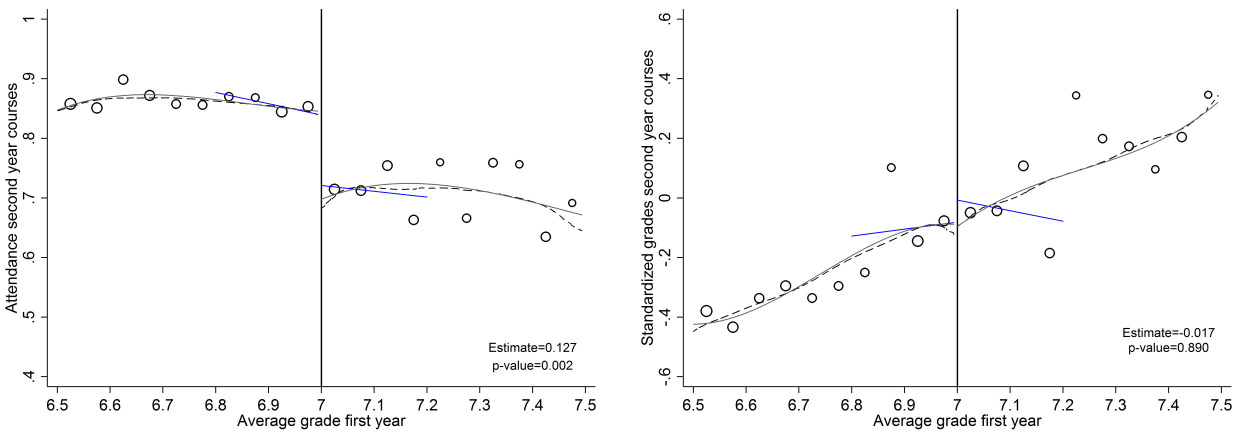


Table 4: **Basic Descriptives by Course Compliance.**

	Courses that follow University Policy			Courses with more Elaborate Policy			Courses which are More Relaxed		
	Grade Range			Grade Range			Grade Range		
	[6.9-7)		[7-7.1]	[6.9-7)		[7-7.1]	[6.9-7)		[7-7.1]
Course level (second year)									
Observations	163		148	216		202	248		232
Grades (std)	-0.128	***	0.130	0.042	**	0.224	-0.112		-0.046
Attendance tutorials	0.897	***	0.590	0.955		0.944	0.848	***	0.713
Avg. TA quality	4.138		3.895	3.903		4.171	4.283		3.931
P[lecture attendance]	0.840		0.640	0.914		0.957	0.937		0.918

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Stars denote the statistical significance for the difference in means.

Table 5: **It is the Forcing that Worsens Performance.**

Cohort	GPA $\in [6.9 - 7.0)$		GPA $\in [7.0 - 7.1]$
2009 - 2013	6.39	$p = 0.001^{***}$	6.77
	$p = 0.057^{*}$		$p = 0.217$
2014	6.75	$p = 0.303$	6.89

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Local averages of (unstandardized) grades for a bandwidth of 0.1, p-values denote one-sided statistical significance for the corresponding comparison of the averages.

Table 6: **Incentive and Opportunity for Manipulation.**

	Grades (Standardized)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
% Attendance	0.578***	0.549***	0.383***	0.563***	0.533***	0.372***	0.561***	0.618***
	(5.04)	(5.00)	(4.03)	(4.92)	(4.87)	(3.92)	(4.54)	(4.27)
GPA secondary school							0.0204**	
							(1.97)	
Observations	6505	6530	6565	6505	6530	6565	5386	6505
Adjusted R^2	0.338	0.337	0.339	0.338	0.337	0.339	0.346	0.367
Controls	No	No	No	Yes	Yes	Yes	Yes	No
Attendance	$\geq$0.7	$\geq$0.15	$\neq$0	$\geq$0.7	$\geq$0.15	$\neq$0	$\geq$0.7	$\geq$0.7

Notes: t statistics in parentheses, standard errors are clustered on the student level
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The sample of students used for these regressions corresponds to the sample used for the main analysis (regression discontinuity with a bandwidth of 0.5). All regressions include course-year fixed effects. Controls include date of birth, distance to university, gender, and nationality. Column (7) includes GPA for secondary school as an extra control variable, which is only available for the sample of Dutch students, and column (8) includes student fixed effects.

Table 7: **Similarity of Students around the Cutoff.**

	Nationality	Age	Distance	Gender	Level of Secondary Education	Track Secondary Education	Grades Secondary Education
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				A: Local linear regression			
Average 1^{st}-year Grade is Below 7	0.0359	0.319	1.661	0.266**	0.0214	0.241	-0.209
	(0.47)	(1.47)	(0.23)	(2.59)	(1.14)	(1.14)	(-0.83)
Observations	312	312	312	312	248	248	248
Adjusted R^2	0.010	0.004	-0.002	0.002	0.006	0.274	0.003
				B: Third order polynomial			
Average 1^{st}-year Grade is Below 7	0.0103	0.408	-3.903	0.279**	0.00429	0.170	-0.348
	(0.11)	(1.44)	(-0.41)	(2.14)	(0.21)	(0.63)	(-0.82)
Observations	723	723	723	723	592	592	592
Adjusted R^2	0.004	-0.005	0.007	0.000	0.017	0.258	0.008
				C: Third order polynomial			
Average 1^{st}-year Grade is Below 7	0.00258	0.304	1.256	0.0905	0.0422*	0.245	-0.313
	(0.04)	(1.63)	(0.22)	(1.15)	(1.72)	(1.47)	(-1.29)
Observations	1431	1431	1431	1431	1177	1177	1177
Adjusted R^2	0.014	0.000	0.009	0.001	0.011	0.350	0.065
Mean Outcome Var.	0.13	20.3	22.94	0.29	0.98	2.50	0.00

Notes: t statistics in parentheses, standard errors are robust.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
All regressions are on the student level and include year fixed effects. The top panel displays local linear regressions with the optimal bandwidth of 0.2 around 7, the mid panel shows regressions for the optimal bandwidth of 0.5 with the third order polynomial, and the bottom panel includes all observations. Column (5) until (7) only use the sample of Dutch students.

