The Effect of Compulsory Tutorial Groups on Student Performance

Matthijs Oosterveen* Sacha Kapoor† Dinand Webbink‡

May 20, 2016

FIRST DRAFT, PLEASE DO NOT CITE OR REDISTRIBUTE

Abstract

This paper exploits a natural experiment at the Erasmus University Rotterdam to evaluate the effectiveness of small-scale teaching groups that are typically used at universities in undergraduate courses. Economic Bachelor students who had a GPA below a seven in their first Bachelor year, were obliged to attend 70% of the tutorials per course in their second year. We use a regression-discontinuity design to identify the intention-to-treat effect of the policy as well as a well-defined local average treatment effect of actual attendance on grades. We find that the policy of mandatory tutorials caused an increase in attendance of second year Bachelor students of roughly 20%. Remarkably, point estimates for the intention-to-treat are consistently negative, but small, and in some cases statistically significant. In addition, we find a negative effect of attendance upon grades, sometimes statistically significant.

*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

1 Introduction

This paper exploits a natural experiment to evaluate the effectiveness of the small-scale tutorial groups that are typically used at universities in undergraduate courses. At a majority of the universities, lectures for undergraduate courses are given to large groups of students, where the basics of the material is explained. Subsequently students are divided into smaller groups, the tutorial groups, where there is the opportunity to discuss the material more in depth: making exercises, discussing problem sets and asking questions. This is also true for the Erasmus University Rotterdam, which is ranked consistently amongst the top universities of continental Europe.[1] The purpose of this small-scale teaching is mainly to improve the learning process of students, but it is also used to to attract potential students. Considerable costs are made for this small-scale teaching, The National Center for Educational Statistics estimated that American colleges and Universities spend 32 billion dollar on student instruction alone, while their effectiveness is not obvious for several reasons. First of all, attending the tutorial groups might serve as a substitute for self-study. If self-study is more effective, tutorials could actually have a potential negative effect. After having attended the lectures, the marginal benefits of attending a tutorial session on similar material might be so low that grades are unaffected. Second, and related, students experience opportunity cost while traveling to university and attending the tutorial groups, time they could have spend different otherwise. Third, tutorial groups are mostly taught by senior students. To the extent that they do not fully grasp the material themselves, it is unclear whether they can actually clarify the material for the students. This argument seems especially relevant, as in the context of of primary education it has been shown that the quality of the teacher matters for the pupils' outcomes.

A large empirical literature has attempted to evaluate the effect of attendance in class on grades and they have all found a positive relationship between class attendance and performance. However, most of this research is plagued with basic endogeneity issues that come with the use of observational data. It is very difficult to disentangle the effect of attending a class from motivation. More motivated students are likely to attend more classes, as motivation shows a positive association with grades, an OLS-estimate of grades upon attendance is likely to be biased upwards.

[1] Worldwide it is ranked on the 126^{th} place in the QS university overall ranking and on the 40^{th} position on the subject Economics & Econometrics (QS Ranking, 2015).

We exploit a natural experiment present at the Economics and Business Economics Bachelor at the Erasmus University Rotterdam to evaluate the effectiveness of the tutorial groups. Economic Bachelor students who had a GPA below a 7 in their first Bachelor year, were obliged to attend 70% of the tutorials per course in their second Bachelor year. Skipping more than 30% of the tutorials resulted in exclusion from the exam. Students with a GPA of a 7 or higher in their first Bachelor year, did not have to attend the tutorials, but were free to do so. This rule was in place for five academic years, from 2009-2010 until 2013-2014. It allows one to estimate the average of students' second year grades just above and below the cut-off to find the causal impact of this compulsory-schooling policy. More formally, we will use a regression-discontinuity design to identify the intention-to-treat effect of the policy as well as a well-defined local average treatment effect of actual attendance on grades. The identifying assumption is the inability of individuals to imprecisely control the average grade of the first year (see Lee (2008) and Lee and Lemieux (2010)). Even if some individuals have a high ex ante probability to locate themselves near the threshold, these individuals will have the same probability of having a score just above or below the threshold, which causes local randomization. The ex ante probability determines the contribution of each individual to the measured treatment effect (Lee and Lemieux, 2010).[2] We show that all baseline characteristics are continuous and there is no discontinuity of the density of number of students at the threshold

We find that the policy of mandatory tutorials caused a strong significant jump in attendance of second year Bachelor students. Depending upon the courses, the jump in attendance ranges from roughly 12% to 30%, where with the more technical courses the policy is less binding. However, point estimates for the intention-to-treat are consistently negative, but small, and in some cases significantly so. When we treat the cut-of as an instrument for second year attendance, we (mechanically) find a larger negative effect of attendance upon grades, in some cases significant. Consistently with this negative effect of the compulsory tutorials on grades, we find a larger negative effect on grades for (clusters of) courses that show a larger positive jump in attendance around the threshold.

Our results indicate that previous studies have overestimated the effect of attendance upon grades. Considerable costs are made on student instruction, whereas we can find

[2] As such, note that the RD gap is not only informative for the observations right around threshold.

no evidence of a positive effect of tutorial groups on grades. In fact, in all our specifications we find consistently negative point estimates, mostly significant so.

The paper proceeds as follows. In Section 2 we explain the setting of the natural experiment, in Section 3 we briefly explain the data, Section 4 outlines the empirical strategy and Section 5 provides several tests of the identifying assumption. Section 6 and Section 7 discuss and interpret the results and Section 8 concludes.

