

Ostermaier, Andreas; Beltz, Philipp; Link, Susanne

## Conference Paper

# Do university policies matter? Effects of Course Policies on Performance

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2013: Wettbewerbspolitik und Regulierung in einer globalen Wirtschaftsordnung - Session: University Enrollment and Student Achievement, No. A07-V2

### Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

*Suggested Citation:* Ostermaier, Andreas; Beltz, Philipp; Link, Susanne (2013) : Do university policies matter? Effects of Course Policies on Performance, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2013: Wettbewerbspolitik und Regulierung in einer globalen Wirtschaftsordnung - Session: University Enrollment and Student Achievement, No. A07-V2, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel und Hamburg

This Version is available at:

<https://hdl.handle.net/10419/79924>

### Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

### Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

**DO UNIVERSITY POLICIES MATTER?  
EFFECTS OF COURSE POLICIES ON PERFORMANCE**

Philipp Beltz, Susanne Link, and Andreas Ostermaier\*

**Abstract**

We benefit from the Bologna reform to show how course and program policies affect academic achievement. We examine two similar programs at the business school of a major European university, which were both reformed. Time lags in the reforms allow us to estimate the difference in the differences of student performance in a compulsory second-year course. Performance fell as the impact of the course on the graduation certificate decreased and the time until students received the certificate increased. More students failed as they were allowed to resit the exam more often. Both effects depend on ability. We conclude that program policies matter and universities should be aware of their effects. (JEL: I21, I23, I28)

## 1 Introduction

Improving the academic achievement of students is a major concern of researchers and policy makers alike. Research has identified several drivers of academic performance, both individual and social, such as ability, family background, or school resources ([Hanushek, 2003](#)). Performance also depends on effort, which,

---

\*Acknowledgments: This is a revised and extended version of [Beltz et al. \(2012\)](#). We thank Joshua Angrist, Uschi Backes-Gellner, Eberhard Feess, Eric Hanushek, Hans-Ulrich Küpper, Ludger Wößmann, seminar participants at the Ifo Institute for Economic Research, participants of the 74<sup>th</sup> Annual Meeting of the German Academic Association for Business Research, the 26<sup>th</sup> Annual Conference of the European Society for Population Economics, and the 13<sup>th</sup> Symposium of the German Economic Association of Business Administration for helpful comments. Susanne Link gratefully acknowledges financial support from the German Research Foundation through GRK 801.

unlike ability or the family background, is both variable and susceptible to incentives (Levitt et al., 2012). This makes effort a primary starting point for improving academic achievement.

The Bologna Process offers us the opportunity to show that policies for programs or courses affect performance and may serve as incentives.<sup>1</sup> Grading is an example of such policies, which indeed include all sorts of prerequisites, requirements, credit points, the number of resits students can take, and the like. We use unique data from reforms of program policies at a major European university to show that performance decreased when the impact of a course on the (graduation) certificate decreased and the time until this certificate was awarded to students increased. We also find that students who were allowed to resit exams more often took this opportunity. However, students responded differently to the new policies depending on their ability.

The course and program policies of some university are probably not the first thing one would think of when talking about a process on the European level which transformed entire systems of higher education. However, even the effects of national and international reforms have their causes in individual behavior. While we examine effects on a rather disaggregate level, our study may still help understand the effects of the Bologna Process. Nonetheless, the Bologna Process is not the focus of this study; instead, it is an occasion for us to investigate effects which are steadily at work in universities but can more readily be observed under the shock of the reforms that the Bologna Process launched.

To the best of our knowledge, we are first to study program and course policies as incentives. By contrast, there is substantial research on *monetary* incentives. Among secondary school students, monetary rewards were found to

---

<sup>1</sup>“Program” refers to a set of courses—some compulsory, some elective—that students have to take in order to earn a degree, like in an MBA program. By contrast, a “course” is a lecture, tutorial, seminar, or a combination of those, for which one overall grade is reported.

increase test scores and other outcomes, such as completion rates (Angrist et al., 2002, 2006; Kremer et al., 2009; Angrist and Lavy, 2009). Among university students, there is evidence that merit-based scholarship programs and rewards raise enrollment rates (Cornwell et al., 2006) and that continuation fees prompt students to graduate within the scheduled time (Garibaldi et al., 2012). However, some studies fail to find effects (e.g., Fryer, 2011), while others suggest that design issues such as timing are crucial (Levitt et al., 2012). Leuven et al. (2010) show that the effects of monetary rewards depend on ability, which we find for course and program policies as well.