Figure 3: **No Bunching Just Above 7.** RD plot of the density for the number of students.

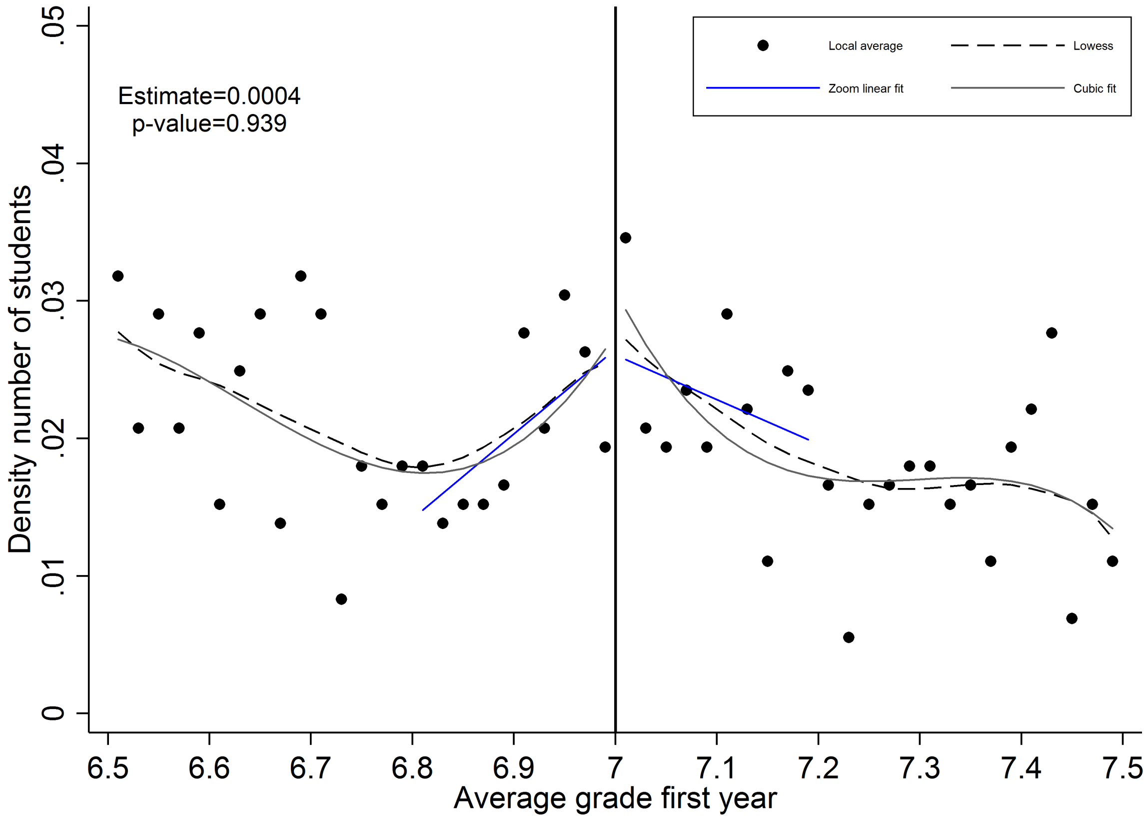


Figure 4: **Grades After Forced Attendance is Abolished.** Data is from the 2014-15 academic year. All courses on top. Courses that previously complied on the bottom.