2 The Context

We exploit a natural experiment present at the Economics faculty of the Erasmus University Rotterdam. Since the Bologna Accords in 1999, universities in Europe are following a Bachelor-Master system. A Bachelor program takes on three to four years, and all students within the same Bachelor follow the same courses, with some degrees of freedom at the end of the Bachelor. After finishing the Bachelor program, students can specialize and they are allowed to follow one of the various Master programs.[3] This system also goes for the the Erasmus University Rotterdam, where the Department of Economics is a supplier of an Economics and Business Economics Bachelor. In the academic year 2013-2014, the economics Bachelor had an influx of approximately 800 students, whereas in total N_2 students started an economics Bachelor in the Netherlands. The first two years of the Bachelor program, all students follow the same 10 courses per year, where they are being taught the basics in economics (micro and macro) business economics (marketing and accounting) and econometrics (mathematics and statistics). In the third Bachelor year, students can choose their own minor and major and subsequently they can continue this specialization by choosing a Masters program. The economics Bachelor is both provided in Dutch and English, both programs know the exact same structure and courses. The difference is that the Dutch version of the economics Bachelor solely attracts Dutch students and is roughly 2.5 times the size of the English version.

The first two Bachelor years follow the same structure. An academic year is divided into five education blocks, of eight weeks each (seven weeks of teaching and one week of exams). In each block the students follow a small and a large course, from which the

[3]An undergraduate program is comparable with a Bachelor, whereas a graduate program is the equivalence of a Master.

student can earn 4 and 8 ECTS respectively.[4] In both Bachelor years, courses are being taught through lectures and small-scale tutorial groups. There are three (two) lectures per week for the 8 (4) ECTS course and they are taught by (assistant-)Professors. Attendance is voluntarily, which causes there to be a high standard deviation in the number of students attending per course and lecture.[5] Subsequently, there are two (one) tutorial groups per week for the 8 (4) ECTS course, where assignments of that weeks material are discussed. The tutorial groups are mostly given by senior students, which are sometimes PhD-students, and last for 1 hour and 45 minutes. See Table A.1 for an overview of the specific courses in the first and second Bachelor year. Table A.1 also shows that the University classifies the first- and second-year courses into three clusters: Economics (A), Business Economics (B) and Econometrics (C). This 'natural' classification will be used to identify the effect of the mandatory-tutorials policy for different type of courses.

In the first Bachelor year students have to attend 70% of the tutorials, otherwise they are excluded from making the exam. As there is some freedom for a student in deciding to attend a tutorial, there is also variation in attendance. Using this variation to explain grades would be very likely to cause an upwards biased estimate, as motivation as an omitted variable positively correlates with both attendance and grades. However, in the second year of the Bachelor, students have to attend 70% of the tutorials unless they meet two requirements: (i) GPA of a 7 or higher in the first year *and* (ii) completed the whole first Bachelor year within the first year (passed all 10 courses).[6] All the other students, (i) who had a GPA lower than a 7 *or* (ii) did not complete the whole first Bachelor year within the first year, have to attend 70% of the tutorials. Table 1 gives an overview of the groups of students that have to comply with the treatment. Students were exempted from making the exam if he or she did not comply. This mandatory-tutorials policy was in place for five academic years, it started in 2009 and lasted until 2013. Students had to register themselves for the tutorials and were able to choose between the tutorial groups

[4]ECTS is an abbreviation for European Transfer Credit System, which is a common measure for student performance to accommodate the transfer of students and grades between European Universities. One ECTS roughly equals 28 hours of studying.

[5]This is anecdotal evidence, as there is no data on attendance at the lectures.

[6]The classification of courses within clusters are important for determining whether a student completed the first Bachelor year, as a student is able to compensate an insufficient grade between 4.5 and 5.4 if the grades for the other courses within the cluster are sufficient. Courses of the second year are also classified within clusters, as this compensation rule is also applicable for the second year. Students are not affected differently on either side of the threshold, as such this is not a problem for our identification strategy. On the contrary, this natural classification of courses is a possibility to study the heterogeneity of the policy among courses (see Table A.1 for an overview of the courses and classification).

with different time slots. At the time of registration, students did not know up front which student-assistant taught at which tutorial group and they were not able to change group after the registration period ended. A consequence is that every course is followed with a different set of peers.

Table 1: Overview of treatment

	GPA	GPA< 7	GPA $\geq$ 7
Completed Bachelor 1			
YES		✓	
NO		✓	✓

Notes: A ✓means a student has to comply with the mandatory-tutorials policy within the course.

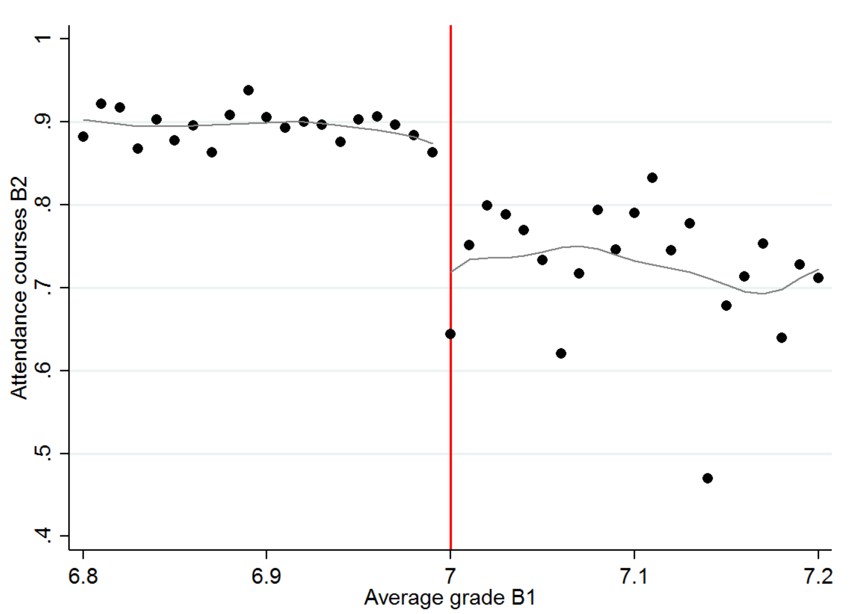
Taking this policy at face value, it ‘randomly assigns’ mandatory schooling of more than 3.5 (26) hours per week (block) near the threshold.[7] However, as students with a GPA above a 7 were allowed to go to the tutorials, effectively this policy induced less exogenous variation in attendance. Figure 1 below shows the jump in attendance near the threshold, where students just below a GPA of seven attend roughly 15%-point more tutorials than their peers with a GPA just above a seven. This corresponds to roughly two (one) extra tutorials for an 8 (4) ECTS-course, which is more than 5.5 hours of schooling per block.[8] Following Figure 1, our main approach is to rank students according their GPA in the first Bachelor year, as to compare students (just) above and below the threshold of a seven, and to use the discontinuity to find the causal effect of attendance on grades.