Our contribution is twofold. On the one hand, we offer first insights into an unexplored field. We thus contribute to the burgeoning behavioralist literature on education (Levitt et al., 2012). On the other hand, the effects of course and program policies are highly relevant for universities and policy makers. Universities must design policies when they set up courses or programs and should be aware of their effects. If used properly, they may be effective and inexpensive means to enhance performance. Moreover, although monetary rewards (e.g., reductions or refunds of tuition fees or scholarship programs) are widespread, there are countries where education is free and these rewards are less relevant. Input-based measures, such as reducing class sizes, which affect individual performance indirectly, are as well costly and their (marginal) effectiveness is controversial (e.g., Hoxby, 2000; Hanushek, 2003).<sup>2</sup> It is therefore necessary and worthwhile to examine course and program policies.

We consider the business school of the University of Munich, which offers two similar programs, both of which were subject to reforms in the Bologna Process.

While the policies for the first program were changed as early as 2005, and again

---

<sup>2</sup>Recent research shows that academic achievement also depends on the *composition* of the class (De Giorgi et al., 2012). However, from a practical viewpoint, the composition of classes may be even harder to manipulate than their sizes, and certainly much harder than course and program policies.

in 2008, the reform of the second program was delayed until 2010. We benefit from this delay and use a difference-in-differences approach to estimate the effects of the reform on performance in a second-year course, which is compulsory for both programs. All students attend the same course, are taught by the same instructors, use the same textbooks and teaching materials, and have a nearly identical curriculum when they take the exam, which corroborates the common trend assumption. Technically, we interpret the revisions of the policies as natural experiments, where the students in the reformed program are the treatment group, those in the other program, the control group.

The first reform in 2005 introduced a Bachelor program and thus doubled the time until students received their first certificate and halved the impact of the course considered on this certificate. The second reform in 2008 revised the Bachelor program and allowed students to resit the exam more often. After the reforms, performance decreased and the portion of students who failed increased. However, students responded differently to the reforms depending on ability. These results are robust to different specifications and tests. They hold when we restrict our sample to groups of students that are less prone to selection and match students along the vector of observable characteristics. We show that students in the second program, conditional on age and time at university, responded similarly to those in the first program when the same reforms applied to them in 2010.

The remainder of this article is organized as follows: We first describe the institutional setting and the reforms that we consider (Section 2). We then predict how students would respond to the reforms (Section 3). We go on to describe the data (Section 4) and present our identification strategy and the empirical model (Section 5). We then report the baseline results for both reforms

along with the subgroup analyses (Sections 6) as well as the results of the tests for robustness (Section 7). Finally, we conclude with a discussion of our findings and their implications (Section 8).

## 2 Institutional Setting

In 1999, the education ministers of 29 European countries agreed in Bologna to create the European Higher Education Area, which required them to harmonize their national university systems. The so-called Bologna Process obliged German universities, which up to that point did not distinguish between Bachelor and Master degrees, to introduce sweeping reforms across their degree programs. These reforms affected grading, credit points, number of resits, and similar program and course policies. While the Bologna Process produced a host of effects, it is those program and course policies we are interested in.

The business school of the University of Munich offers undergraduate programs in business administration and business education.<sup>3</sup> Students of both programs are trained to take management positions, but the business education program prepares them also to become teachers at vocational schools. The curricula diverge therefore toward the end, while both programs are nearly identical at the beginning. Until the Bologna reform, both programs were divided into two stages or periods of study. Students received the *Vordiplom* certificate at the end of the first period, and the *Diplom* certificate at the end of the second period, when they graduated. Graduates were awarded a degree called *Diplom*. Table 1 gives an overview of the first period of study, which is relevant for our investigation.

Insert Table 1 about here.

---

<sup>3</sup>The traditional German names are “Betriebswirtschaftslehre” for business administration and “Wirtschaftspädagogik” for business education. We use “business education” rather than the formal translation “Human Resource Education & Management” for convenience.

Program policies specify which courses students had to take and which exams to pass in order to earn their certificates and thus their degree, as depicted in Table 1. Both the *Vordiplom* and *Diplom* certificates reported an overall grade, which averaged the grades students had obtained in the courses required for that certificate and which we refer to as grade point average (GPA).<sup>4</sup> While there was no *Vordiplom* degree, the *Vordiplom* certificate was a prerequisite to qualify for the second period of study. The *Vordiplom* drew a line under the first period of study as grades earned until then did not count toward the *Diplom* certificate.