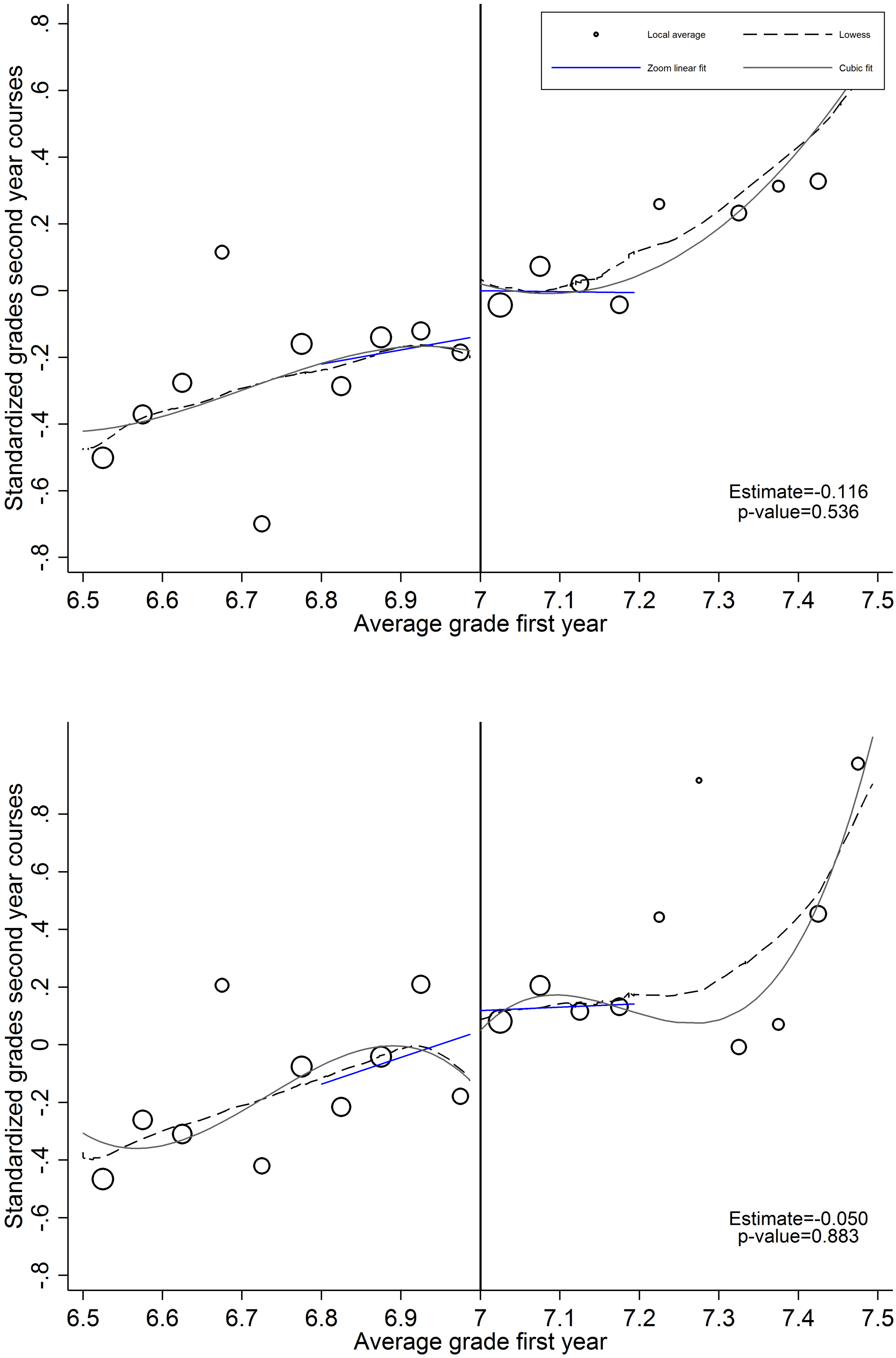


Table 8: **Sample Selection.**

	Number of Courses			Completed Course Evaluation			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Average 1^{st}-year Grade is Below 7	0.124	-0.0137	0.0359	-0.0885	-0.0601	-0.0735	-0.0713
	(0.47)	(-0.09)	(0.10)	(-1.48)	(-0.82)	(-1.11)	(-0.90)
Intercept$^+$	9.166***	9.198***	9.166***	0.202***	0.186***	0.180***	0.170***
	(54.11)	(86.21)	(40.80)	(4.78)	(3.89)	(4.15)	(3.40)
Observations	312	723	723	2152	4949	551	1287
Adjusted R^2	0.040	0.035	0.032	0.055	0.071	-0.010	0.017
Bandwidth	0.2	0.5	0.5	0.2	0.5	0.2	0.5
Polynomial	1^{st}	1^{st}	3^{rd}	1^{st}	3^{rd}	1^{st}	3^{rd}

Notes: t statistics in parentheses, standard errors are robust (columns (1) until (3)) or clustered on the student level (columns (4) until (7))

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Columns (1) until (3) include year fixed effects, whereas column (4) until (7) include course-year fixed effects. All regressions do not include further controls.

$^+$ The intercept is calculated via regressions which exclude year or course-year fixed effects - such that it approximates the outcome mean near the threshold.

Table 9: **Forced Attendance Decreases Grades.**

	Grades (standardized)					
	Courses that follow the University Policy		Courses with more Elaborate Policy		Courses which are More Relaxed	
	(1)	(2)	(3)	(4)	(5)	(6)
	A: Local linear regression					
Average 1^{st}-year Grade is Below 7	-0.360***	-0.337***	-0.150	-0.163	-0.0172	0.00905
	(-2.90)	(-2.70)	(-1.18)	(-1.25)	(-0.14)	(0.07)
Observations	551	551	748	748	853	853
Adjusted R^2	0.178	0.177	0.094	0.094	0.201	0.200
	B: Third order polynomial					
Average 1^{st}-year Grade is Below 7	-0.429***	-0.422***	-0.166	-0.166	0.0280	0.0438
	(-2.70)	(-2.65)	(-1.02)	(-1.01)	(0.17)	(0.27)
Observations	1287	1287	1679	1679	1983	1983
Adjusted R^2	0.217	0.216	0.153	0.156	0.251	0.250
Controls	No	Yes	No	Yes	No	Yes

Notes: t statistics in parentheses, standard errors are clustered on the student level
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
All regressions include course-year fixed effects. The top panel uses a bandwidth of 0.2 around 7. The bottom panel uses a bandwidth of 0.5.

Table 10: **Forced Attendance Increases Attendance.**

	Attendance (% Tutorials Attended)					
	Courses that follow the University Policy		Courses with more Elaborate Policy		Courses which are More Relaxed	
	(1)	(2)	(3)	(4)	(5)	(6)
	A: Local linear regression					
Average 1^{st}-year Grade is Below 7	0.299***	0.285***	0.00312	0.000733	0.127***	0.107***
	(6.32)	(6.22)	(0.19)	(0.04)	(3.11)	(2.81)
F-test	39.95	38.69	0.04	0.00	9.67	7.90
Observations	551	551	748	748	853	853
Adjusted R^2	0.365	0.376	0.149	0.174	0.158	0.185
	B: Third order polynomial					
Average 1^{st}-year Grade is Below 7	0.345***	0.333***	-0.000408	-0.000905	0.153***	0.143***
	(5.81)	(5.72)	(-0.02)	(-0.05)	(3.00)	(2.90)
F-test	33.76	32.72	0.00	0.00	9.00	8.41
Observations	1287	1287	1679	1679	1983	1983
Adjusted R^2	0.407	0.411	0.144	0.150	0.177	0.187
Controls	No	Yes	No	Yes	No	Yes

Notes: t statistics in parentheses, standard errors are clustered on the student.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
All regressions include course-year fixed effects. The top panel uses a bandwidth of 0.2 around 7. The bottom panel uses a bandwidth of 0.5.