Table 1, however, shows there is a *two-by-two treatment-control set-up* due to the two requirements (GPA and complete Bachelor one). This has three consequences. First of all, with our main approach we should limit our sample to students that completed their Bachelor one within the first year. As students that scored a GPA above a seven, but did not complete their first year, still had to comply with the 70%-attendance rule. Second, this two-by-two set-up provides us with an extra test for the validity of our identification

[7] There are 3 tutorials of 1.75 hour per week, 5 education blocks of each 7 weeks and overall students have to attend 70% of the tutorials. This makes $3 * 1.75 * 0.7 = 3.675$ hours per week, $3.675 * 7 = 25.725$ hours per block and $25.725 * 5 = 128.625$ hours per year.

[8] As there are $7 * 3 = 21$ tutorials per block, this corresponds to $21 * 0.15 = 3.15$ extra tutorials. As one tutorials lasts 1.75 hours, this causes $3.15 * 1.75 \approx 5.5$ extra hours of schooling per block.

Figure 1: Visual representation of first-stage



Notes: This figure is based upon students that completed Bachelor one within the first year. For chosing the bandwith we specify a regression of grade in second year on K' bin dummies and $2K'$ bin dummies (reduced-form equation). As the first model is nested in the latter, we can perform an F-test. If the null hypothesis is not rejected, we choose the model with K' bin dummies. We start with 2 bin dummies and stop at the first number of bin dummies we do not reject the null hypothesis. We stop rejecting at a bandwith of 0.0625, the figure is based on a bandwith of 0.05. There is little harm in undersmoothing, as this shows the raw variation in the data. The size of the dot indicates the amount of observation used to calculate the mean.

strategy. That is, we could compare students just above and below the cut-of that did not complete their Bachelor 1. As both groups are treated, one would expect the attendance rate to be above 70%, without a discontinuity around the threshold. Moreover, bunching would be revealed by significant differences in background characteristics near the threshold. One might argue this test is unlikely to reveal potential bunching of the group of interest (which are students who did complete their Bachelor one), as students that are very unlikely to complete their Bachelor one within the first year do not have an incentive to be above the threshold of a seven. Indeed, this latter group has to comply with the mandatory tutorials independent of their GPA-score. Therefore we will limit this sample to the students who had the potential, but did not manage, to finish their Bachelor one within the first year. We define this sample to be the group of students that failed to complete the Bachelor one in the first year because they scored an insufficient grade for one of the ten courses. We show our results are not sensitive to the definition of the choice of this control-sample.[9]

Third, to exploit the full potential of the two-by-two set-up and to increase the precision of our estimates, we can use a local difference-in-difference design (DiD). In the analogy of a DiD, the treatment group are the students who completed their first Bachelor year, where the control group are students who did not. Subsequently the GPA-score represents whether the students is actually observed being in the treatment or control (this represent the time-variable in the DiD, GPA<7 means $T=1$ and GPA≥ 7 means $T=0$). This allows us to compare, near the threshold, the change in grades of the students who completed their Bachelor one with the change in grades of the students who did not complete their Bachelor one. Intuitively, this removes any potential bias that would remain near the threshold. Potential unobserved differences between students around the threshold within the treatment group are filtered out by the comparison with the control group around the threshold. Also here, we can limit the sample of the control group to the students who had the potential to finish their Bachelor one but did not succeed in doing this, as to increase the comparability between the treatment and control group. In other words, this limited group also had an incentive to select themselves just above the threshold and thus also corrects for it.

3 Data

We use administrative data provided by the faculty of the Economics department of the Erasmus University Rotterdam. Between 2008 and 2014, we observe course participation and grades on the student-level for all Bachelor years. This provides us with information on students' GPA in the first Bachelor year and whether they actually completed their first Bachelor year, such that we can determine where the student is located within Table 1. We observe the grades of the students in their subsequent Bachelor years, such that we can identify the effect of the mandatory tutorials policy (intention-to-treat effect). We

[9]There are some degrees of freedom on how to define the group of students who had the potential, but did not manage, to finish their first Bachelor year. This is due to the maximum of three resits within the resit period, which is at the very end of the academic year (and is also referred to as block six) and due to the fact of the five education blocks and the possible continuous upgrading of students' belief on the possibility of completing Bachelor one with a GPA of a seven. Therefore, students who failed to complete their Bachelor one because of one course seemed to be the most transparent and practical approach.

have several background characteristics, such as nationality, age, gender, performance on secondary school and living address which we will use as a check for the identification strategy and to reduce the variance of the error term.

To be able to measure the causal effect of actually attending the tutorial groups requires data on attendance. For every tutorial session of a first and second year Bachelor course, attendance is registered through handing out an attendance list on which students can indicate their presence by their signature. Subsequently, the student assistant has the task to update the attendance on an online university-portal. This serves two purposes. First, students can be excluded from making the exam if attendance is below 70% and second the student is able to visually inspect how many tutorials he has been attending. In theory, using this online-stored attendance data gives us the attendance rate for every student for every course he has attend in the first and second year of the Bachelor. While merging these two sources of data, we experienced a match of 94% for our sample. For 6% of the courses in our sample for which we have a valid grade, we do not have data on attendance. To keep the sample constant, we work with this 94% of the whole sample throughout the paper. The intention-to-treat effect does not qualitatively differ for all our analysis. We construct the main attendance-variable on the student level, which is simply the percentage of tutorials attended per course.