The business administration program was adapted to the Bologna system in two steps, which involved a major reform in 2005 and a minor reform in 2008. In 2005, the four-year *Diplom* program was replaced with a three-year Bachelor program.<sup>5</sup> Since then, the Bachelor certificate is the first and only certificate that students receive upon completing their program. (They can still retrieve transcripts of records at any time, but this was also possible before the Bachelor program was introduced.) While the first three semesters of the Bachelor program were nearly identical to the first three semesters of the *Diplom* program, their completion was no longer marked by anything equivalent to the *Vordiplom*. In contrast to the *Diplom* program, now all grades count toward the final Bachelor GPA.

In 2008, the newly introduced Bachelor program was revised and, as a result, the number of times that students were allowed to resit an exam was raised for most courses. According to the 2005 policies, students in the Bachelor program were allowed a maximum of three attempts for exams, like those in the *Diplom* program. After the revision, students may now take exams as often as they want

---

<sup>4</sup>More precisely, German universities award usually grades ranging from 1.0 (excellent) to 4.0 (sufficient) as well as 5.0 (failed). The overall grade in the *Vordiplom* and *Diplom* certificates was a weighted mean of the grades obtained in a number of courses. As German grades are numbers, they are not translated into grade points. We use the common term GPA for convenience.

<sup>5</sup>In addition, a Master program and degree were introduced. Most of the courses to be taken in the last year of the *Diplom* program became part of the new Master program.

as long as they graduate within three or at most four years, which adds up to a maximum of eight attempts, as they can take the exam each semester. Students who fail at their final attempt must not continue on this or a related program at *any* university in Germany. Students thus realized early whether they had chosen the right program.<sup>6</sup> In 2010 the business education program was reformed in one single step, which combined both reforms of the business administration program. The 2010 reform restored the situation before 2005, when the organization of both programs was identical.

For our analysis, we interpret both reforms as natural experiments. Accordingly, the reforms are the treatments. In either experiment, the students of business administration form the treatment group, whereas those of business education are the control group. Response to the treatments is observed in terms of performance in the compulsory course on “production & organization,” which students were advised to take in their third semesters according to 1. Both the timeline of the reforms and the design of the experiments is depicted in Figure 1.

Insert Figure 1 about here.

## 3 Theoretical Predictions

### 3.1 Student Responses to Incentives

The reforms can be more easily assessed if we think of performance or academic achievement as a function of family, peer, and school inputs as well as student characteristics, such as ability (Hanushek, 1986, 2002). Research based on such models has often focused on school resources, which translated into providing institutions with more resources and especially into reducing class sizes (Hoxby,

---

<sup>6</sup>The limited number of resits helped screen out students who were not suitable for the program of their choice. German universities charge little or no tuition fees so that students might overuse their resources and drop out late otherwise.



2000). The success of this input-based approach is contestable (e.g., [Hanushek, 1996, 2003](#); [De Giorgi et al., 2012](#)). Interest has therefore been growing in incentives both for institutions and individuals involved in education, such as accountability policies or competition (e.g., [Besley and Ghatak, 2006](#); [Bishop, 1997](#); [Bishop and Wößmann, 2004](#); [Jürges et al., 2005](#); [Wößmann, 2003](#)). Monetary incentives for students are an example of these incentives.

The purpose of incentives for students is to increase their effort and, as a result, performance. There is evidence that effort drives academic achievement. For example, the OECD's Programme for International Student Assessment (PISA) revealed that truancy and inattention, which indicate insufficient effort, correlate with poor reading and mathematical skills ([Bishop, 2004](#)); conversely, attendance was found to enhance performance ([Bjorvatn and Tungodden, 2010](#); [Durden and Ellis, 1995](#)). Likewise, in some studies the effects of monetary rewards could be attributed to increased effort ([Angrist et al., 2002, 2009](#)). Although effort does not necessarily translate into academic achievement ([Fryer, 2011](#)), the link is on the whole intuitive. Accordingly, effort has been included in the inputs of education production functions (e.g., [Akerlof and Kranton, 2002](#); [Bishop, 2004](#); [Bishop and Wößmann, 2004](#)).