Table 11: **Forced Attendance Decreases Passing Probability.**

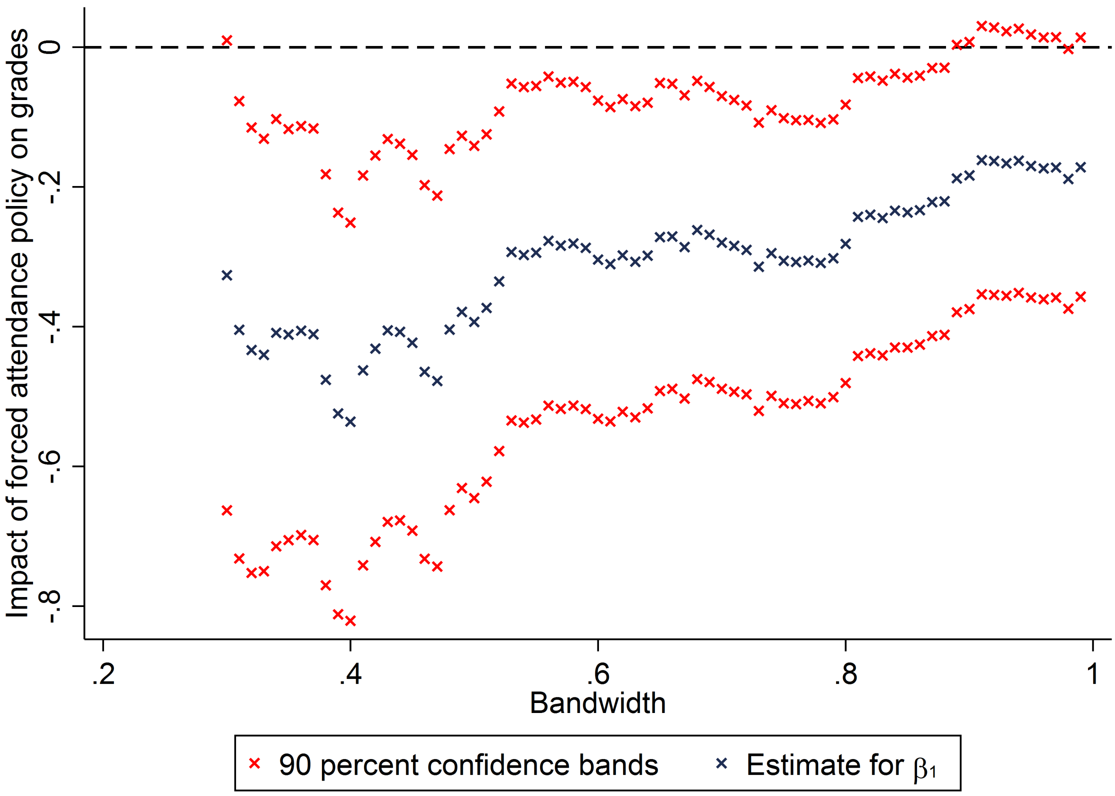
	Passing Probability (1=sufficient grade, 0=insufficient grade)			
	Local linear regression		Third order polynomial	
	(1)	(2)	(3)	(4)
		A: OLS		
Average 1^{st}-year Grade is Below 7	-0.129** (-2.25)	-0.116** (-2.02)	-0.108 (-1.49)	-0.0991 (-1.38)
Adjusted R^2	0.076	0.077	0.085	0.083
		B: Probit		
Average 1^{st}-year Grade is Below 7	-0.160** (-2.37)	-0.143** (-2.15)	-0.130* (-1.67)	-0.120 (-1.56)
Observations	551	551	1287	1287
Pseudo R^2	0.085	0.095	0.125	0.127
Controls	No	Yes	No	Yes

Notes: t and z statistics in parentheses, standard errors are clustered on the student level

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

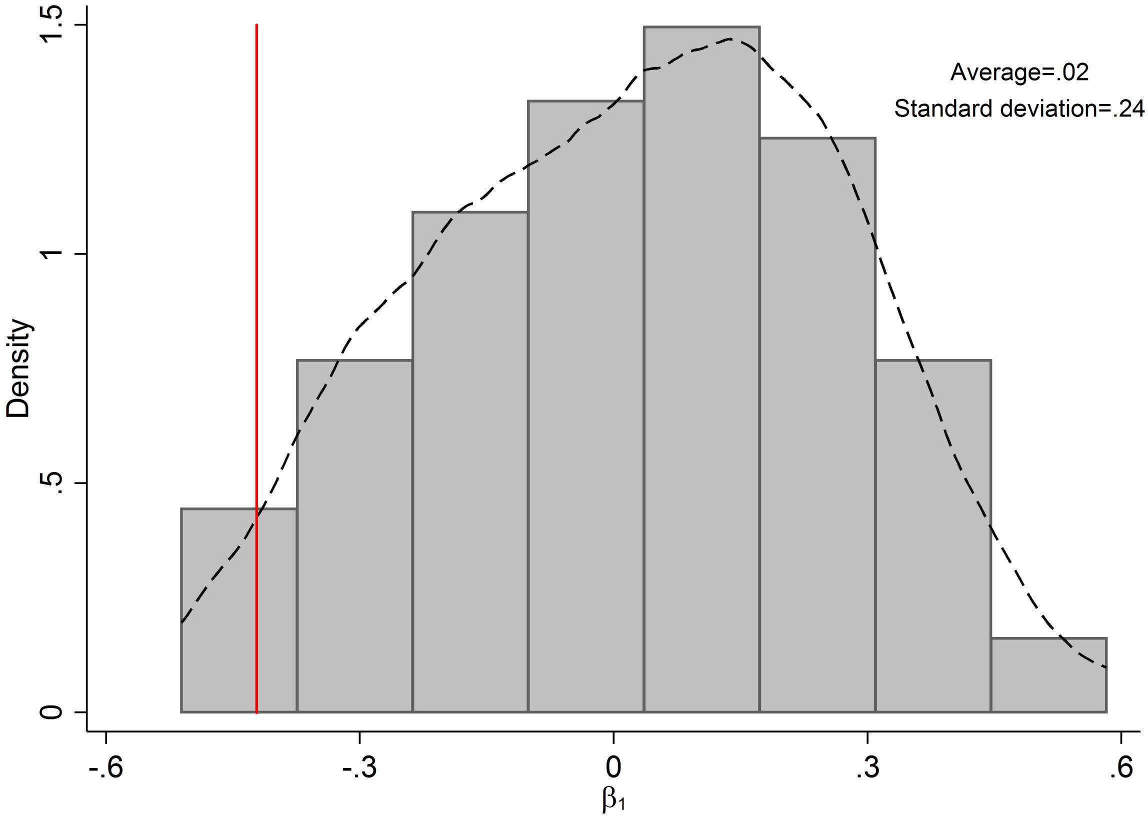
All regressions include course-year fixed effects. Column (1) and (2) use a bandwidth of 0.2 around 7, whereas column (3) and (4) use a bandwidth of 0.5. For the Probit analysis (bottom panel) the average marginal effects are displayed.