To the extent that the administrative data contains no measurement error, the attendance rate is likely to have some for two reasons. First, one is able to let a fellow-student sign the attendance list for him or her. This type of fraud would mostly (or even only) be committed by the group of students who have to comply with the mandatory-tutorials policy, as only this group has an incentive to be present. Second, and perhaps more important, students that do not have to comply with the mandatory-tutorials policy have zero incentive to put their signature on the attendance list. Both reasons cause the effect of attendance on grades to be underestimated. To see this, note that the effect of attendance upon grades will be estimated through an IV procedure, where a simple Wald-estimate equals: $\frac{E[G|D=1]-E[G|D=0]}{E[A|D=1]-E[A|D=0]}$, where G is the grade, A is the attendance rate and $D = 1$ if one has to comply with the 70%-attendance rule. The first reason would lead to overestimating the percentage of tutorials attended by students who had to comply with the mandatory-tutorials policy ($E[\widehat{A|D=1}] > E[A|D=1]$). The second reason would lead to underestimating the percentage of tutorials attended by students who did not have to comply with the mandatory-tutorials policy ($E[\widehat{A|D=}0] < E[A|D=0]$). Both cause the effect of the mandatory tutorials policy on attendance to increase, which increases the denominator and thereby decreases the Wald-estimate. However, for two reasons we believe measurement error is not that big of a problem as to drive our results. First of all, student-assistants have the task to prevent the type of fraud mentioned above.[10] Second, in our data we do observe that students who did not have to comply with the mandatory-tutorials policy are in fact attending roughly 40% to 70% of the tutorials. Nevertheless, to the extent that the attendance data contains some measurement error, this would cause the effect of attendance upon grades to be underestimated.

4 Empirical Design

The GPA-threshold provides us with a regression discontinuity design that we are able to exploit via a local difference-in-difference. Consider the grade of student i on course j in year t (G_{ijt}), which we consider to be a function of a constant, the variable D_{it}, which equals 1 if the GPA of student i in year t is below a seven, a polynomial expansion of the original assignment score (GPA_{it}), a vector of control variables for student i (x_i) and an error term (ϵ_{ijt}). We also control for course-year interaction-dummies (C_{jt}), as this leads to the relevant comparison.[11] To estimate the effect of the mandatory-tutorial policy (intention-to-treat), we can estimate the following equation by OLS for students who completed their Bachelor 1 within the first year:

$$G_{ijt} = \beta_0 + \beta_1 D_{it} + f(GPA_{it} - 7) + f(GPA_{it} - 7)D_{it} + C_{jt} + x_i^{'}\beta + \epsilon_{ijt} \qquad (1)$$

The main identifying assumption to get an unbiased estimate of the mandatory-tutorials policy is that the bias cancels in the limit, or more formally $\lim_{GPA_{it}\uparrow 7} E[\epsilon_{ijt}|GPA_{it}] - \lim_{GPA_{it}\downarrow 7} E[\epsilon_{ijt}|GPA_{it}] = 0$, such that a comparison of means close to the cut-of results in the causal estimate β_1. Indeed, in the absence of the treatment individuals around the cut-of would have been similar. Increasing the interval around the cut-of will bias the estimate of the treatment effect, especially if GPA is itself related to the outcome variable. However, if we are willing to make an assumption about this relationship ($f(\cdot)$), then

[10] In particular, after talking with several student assistants, we found out that quite a few of them count the number of students present in the room, to make sure this number coincides with the number of signatures on the attendance list.

[11] Note that the Dutch and English Economics program are taught via separate courses, hence this interaction term causes us to only compare students within the same program.

we can use more observations and extrapolate from above and below the cut-of point to what a tie-breaking randomized experiment would have shown (Van der Klaauw, 2002). We use the observations of the running variable between XXX and XXX. To select $f(\cdot)$, which is allowed to be different on either side of the threshold, we estimate Equation (1) while adding K' bin dummies and excluding the controls x_i'. Subsequently we include higher-order polynomials until the bin dummies are jointly insignificant. Jointly insignificant bin dummies provides evidence for (i) the polynomial being a good fit and (ii) there to be no other discontinuities in the data. To check for the sensitivity of our results, we follow Imbens and Limieux (2008) in estimating a local linear regression with a rectangular kernel. We select the bandwidth for the local regressions via the rule-of-thumb method as described in Lee and Lemieux (2010).[12]

Subsequently, and perhaps more interesting, we estimate a causal effect of actual attending the tutorial groups on grades. Here we opt for a fuzzy regression discontinuity design, where in Equation (2) the mandatory-tutorials policy (D_{it}) is used as an instrument for the percentage of tutorials attended (AT_{ijt}). Around the cut-of the mandatory-tutorials policy is a locally valid instrumental variable, provided that it strongly associates with attendance. Subsequently, in Equation (3) this exogenous variation in attendance is used to measure the effect on grades:

$$AT_{ijt} = \gamma_0 + \gamma_1 D_{it} + f(GPA_{it} - 7) + f(GPA_{it} - 7)D_{it} + C_{jt} + x_i^{'}\gamma + \eta_{ijt} \tag{2}$$

$$G_{ijt} = \alpha_0 + \alpha_1 \hat{AT}_{ijt} + f(GPA_{it} - 7) + C_{jt} + x_i^{'}\alpha + v_{ijt} \tag{3}$$

One can obtain the reduced form (Equation (1)) by substituting Equation (2) into Equation (3). Note that the instrument set in the first stage can be extended with interaction terms, $\{D_{it}, D_{it} * f(\cdot)^n\}$, to match the number of endogenous variables, $\{AT_{it}, AT_{it} * f(\cdot)^n\}$, if we want the treatment to have a different impact across the distribution. The parameter α_1 measures a well-defined 'weighted' local average treatment effect (LATE).