Unlike other drivers of academic performance, such as ability or family background, effort is variable, because students can choose how much effort they make. Effort choices depend both on the individual and the situation, which makes them susceptible to incentives. Learning can be experienced as rewarding in itself, but it is also a means of obtaining other rewards, such as recognition or the prospect of higher earnings. The motivation to learn may be intrinsic or extrinsic. In either case, students choose the level of effort that maximizes their expected net benefit, that is, the difference between the benefits and costs of

studying (Akerlof and Kranton, 2002), where the benefits consist of both intrinsic and extrinsic rewards, while the costs may be monetary or non-monetary (e.g., tuition fees, but also time, strain, stress, etc.).

Students will respond to changes of the policies which alter their cost–benefit ratio by adjusting their effort; in fact, unless they are exclusively driven by intrinsic motivation, most modifications of the policies will shift their cost–benefit ratio. This approach helps explain how students responded to the two reforms of program and course policies.

### **3.2 The 2005 Reform: Introduction of the Bachelor Program**

The major reform of the program in business administration in 2005 was the replacement of the *Diplom* with the Bachelor degree and the abolition of the *Vordiplom*. Ever since then, the Bachelor certificate is the first and only certificate that students receive. This is critical because certificates are important rewards of studying. They serve as a “signal” when students apply for jobs, scholarships, or programs at other universities (Hanushek, 2002).

The effects of abolishing the *Vordiplom* were particularly hard to predict because it was not a degree, but rather a milestone required to enter the second period of study. On the one hand, one might argue that, being no discernible reward, it did not motivate students anyway. Moreover, the grades earned until the *Vordiplom* did not count toward the *Diplom*; according to this argument, the reform might even have increased motivation for those early exams, because grades would now count for the degree certificate. On the other hand, students used their *Vordiplom* certificate much like a degree certificate. For example, they presented it when they applied for jobs or at other universities.

As a matter of fact, the *Vordiplom* was therefore quite similar to a degree, and its abolition affected students’ cost–benefit ratio for the course that we consider

in at least two ways. First, it doubled the time until their effort was rewarded, because the Bachelor program takes three or at most even four years, whereas the *Vordiplom* was regularly awarded after 1.5 and never later than after two years. Second, the reform halved the impact of the grade on the GPA because the courses that count for the Bachelor GPA are more than twice as many as those that counted for the *Vordiplom* GPA.

Although studying may be intrinsically beneficial, benefits such as a better job or higher lifetime earnings are obtained in the future, while students have to make an effort and pay for the costs now. Students will therefore account for time and discount the expected benefits when comparing them to the costs. In this vein, it has been argued that monetary rewards for students might correct for overly high discount rates (Angrist and Lavy, 2009), and there is also experimental evidence that the timing of rewards matters (Levitt et al., 2012). The 2005 reform augmented the time lag between the cost and benefit of studying and thus reduced the expected benefit. Consequently, we expect that students reduced their effort and thus performance fell.

The second effect of the reform bears on the impact of the course on the GPA, which is, for the *Vordiplom* like for the Bachelor certificate, the weighted average of the grades obtained in courses. While the number of credit points reflects the hours that a course requires, it also tells students how much the grade affects their GPA. Students will probably prioritize courses which carry more credit. However, they will not only consider the credit of one course relative to others. Even a course with much credit has little impact if the number of courses factored into the GPA is large, and the weight of each course is small. As a result, students will put less effort into each course as the number of courses increases.

This response seems irrational at first glance, because it impairs the grade

obtained in the single courses and, by implication, the GPA. However, it is intuitive from a *motivational* perspective. The reward that students receive—i.e., a high-class certificate or whatever this means for their future job or earnings—depends on their performance in a number of exams, which in turn hinges on their effort (and other factors). If the impact of a course is small, performance has less impact on the outcome, and effort is less worthwhile. We therefore predict that, because of this argument and the time lag, the 2005 reform decreased the effort and thus performance of students.

### **3.3 The 2008 Reform: Revision of the Bachelor Program**

The minor reform of the Bachelor program in 2008 increased the number of times that students could resit the exam, shifting again their cost–benefit ratio. Before the reform, they were allowed no more than three attempts, whereas they may now resit the exam as often as they want provided that they graduate in three or at most four years. As a result, the number of possible attempts was effectively more than doubled.

When students do not have the chance to take an exam twice, the cost of failing is prohibitive or, put differently, the benefit of succeeding is immense. The cost was particularly high for the students that we consider because in Germany students who failed had to leave their program and could not even go to a different university to continue on it. As the number of attempts increases, the cost of studying remains the same, but the benefit of succeeding decreases for all attempts but the last.