Figure 5: **Reduced Form Estimate Insensitive to Bandwidth Choice.**



Notes: This figure plots the reduced form estimates, and corresponding confidence interval, estimated with a third order polynomial against the bandwidth used to estimate the reduced form. The bandwidth ranges from 0.3 until 1.0, where 0.5 is preferred. The specifications include all control variables.

Figure 6: **Fake Cutoffs.**



Notes: Histogram for the reduced-form estimate at fake cutoffs. Cutoffs are chosen at every 0.005-points for GPA between 6.5 and 7.5, where the bandwidth for the estimation is 0.5. The vertical red line corresponds to the estimate at the true cutoff of seven. The specifications include all control variables.

Table 12: **Students Substitute Self Study for Attendance.**

	Attendance lectures	Lectures+ Tutorials+ Self study	Usefulness lectures	Quality lecturer(s)
	(1)	(2)	(3)	(4)
Average 1^{st}-year Grade is Below 7	0.355*	0.185	0.462	0.00289
	(1.91)	(0.28)	(0.76)	(0.01)
Intercept^{+}	0.523***	3.310***	3.313***	3.845***
	(3.82)	(5.56)	(8.60)	(22.03)
Observations	382	382	203	332
Adjusted R^2	0.055	0.264	0.058	0.168

Notes: t statistics in parentheses, standard errors are clustered on the student level

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

All regressions include course-year fixed effects. The bandwidth used is 1.0 around 7.

$^{+}$ The intercept is calculated via regressions which exclude course-year fixed effects - such that it approximates the outcome mean near the threshold.

Table 13: **Channels Other Than Substitution Effects.**

	Grades (Standardized)			
	(1)	(2)	(3)	(4)
Average 1^{st}-year Grade is Below 7	-0.413**	-0.434**	-0.433***	-0.407**
	(-2.51)	(-2.35)	(-2.71)	(-2.38)
Average 1^{st}-year Grade Among Peers		0.0847		
		(0.79)		
Their Interaction (Treatment×Peers)		0.0206		
		(0.13)		
Average Registration Time Among Peers			0.00169	
			(0.32)	
Its Interaction with Treatment			0.000539	
			(0.06)	
TA Gives Good Tutorials (Average)				0.00611
				(0.16)
Its Interaction with Treatment				-0.00627
				(-0.11)
Observations	1287	1287	1287	1121
Adjusted R^2	0.211	0.217	0.216	0.219

Notes: t statistics in parentheses, standard errors are clustered on the student level

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

All regressions include course-year fixed effects, except column (1) which includes tutorial fixed effects. The peer group average is the leave-out mean. For column (4) some observations are lost, because no students completed the course evaluation for a given TA. Bandwidth of 0.5 is used with a third order polynomial.

Table 14: **Policy has No Impact on Outcomes in the Third Year.**

	Thesis						Third Year Grades (std)	
	Grades (std)		Graduation Date		Graduated (1=yes, 0=no)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	A: Local linear regression							
Average 1^{st}-year Grade is Below 7	-0.312	-0.322	36.57	38.72	0.0404	0.0266	-0.160	-0.133
	(-1.48)	(-1.50)	(0.78)	(0.82)	(0.75)	(0.50)	(-1.29)	(-1.04)
Observations	284	284	284	284	312	312	1885	1885
Adjusted R^2	-0.002	-0.004	0.830	0.838	0.074	0.085	0.221	0.221
	B: Third order polynomial							
Average 1^{st}-year Grade is Below 7	-0.515*	-0.528*	15.91	21.29	0.0763	0.0645	-0.0788	-0.0627
	(-1.89)	(-1.95)	(0.27)	(0.37)	(1.05)	(0.92)	(-0.55)	(-0.44)
Observations	628	628	628	628	723	723	4288	4288
Adjusted R^2	0.007	0.008	0.852	0.856	0.168	0.178	0.253	0.253
Controls	No	Yes	No	Yes	No	Yes	No	Yes

Notes: t statistics in parentheses, robust standard errors
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Column (1) until (6) are on the student level and include year fixed effects, where column (7) and (8) are on the course level and include course-year fixed effects. The top panel uses a bandwidth of 0.2 around 7. The bottom panel uses a bandwidth of 0.5.

Table 15: **Policy has No Impact on Third Year Course Choice.**

	Course Choice	
	First order polynomial	Third order polynomial
	(1)	(2)
Fiscal Economics	0.005	-0.008
	(0.10)	(-0.14)
Management Accounting		0.101
		(0.48)
Financial Accounting	0.062	0.086
	(0.95)	(0.98)
Entrepeneurship	0.060	0.101*
	(1.32)	(1.76)
Marketing	-0.051	-0.037
	(-0.85)	(-0.60)
Urban and Transport Economics	-0.069	-0.079
	(-1.08)	(-0.95)
Applied Economics	-0.037	-0.00
	(-0.65)	(-0.01)
Financial Economics	0.224**	0.181
	(2.00)	(1.13)
Behavorial Economics	-0.018	-0.065
	(-0.64)	(-1.56)
Microeconomics	-0.059	-0.087
	(-1.06)	(-0.84)
Macroeconomics	-0.076	-0.140
	(-1.23)	(-1.56)
Econometrics	-0.0481	-0.053
	(-1.08)	(-0.84)
Observations	295	674
Pseudo R^2	0.064	0.054

Notes: z statistics in parentheses, robust standard errors
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The multinomial logit model is on the student level. The outcome variable is the major choice of the student (twelve options), which is explained by the treatment, a year fixed effect and a polynomial for $f(\cdot)$. Marginal effects are displayed, which can be interpreted as the effect of the policy on the probability of choosing a certain major.