As Lee and Lemieux (2010) make clear, the regression discontinuity design is not only informative for the subpopulation of individuals at the threshold. Rather it can be interpreted as a weighted average treatment effects across all individuals, where the weights

[12]Note that we use the reduced-form to select the order of the polynomials and the bandwidth in the local linear regression. This solution seems reasonable as the first stage seems relatively flat. Therefore the polynomial is expected to be flat and the optimal bandwidth to be wide.

are proportional to the ex ante likelihood that an individual's realization of the GPA will be close to the threshold.[13] In our case, the interpretation of the (weighted) LATE is the subpopulation that went to the tutorials because they on average scored lower than a 7 in their first bachelor year (the compliers) but would not have gone otherwise. However, LATE can only be identified in case the instrument is exogenous, if we have a valid first stage, if it satisfies the exclusion-restriction and the monotonicity-assumption is met (Imbens and Angrist, 1994 and Hahn, Todd and van der Klaauw, 2001). The assumption of exogeneity is met, due to its character of random assignment. The first stage, and the exclusion of the weak-instrument problem, is shown in Section 7. The exclusion-restriction would be met if there is no other mechanism the mandatory-schooling policy has affected the students' grades than through the attendance. This is fundamentally untestable in the presence of unobservables. For arguing that the monotonicity-assumption is met, we need a small explanation. Following Imbens (2010), imagine a model where the grade of student i solely depends upon the presence at the tutorial (suppressing other subscripts):

$$G_i = \beta_0 + \beta_1 AT_i + \epsilon_i \tag{4}$$

Attendance is endogenous ($cov[AT_i, \epsilon_i] \neq 0$), due to, for example, motivation both correlating with attendance and the grade. Now one can think of attendance not as a binary random variable, but as a continuous latent variable (AT_i^*) which describes your utility of attending the tutorials. Subsequently, this latent variable can be modeled through:

$$AT_i^* = \gamma_0 + \gamma_1 D_i + v_i \tag{5}$$

Where D_i reflects the assignment due to the mandatory-tutorials policy. Note, this continuous variable is mapped into a binary variable by the following:

$$AT_i = \begin{cases} 1 & \text{if } AT_i^* \geq 0, \\ 0 & \text{if } AT_i^* < 0. \end{cases} \tag{6}$$

The inclusion of D_i in the equation above can be argued to reflect the benefit of being present at a tutorial. That is, if $D_i = 1$, the utility of being present at the tutorial is

[13] In the most extreme case, all weight are similar across indivuals and the RD gap woud be equal to the overall average treatment effects. In our analysis one might reason the signal to noise of the test is relatively high and the weights might differ across individuals. Ultimately, as we observe one realization of GPA per person, the similarity of the weights is untestable.

higher, since otherwise you are ruled out from the exam. Hence, a rational agent maximizing his utility, would set γ_1 bigger than 0. Other characteristics are captured by v_i, for example, motivation. Students very unmotivated will never go to the tutorials, even if D_i is equal to 1: $v_i < -\gamma_0 - \gamma_1$ (never takers). Very motivated students still will always go to the tutorials, even if D_i is equal to 0: $v_i \geq -\gamma_0$ (always takers). What results are the compliers, who are defined by: $-\gamma_0 - \gamma_1 \leq v_i < -\gamma_0$. This framework rules out the condition of defiers, since (i) if an individual is not present if $D_i = 1$, this implies he will also not be present if $D_i = 0$ (if $v_i < -\gamma_0 - \gamma_1$, then $v_i < -\gamma_0$) and (ii) if an individual is present if $D_i = 0$, this implies he will also be present if $D_i = 1$ (if $v_i \geq -\gamma_0$, then $v_i \geq -\gamma_0 - \gamma_1$). This means we are able to identify a well-defined local average treatment effect.

For all specifications mentioned above we cluster standard errors at the individual level, to allow for a nonzero covariance of the error terms within each individual.[14]

We are able to exploit the regression-discontinuity design in a local difference-in-difference as to increase the precision of our estimates. Looking at the two-by-two treatment-control set-up of Table 1, we create the variable CB_{it} which equals 1 if a student completed the first Bachelor year within the first year (treatment group) and 0 if a student did not complete the first Bachelor year within the first year (control group). Subsequently, we estimate the following regression on our full sample (including students who did not complete their Bachelor one):

$$G_{ijt} = \delta_0 + \delta_1 CB_{it} + \delta_2 D_{it} + \delta_3 (CB_{it} D_{it}) + x_i^{'} \delta + \xi_{ijt} \tag{7}$$

Note that δ_3 recuperates the difference-in-difference estimate we are after. This coefficient compares the changes of outcomes around the threshold for students who completed their Bachelor with the changes of outcomes around the threshold for students for students who did not completed their Bachelor one. Unto the extent that the changes in grades near the threshold would be due to unobservables (and not due to the treatment), the comparison with the control group controls for this, as the effect of these unobservables would also be present in the control group.[15]

[14]Note that students face a different composition of peers per course, as such it seems not appropriate to cluster upon the tutorial-group level.

[15]Note that we also limit out control group to a sampe of students that potentially could have completed their Bachelor one within the first year, as to make the groups as comparable as possible (see Section 2).

5 Testing the Identifying Assumption

The table below compares individuals on background characteristics around the cut-of and shows no significant differences, which makes the identifying assumption plausible. The regression specifications in the table are a simple version of Equation (1), where we do not include any controls or a function of the running variable ($f(\cdot)$). Therefore, the treatment-coefficient indicates the simple difference in means between the left- and right side of the threshold, where it is estimated on a sample with a bandwith of 0.2 (so the running variable varies from 6.8 to 7.2).

Table 2: Table 2

	(1) Gender	(2) Distance	(3) Date of birth	(4) Nationality	(5) Grade secondary school
Treatment	0.0208	34.57	-73.01	-0.132	-0.217
	(0.44)	(0.62)	(-1.13)	(-1.36)	(-1.64)
Constant	0.281***	45.25***	12080.4***	0.623***	6.953***
	(8.80)	(4.24)	(277.67)	(9.26)	(79.69)
Observations	381	381	381	104	303
Adjusted R^2	-0.002	-0.002	0.001	0.008	0.006

t statistics in parentheses, standard errors are robust

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

We do the same test for students who did not complete Bachelor one within the first year, where we limit the sample to the group of 'potential' students. Again, we find no significant differences, which makes the identifying assumption more credible.

WORK IN PROGRESS; MCCRARY TEST. Visually, one can see there is no true discontinuity of the density of the running variable around the threshold, which shows we graphically past the test of McCrary (2008).

6 Mandatory Attendance Policy and Grades

This section shows the results for the intention-to-treat effect of the mandatory tutorials policy. Before showing the results of Equation (1), we visually show a negative, but small, jump in grades around the cut-off in Figure 2. This figure is consistent with a negative point estimate for the effect of mandatory tutorials.

Table 3: Table 3

	(1) Gender	(2) Distance	(3) Date of birth	(4) Nationality	(5) Grade secondary school
Treatment	-0.279*	-3.909	-166.9	-0.500	0.0460
	(-1.75)	(-0.55)	(-0.55)	(-1.41)	(0.25)
Constant	0.429***	16.38**	11577.1***	0.750**	6.865***
	(3.14)	(2.46)	(43.53)	(3.00)	(39.32)
Observations	34	34	34	8	28
Adjusted R^2	0.068	-0.019	-0.020	0.125	-0.035

t statistics in parentheses, standard errors are robust

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

Figure 2: Visual McCrary-test, density of running variable

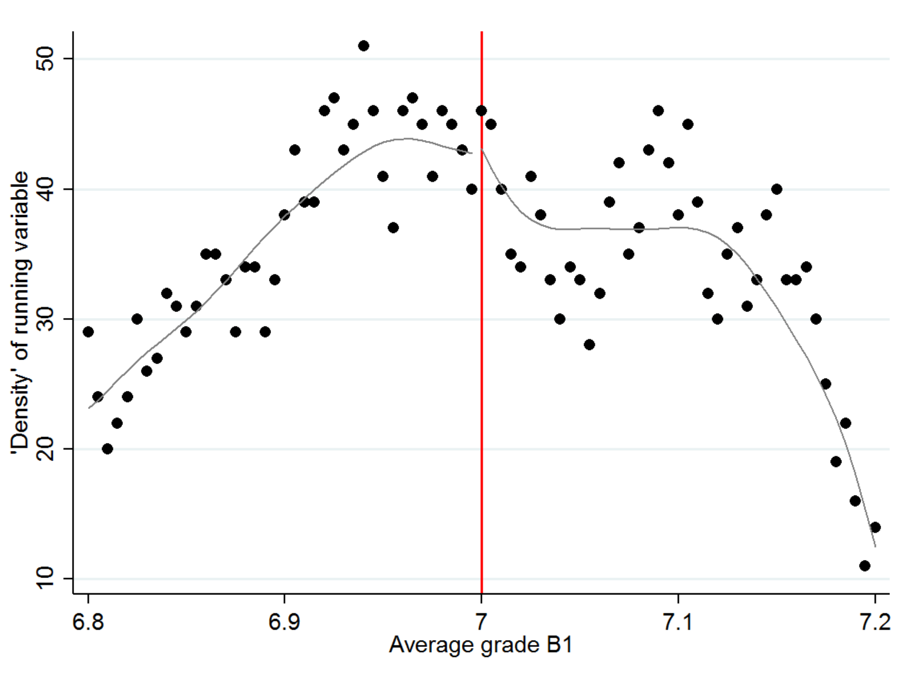
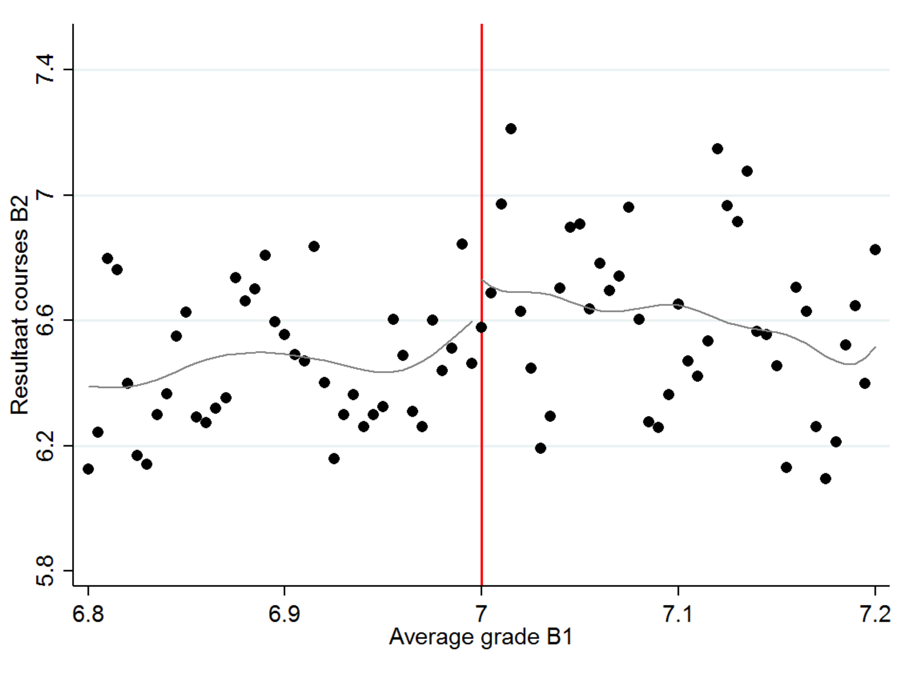