Table 16: **IV Estimates. Impact of Attendance on Academic Performance.**

	Courses that follow University Policy		Course with more Elaborate Policy		Courses which are More Relaxed	
	(1)	(2)	(3)	(4)	(5)	(6)
			A: Local linear regression			
% Attendance	-1.199***	-1.180**	-41.99	-89.32	-0.100	0.162
	(-2.65)	(-2.44)	(-0.16)	(-0.07)	(-0.10)	(0.14)
Observations	551	551	748	748	853	853
Adjusted R^2	0.019	0.018	0.00	0.00	0.197	0.201
			B: Third order polynomial			
% Attendance	-0.905**	-0.929*	-0.639	2.003	0.311	0.425
	(-1.98)	(-1.85)	(-0.02)	(0.13)	(0.37)	(0.49)
Observations	1287	1287	1679	1679	1983	1983
Adjusted R^2	0.00	0.00	0.00	0.00	0.243	0.243
Controls	No	Yes	No	Yes	No	Yes

Notes: t statistics in parentheses, standard errors are clustered on the student level
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
All regressions include course-year fixed effects. Panels A and B have bandwidths of 0.2 and 0.5, respectively. The instrument is a binary variable that indicates whether the average first year grade was below 7.

Table 17: **Identities of Affected Students.**

	Attendance (% Tutorials Attended)				
	(1)	(2)	(3)	(4)	(5)
Average 1^{st}-year Grade is Below 7	0.336***	0.276***	0.335***	0.346***	0.301***
	(5.83)	(3.34)	(5.76)	(5.86)	(4.47)
Distance to University	-0.040**				
	(-2.43)				
Distance×Treatment	0.044***				
	(2.61)				
Low Tuition Fee		-0.063			
		(-1.01)			
Low Fee×Treatment		0.065			
		(1.05)			
Age			0.027**		
			(2.05)		
Age×Treatment			-0.030**		
			(-2.31)		
Female				0.031	
				(0.89)	
Female×Treatment				-0.030	
				(-0.84)	
High School Grade					-0.021
					(-1.10)
High School Grade× Treatment					0.031
					(1.59)
Observations	1275	1275	1275	1275	1057
Adjusted R^2	0.417	0.412	0.413	0.412	0.426

Notes: t statistics in parentheses, standard errors are clustered on the student level
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
All regressions include course-year fixed effects and controls and are estimated on the (optimal) bandwidth of 0.5 with a third order polynomial. Exact nationality is unknown for six students and grades for secondary school is only observed for Dutch students (column (5)).

Table 18: **Estimates Elasticity of Substitution.**

	$ln(\frac{Tutorials}{SelfStudy})$			
	(1)	(2)	(3)	(4)
Distance	0.0847	0.105	0.105	0.0980
	(1.13)	(1.25)	(1.23)	(1.14)
Distance × Age	0.309***	0.227**	0.175**	0.169*
	(3.24)	(2.59)	(1.97)	(1.89)
Observations	382	382	382	382
Adjusted R^2	0.017	0.192	0.208	0.205
Controls	No	$C_{jc}^{(2)}$	(2) & age	(3), gender & nationality

Notes: t statistics in parentheses, standard errors are clustered on the student level

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The bandwidth used is 1.0 around 7. Column (1) includes no other controls, column (2) adds course-year fixed effects (and a constant), column (3) includes course-year fixed effects and age, and column (4) includes the controls in column (3) plus nationality and gender.

Appendix

Table A.1: **No Sample Selection when Matching Grades with Attendance.**

	Grades (standardized)				Matched	
	(1)	(2)	(3)	(4)	(5)	(6)
Matched	-0.0534	0.0219	0.492	0.625**		
	(-0.92)	(0.55)	(1.18)	(2.27)		
Average 1^{st}-year Grade is Below 7			-0.136	-0.139	0.00161	0.00110
			(-1.01)	(-1.02)	(1.05)	(0.46)
Their Interaction (Matched×Treatment)			-0.0310	-0.0334		
			(-0.30)	(-0.47)		
Observations	2314	5345	2314	5345	2314	5345
Adjusted R^2	-0.000	-0.000	0.168	0.210	0.994	0.984
Bandwidth	0.2	0.5	0.2	0.5	0.2	0.5
Polynomial	-	-	1^{st}	3^{rd}	1^{st}	3^{rd}

Notes: t statistics in parentheses, standard errors are clustered on the student level
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Matched is a variable which equals 1 if the grade record found a match with the attendance data and 0 otherwise. As such, column (1) and (2) regress second year grades upon a constant and the matched-variable and shows that grades are similar for matched and nonmatched records. Subsequently, column (3) and (4) show the reduced-form effect is not different between matched and nonmatched records (Matched×Treatment). The final two columns regress the matched-variable upon scoring below 7 in the first year and thereby show the policy is unable to explain whether or not a record is matched. Columns (3) until (6) include course-year fixed effects.

Table A.2: **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)
Lecturer has good command of English	1-5	Quality lecturer(s)
Lecturer can be easily contacted	1-5	Quality lecturer(s)
Lecturer provides sufficient assistance	1-5	Quality lecturer(s)
TA gives good tutorials	1-5	Quality TA
TA can be easily contacted	1-5	Quality TA
TA provides sufficient assistance	1-5	Quality TA
TA has good command of English	1-5	Quality TA

Notes: The measurement for every question reflects a Likert scale, where 1 equals strongly disagree and 5 equals strongly agree, with the two exceptions being total study time (1 being 0 hours, 2 being $[1-5]$ hours, 3 being $[6-10]$ hours and 10 being ≥ 40 hours) and lecture attendance (1 being yes and 0 being no). Within a category most questions show a correlation of above 0.6. We simply take the mean for every question within a category and thereby ignore potential missing values within a category. The more sophisticated approach of calculating the principal components gives qualitatively similar results. In fact, the rowmeans and corresponding principal components all show correlations of above 0.97. However, the principal component analysis strongly reduces the amount of observations due to the missing values within categories.

Table A.3: **Similarity of Students around the Cutoff.** In secondary school grades.