Figure 3: Visual representation of the intention-to-treat



Before resorting to estimation of Equation (1) and Equation (7), we use a nonparametric approach to estimate the intention-to-treat effect. Remember, the identifying assumption is that the bias cancels in the limit. As such, we can just take the means of the second year grades on each side of the GPA-threshold to get a first idea of the direction and size of the effect of mandatory-tutorials. Table 4 shows the means of second-year grades of students who are located near the threshold with a bandwith of 0.1 (the GPA running variable varies from 6.9 to 7.1). The regression-discontinuity estimate related to Equation (1) can be interpreted as the differences in means between students on the left and right side of the threshold. A non-parametric estimate of the intention-to-treat effect is therefore $\lim_{GPA_{it}\uparrow 7} E[G_{ijt}|GPA_{it}] - \lim_{GPA_{it}\downarrow 7} E[G_{ijt}|GPA_{it}] = 6.408 - 6.626 = -0.234$. Alternatively, one can use the local difference-in-difference as described by Equation (7). Taking the course without tutorials as an extra control group, we find the non-parametric estimate to be: $(\lim_{GPA_{it}\uparrow 7} E[G_{ijt}|GPA_{it}, CB_{it} = 1] - \lim_{GPA_{it}\downarrow 7} E[G_{ijt}|GPA_{it}, CB_{it} = 1]) - (\lim_{GPA_{it}\uparrow 7} E[G_{ijt}|GPA_{it}, CB_{it} = 0] - \lim_{GPA_{it}\downarrow 7} E[G_{ijt}|GPA_{it}, CB_{it} = 0]) = (6.411 - 6.624) - (7.397 - 7.285) = -0.324$. The two procedures give very similar results, as we cannot conclude they are significantly different from each other. This is a confirmation of the identifying assumption of the regression-discontinuity design.

Subsequently, the table below shows the estimates corresponding to Equation (1). Sim-

Table 4: Non-parametric approach for the intention-to-treat

Courses with tutorials \ GPA	6.9 $\leq$GPA$<$ 7	7 $\leq$GPA $<$ 7.1
YES	6.411	6.624
NO	7.397	7.285

Notes: This table shows the means of the grades of the second-year Bachelor courses, where a distinction is made between courses with and without tutorials.

ilar to Figure 2, we opt for a bandwith of 0.2 for the running variable and for a second order polynomial for $f(\cdot)$, which is interacted with the treatment variable in column (3) and (4). Column (2) and (4) include additional controls, namely gender, nationality, date of birth, distance towards the university and grade on secondary school. Note that because of the control 'grade on secondary school' we lose some observations, as we do not have this control variable for international students. The treatment-estimate consistently shows a negative sign, sometimes significant so. The coefficient in column (1) can be interpreted as follows: going from untreated to treated decreases your grade on a course in the second year of the Bachelor by 0.25 points. Grades are measured on a scale from 0 to 10, where 5.5 is a sufficient grade. The mean of the second year grades is 6.24 and the standard deviation is 1.62.

Table 5: Table 4

	(1) Grades	(2) Grades	(3) Grades	(4) Grades
Treatment	-0.243*	-0.294**	-0.295	-0.328
	(-1.91)	(-2.09)	(-1.54)	(-1.65)
Constant	6.670***	5.151***	6.643***	5.168***
	(79.45)	(18.21)	(54.54)	(16.69)
Observations	3456	2665	3456	2665
Adjusted R^2	0.006	0.024	0.006	0.023
Polynomial	Quadratic	Quadratic	Quadratic	Quadratic
Interacted polynomial	NO	NO	YES	YES
Control variables	NO	YES	NO	YES

t statistics in parentheses, standard errors are clustered at the individual level

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

7 Attendance and Grades

7.1 LATE

This section shows the IV-estimates of Equation (2) and Equation (3), where we use the treatment as a local instrumental variable for attendance. In the following subsection we show that this procedure is able to recover a well-defined local average treatment effect.

The table below shows the actual estimates of the first- and second stage. The specifications are similar to the models estimated in Section 6, as we opt for a bandwith of 0.2 for the running variable and choose a quadratic function for $f(\cdot)$. The estimates of the first stage show the jump in attendance is roughly 15%-point. F-values show the treatment-variable is strongly related with actual attendance. The second-stage estimates show a negative point estimate of actual attendance on grades, but insignificantly so.

Table 6: Table 5

	(1) Attendance First stage	(2) Attendance First stage	(3) Grades Second stage	(4) Grades Second stage
Treatment	0.111***	0.172***		
	(3.26)	(3.33)		
Attendance			-2.112	-0.565
			(-1.25)	(-0.46)
Constant	0.842***	0.784***	7.113***	6.573***
	(13.32)	(10.90)	(10.59)	(12.51)
Observations	1558	1558	1558	1558
Adjusted R^2	0.138	0.144	0.000	-0.009
F-value	10.67	11.09		
Polynomial	Quadratic	Quadratic	Quadratic	Quadratic
Interacted polynomial	NO	YES	NO	YES
Control variables	YES	YES	YES	YES

t statistics in parentheses, standard errors are clustered at the individual level

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

Subsequently, we naturally split the sample of second-year Bachelor courses into General Economics, Business Economics and Econometrics (Cluster A, B and C respectively, see Table A.1 for the classification of the respective courses). The table below first shows the intention-to-treat estimates for the three clusters of courses separately. We use the same specification as above, choosing for a quadratic polynomial and a bandwith of 0.2 for the running variable. We again see negative point estimates throughout, which is strongly significant for the cluster Business Economics on a 2% level.