	Grades (Standardized)							
	Dutch	English	Economics	General science	Civic education	History	German	Man. and org.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	A: Local linear regression							
Average 1^{st}-year Grade is Below 7	-0.247	-0.179	0.0918	-0.0128	0.208	0.0788	0.423	0.0582
	(-0.96)	(-0.67)	(0.43)	(-0.05)	(0.79)	(0.26)	(1.49)	(0.20)
Observations	241	234	228	232	229	187	165	154
Adjusted R^2	-0.001	0.004	-0.003	-0.018	0.005	-0.030	0.020	-0.001
	B: Third order polynomial							
Average 1^{st}-year Grade is Below 7	-0.324	-0.190	0.166	0.0385	0.424	0.0394	0.574	0.183
	(-0.82)	(-0.50)	(0.53)	(0.09)	(0.97)	(0.09)	(1.39)	(0.43)
Observations	578	560	554	552	543	457	393	381
Adjusted R^2	-0.004	0.004	0.013	-0.007	0.011	-0.006	0.035	0.007
	C: Third order polynomial							
Average 1^{st}-year Grade is Below 7	-0.259	-0.131	-0.0290	0.0308	0.0249	0.125	0.191	-0.0319
	(-1.20)	(-0.63)	(-0.15)	(0.13)	(0.10)	(0.63)	(0.87)	(-0.12)
Observations	1144	1113	1095	1087	1075	917	809	752
Adjusted R^2	0.037	0.035	0.096	0.036	0.046	0.051	0.083	0.078

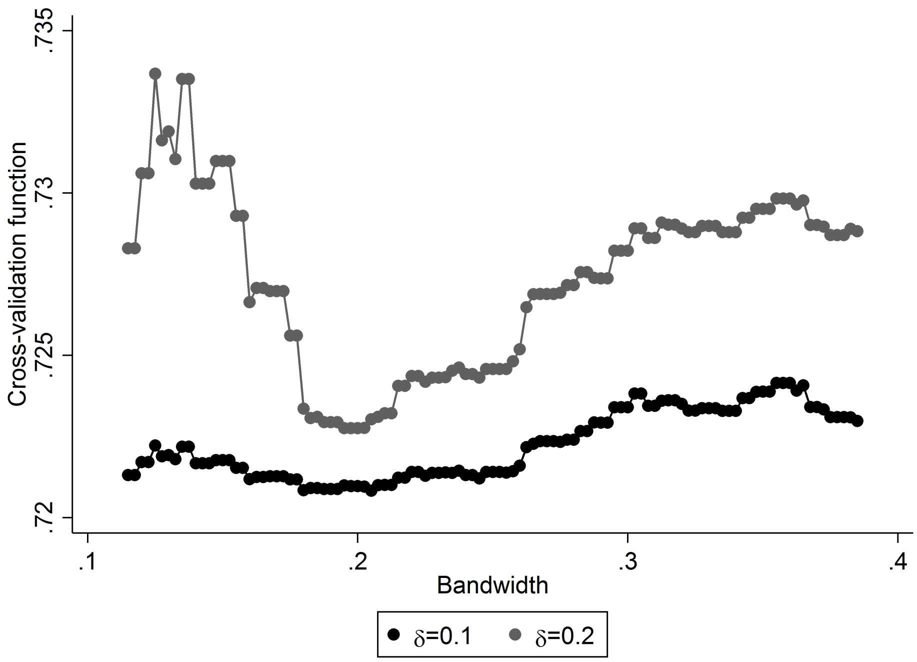
Notes: t statistics in parentheses, robust standard errors
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
All regressions are on the student level and include year fixed effects. Panel A displays local linear regressions with the optimal bandwidth of 0.2 around 7, Panel B shows regressions for the optimal bandwidth of 0.5 with the third order polynomial, and Panel C includes all observations. Observations differ per column, as not all students have followed the same courses.

Table A.4: **No Bunching Just Above 7.** Tested through the method proposed by McCrary [2008].

	Counts of Number of Students		
	Local linear regression	Second order polynomial	Third order polynomial
	(1)	(2)	(3)
	A: Binsize as suggested by McCrary [2008]		
Average 1^{st}-year Grade is Below 7	0.000444	-0.00186	-0.00302
	(0.08)	(-0.37)	(-0.43)
Observations	20	50	50
Adjusted R^2	0.142	0.197	0.192
	B: Bins two times as small		
Average 1^{st}-year Grade is Below 7	0.0000447	-0.00109	-0.00206
	(0.01)	(-0.36)	(-0.51)
Observations	40	100	100
Adjusted R^2	0.004	0.081	0.073
	C: Bins four times as small		
Average 1^{st}-year Grade is Below 7	-0.0000204	-0.000582	-0.00109
	(-0.01)	(-0.36)	(-0.51)
Observations	80	200	200
Adjusted R^2	-0.008	0.028	0.023

Notes: t statistics in parentheses, robust standard errors
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The local linear regression is estimated on the optimal bandwidth of 0.2 around 7, whereas the second- and third order polynomial is estimated on the optimal bandwidth of 0.5. The panels refer to the different binsize as to compute the histogram for the number of students. Panel A uses the plug-in estimate of McCrary [2008], panel B and C subsequently undersmooth and compute bins two and four times as small respectively. Results are robust to the binsize.

Figure A.1: **Selection of Optimal Bandwidth for the Local Linear Regression.**



Notes: The figure plots the cross-validation function against the bandwidth size. We follow Imbens and Lemieux [2008], and estimate separate regressions on either side of the cutoff with a linear polynomial for a given bandwidth (b): $G^{(2)}_{ijc} = \beta_0 + \beta_1(\bar{G}^{(1)}_{ic} - 7) + \epsilon^{(2)}_{ijc}$. Let the predicted values on the left and right side of the cutoff be denoted by $\widehat{G^{(2),l}_{ijc}(\bar{G}^{(1)}_{ic})}$ and $\widehat{G^{(2),r}_{ijc}(\bar{G}^{(1)}_{ic})}$ respectively. Subsequently we stack predicted values from both sides of the cutoff and define the cross-validation criterion as the mean squared prediction error: $CV(b) = \frac{1}{N}\sum_{(i,j,c)} \big(G^{(2)}_{ijc} - \widehat{G^{(2)}_{ijc}(\bar{G}^{(1)}_{ic})}\big)$. The optimal bandwidth b is the one that minimizes this function: $b^* = \arg\min_b CV(b)$. As we are interested in the performance of the regression function near 7, we discard observations from the tails while minimizing the criterion function. In particular, we minimize the criterion while only taking into account students within the absolute distance of δ from the cutoff: $CV(b) = \frac{1}{N}\sum_{(7-\delta)\leq\bar{G}^{(1)}_{ic}\leq(7+\delta)} \big(G^{(2)}_{ijc} - \widehat{G^{(2)}_{ijc}(\bar{G}^{(1)}_{ic})}\big)$. The figure shows results for δ equal to 0.1 and 0.2, which roughly represent 10 and 20 percent of the observations at both sides of the cutoff. Where the criterion (prediction error) is shifted downwards as whole for the smaller δ, both have their minimum at a bandwidth of 0.2.

Table A.5: **Calculations of the Optimal Bandwidth Using Methods of Calonico et al. [2016].**

		First order polynomial		Third order polynomial	
		Left of 7	Right of 7	Left of 7	Right of 7
Mean squared error	MSErd	0.233	0.233	0.414	0.414
	MSEtwo	0.278	0.432	0.447	0.639
	MSEsum	0.332	0.332	0.492	0.492
	MSEcomb1	0.233	0.233	0.414	0.414
	MSEcomb2	0.278	0.332	0.447	0.492
Coverage error rate	CERrd	0.147	0.147	0.248	0.248
	CERtwo	0.176	0.273	0.269	0.383
	CERsum	0.210	0.210	0.296	0.296
	CERcomb1	0.147	0.147	0.248	0.248
	CERcomb2	0.176	0.210	0.269	0.296

Notes: The table displays estimates for the optimal bandwidth size for both the local linear regressions and the third order polynomial, while using the methods as described in Calonico et al. [2016]. For the local linear regression the result corresponds with the (simple) cross-validation method of Figure A.1, the optimal bandwidth seems to hover around 0.2. For the third order polynomial the optimal bandwidth is between 0.4 and 0.6 for the MSE methods, while being significantly smaller for the CER methods. As such, for the third order polynomial we start out with a bandwidth of 0.5, but check for robustness.

Table A.6: **Results Regression Discontinuity (All 8 Eligible Courses).**

	Local linear regression		Third order polynomial	
	(1)	(2)	(3)	(4)
	A: Forced Attendance Decreases Grades			
Average 1^{st}-year	-0.152	-0.139	-0.158	-0.151
Grade is Below 7	(-1.61)	(-1.47)	(-1.30)	(-1.25)
Adjusted R^2	0.165	0.166	0.209	0.209
	B: Forced Attendance Increases Attendance			
Average 1^{st}-year	0.130***	0.119***	0.152***	0.146***
Grade is Below 7	(4.64)	(4.51)	(4.32)	(4.27)
F-test	21.53	20.34	18.66	18.23
Adjusted R^2	0.305	0.318	0.308	0.312
	C: IV Estimates			
% Attendance	-1.130	-1.113	-0.544	-0.477
	(-1.44)	(-1.27)	(-0.62)	(-0.55)
Observations	2152	2152	4949	4949
Adjusted R^2	0.084	0.086	0.00	0.00
Controls	No	Yes	No	Yes

Notes: t statistics in parentheses, standard errors are clustered on the student level

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

All regressions include course-year fixed effects. We estimate the local linear regression on the preferred bandwidth of 0.2 around 7 and the third order polynomial on the preferred bandwidth 0.5.

Table A.7: **Results of Local Linear Regressions.** Restricting the polynomial to be similar on both sides of cutoff.

	All Courses		Courses that follow the University Policy		Courses with more Elaborate Policy		Courses which are More Relaxed	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				A: Forced Attendance Decreases Grades				
Average 1^{st}-year Grade is Below 7	-0.162*	-0.151	-0.367***	-0.346***	-0.164	-0.178	-0.0251	-0.00103
	(-1.74)	(-1.61)	(-3.04)	(-2.84)	(-1.33)	(-1.42)	(-0.20)	(-0.01)
Adjusted R^2	0.165	0.165	0.179	0.178	0.094	0.094	0.201	0.200
				B: Forced Attendance Increases Attendance				
Average 1^{st}-year Grade is Below 7	0.131***	0.121***	0.304***	0.292***	0.00227	0.000420	0.127***	0.109***
	(4.40)	(4.30)	(5.93)	(5.88)	(0.14)	(0.03)	(2.96)	(2.69)
F-test	19.36	18.49	35.16	34.57	0.02	0.00	8.76	7.23
Adjusted R^2	0.305	0.318	0.365	0.375	0.150	0.175	0.159	0.186
				C: IV Estimates				
% Attendance	-1.234	-1.252	-1.208***	-1.186***	-72.20	-423.6	-0.198	-0.00944
	(-1.56)	(-1.45)	(-2.77)	(-2.59)	(-0.14)	(-0.03)	(-0.20)	(-0.01)
Observations	2152	2152	551	551	748	748	853	853
Adjusted R^2	0.035	0.032	0.008	0.013	0.00	0.00	0.188	0.200
Controls	No	Yes	No	Yes	No	Yes	No	Yes

Notes: t statistics in parentheses, standard errors are clustered on the student level
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
All regressions include course-year fixed effects. We estimate the local linear regression on the preferred bandwidth of 0.2 around 7.

Table A.8: **The Absence of Other Channels.** The remaining categories of course evaluations are regressed upon the policy.

	General	Structure	Fairness	Quality TA
	(1)	(2)	(3)	(4)
Average 1^{st}-year Grade is Below 7	-0.0785	0.108	0.417	0.436
	(-0.25)	(0.39)	(1.42)	(1.14)
Intercept^{+}	3.915***	3.999***	3.449***	3.893***
	(20.47)	(18.18)	(14.39)	(12.99)
Observations	382	382	339	344
Adjusted R^2	0.080	0.185	0.408	0.018

Notes: t statistics in parentheses, standard errors are clustered on the student level

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

All regressions include course-year fixed effects. The bandwidth used is 1.0 around 7.

$^{+}$ The intercept is calculated via regressions which exclude course-year fixed effects - such that it approximates the outcome mean near the threshold.