Table 7: Table 6

	(1) Grades	(2) Grades	(3) Grades
Treatment	-0.148	-0.368**	-0.246
	(-0.94)	(-2.35)	(-1.42)
Constant	6.338***	6.838***	6.917***
	(62.02)	(69.63)	(62.73)
Observations	1270	1234	952
Adjusted R^2	0.002	0.010	0.006
Polynomial	Quadratic	Quadratic	Quadratic
Interacted polynomial	NO	NO	NO
Control variables	NO	NO	NO
Cluster	Gen. Eco	Bus. Eco	Metrics

t statistics in parentheses, standard errors are clustered at the individual level
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Consistently with a negative effect of tutorial-attendance upon grades, the table below shows that for the cluster Business Economics the jump in attendance is largest, namely 26%-points. While the jump in attendance for the Econometrics-cluster is indistinguishable from zero. As such, it seems to be that this policy was most binding for the non-technical courses. Moreover, computing the IV-estimate shows a negative and significant effect of the tutorials for the Business-economics cluster, whereas the IV-estimates are insignificant (but negative) for the other two clusters of courses. This estimate should be interpreted as follows: an increase of attendance by 10%-point decreases the grade on a second-year Bachelor course by 0.16 point.

7.2 Characterizing Compliers

To see why the IV-estimate of the second stage (α_1) does not equal ATE, we switch to a heterogeneous framework. This means that the parameters potentially differ per

Table 8: Table 7

	(1) Attendance First stage	(2) Grades Second stage	(3) Attendance First stage	(4) Grades Second stage	(5) Attendance First stage	(6) Grades Second stage
Treatment	0.159***		0.256***		0.0186	
	(3.63)		(5.54)		(1.16)	
Attendance		-0.318		-1.606**		-11.69
		(-0.26)		(-2.21)		(-0.72)
Constant	0.695***	6.607***	0.623***	7.900***	0.952***	18.08
	(19.99)	(6.97)	(15.99)	(14.16)	(78.63)	(1.16)
Observations	751	751	738	738	558	558
Adjusted R^2	0.090	0.000	0.217	0.000	0.002	0.000
F-value	13.177		30.692		1.346	
Polynomial	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic
Interacted polynomial	NO	NO	NO	NO	NO	NO
Control variables	NO	NO	NO	NO	NO	NO
Cluster	Gen. Eco	Gen. Eco	Bus. Eco	Bus. Eco	Metrics	Metrics

t statistics in parentheses, standard errors are clustered at the individual level
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

individual, so formally we have α_{1i} in the first stage regression. If attendance would be exogenous in the first place, we would still be able to measure an ATE(T), since $ATE(T) = E[G_i|AT_i = 1] - E[G_i|AT_i = 0] = E[\alpha_{1i}] = \frac{1}{n}\sum_{i=1}^{n}\alpha_{1i} = \bar{\alpha_1}$. Since AT_i is exogenous, this average can still be interpreted as an ATE(T). Now consider AT_i as an endogenous variable and one uses 2SLS in order to get a consistent estimate. In a heterogeneous framework, the 2SLS-estimator equals:

$$\alpha_{1,2sls} = \frac{cov[G_i, D_i]}{cov[AT_i, D_i]} = \frac{\frac{1}{n}\sum_{i=1}^{n}\alpha_{1i}\gamma_{1i}}{\frac{1}{n}\sum_{i=1}^{n}\gamma_{1i}} \tag{8}$$

This boils down to ATE(T) if and only if $\alpha_{1i} = \alpha_1 \forall i$ and/or $\gamma_{1i} = \gamma_1 \forall i$. Hence, in a heterogeneous framework the 2SLS-estimator equals a weighted average of individuals' treatment effects, with larger weight for whom the instrumental variable is most influential. Under the assumptions mentioned above this weighted average measures a LATE. This exercise makes clear that homogeneity in the first stage means LATE equals ATE. Thus to characterize the LATE, we do something unusual. We redo the first stage and include interaction effects between D_i and observables to find for which individuals γ_{1i} is large or small. Equation (8) makes clear that individuals with a large γ_{1i} contribute to the LATE-estimator and individuals with a small γ_{1i} do not contribute to the LATE-estimator.[16] Whereas the monotonicity assumption is also fundamentally untestable, we would not want the total effect of D_i to become negative. Indeed, this causes the explanation below Equation (6) to break down, since γ_{1i} is not positive for all individuals.

WORK IN PROGRESS; RESULTS OF INTERACTION; EXPLAIN WHICH INDIVIDUALS CONTRIBUTE MOST TO LATE (ON OBSERVABLES)

8 Conclusion

WORK IN PROGRESS

References

[16]In the extreme case, let $\gamma_{1i} = 0$, where the contribution of student i equals zero.

Appendix

Table A.1: Overview of courses in Bachelor 1 and 2

Cluster	Bachelor 1 Courses	ECTS	Block	Bachelor 2 Courses	ECTS	Block
A	Microeconomics	8	2	Applied Microeconomics	8	3
	Macroeconomics	8	3	International Economics	8	1
	Organisation and Strategy	8	5	History of Economic Thought	4	3
B	Financial Information Systems	4	1	Intermediate Accounting	8	5
	Marketing	8	4	Behavioral Economics	4	4
	Financial Accounting	8	5	Finance 1	8	2
C	Mathematics I	4	1	Methods & Techniques	8	4
	Mathematics II	4	3	Research Project	4	5
	Applied Statistics I	4	4	Applied Statistics II	4	2
	ICT	4	2	Economics of Ageing	4	1

Notes: This is for the academic year 2013-2014, but the courses and the grouping within the clusters of courses where almost identical from academic year 2008-2009 onwards.