This may have two effects, both of which compromise performance. On the one hand, students may generally make less effort and prepare themselves worse for their exam. On the other hand, as they must not resit the exam unless they have failed, they may decide to submit blank exam sheets and fail deliberately,

so that they can take the exam again and improve. Students who decided to fail either had not to answer at all or to cancel their answers. Like an overall decrease in effort and performance, this implies that they scored lower, failed more often, and took more attempts to pass the exam. However, unlike the overall decrease in performance, which is probably not intentional, the submission of a blank exam is obviously a choice. It is therefore important to distinguish both effects.

To summarize, we expect that the 2008 reform caused, in particular, the rate of blank submission to increase and, in general, performance to decrease. We expect that the decrease in performance resulted in lower test scores and a higher rate of failure. While both effects were driven by the increase in blank submission, we expect that they exceeded this effect after accounting for blank submission.

## 4 Data

We examine the effects of these reforms on performance in a course on “production & organization” which is compulsory for students of business administration and business education, no matter in which program. Students were advised to take the course in their third semester at university (see Table 1). Contents are general and the one-hour written exam tests both quantitative and verbal skills. The course is the same for all students and it is offered every year. Since two teams of instructors alternate, we consider only the courses and exams held by the same team.

We collected data on the students who took the exam in 2006, 2008, 2010, and 2012. The data were compiled from two sources. The office of the university registrar collects personal data from students when they apply for admission, such as age, gender, country of birth, or prior qualifications. The office of the registrar of the business school keeps academic records and files exams. However,

academic records only report grades, whereas we wanted to rely on the more finely partitioned test scores. We therefore retrieved the exams from the file room and digitized the data on performance for each observation. To combine information on performance and personal characteristics, we finally matched the data from these sources.

The students of business administration who took the exam in 2006 were enrolled in the *Diplom* program, those who took it in 2008, in the Bachelor program, and those who took it in 2010, on the revised Bachelor program. The students of business education, whether they took the exam in 2006, 2008, or 2010, were enrolled in the *Diplom* program. All students who took the exam in 2012 were in Bachelor programs, and the newly introduced Bachelor program in business education corresponded to the revised Bachelor program in business administration.

Few observations were discarded for different reasons. Two students in the *Diplom* program took the exam only in 2008 rather than 2006, sixteen in the Bachelor program, in 2010 rather than 2008. These students did not respond to the reforms because they were in the “wrong” program and were therefore not considered. Exchange students were excluded for the same reason. Moreover, we discarded nine observations because information on the program was missing. The remaining database totals nearly 1,600 observations

The outcome variables are test scores, the rates of failure, and the rate of blank submission. In addition to test scores and the programs students were enrolled in, our database contains information on demographic characteristics, namely gender, age, and country of birth. It also contains information on the number of semesters a student had officially been enrolled in his or her program when he or she took the exam (“semester”) and on his or her secondary education certificate

(“high school GPA”). We miss information on age for 1% of the students and on high school GPAs for 10%.

Test scores range from 0 to 120 points. High school GPAs were rescaled so that higher numbers mean better grades, and range from 1.0 to 4.0.<sup>7</sup> Exams scoring ten points or less were defined as “blank” to indicate that the examinee failed deliberately. Students who failed deliberately either did not answer at all or cancel their answers before submitting. Nonetheless, students who scored only slightly above zero were presumably sure enough that they would not pass or they just forgot to cancel some of their answers. Summary statistics are reported in Table 2.

Insert Table 2 about here.

Table 2 shows that in each year between 60 and 80 percent of the students were enrolled in the business administration program, the rest in the business education program. The mean test score varies considerably between the exams, ranging from slightly above 60 in 2010 to around 80 in 2012. The rate of failure was about 25 percent in 2006, but fell to about 15 percent in the following years. The rate of blank submission rose strikingly from zero in 2006 to 6% in 2010 among students of business administration and remained above zero in 2012 for both programs.

About half of the students of business administration and between 60 and 70% of those of business education were female. The proportion of students born abroad ranged from 4 to 11% in business administration, from 0 to 3% in business education. The high school GPAs were about 3.0 (“good”), with major increases from 2008 to 2010 among business administration students, and from 2010 to

---

<sup>7</sup>Similar to grades at universities (see Footnote 4), high school GPAs range from 1.0 to 4.0 and 6.0, where grades worse than 4.0 mean that the student failed. Since students who failed at high school must not enter university, high school GPAs at university range from 1.0 to 4.0 and it seemed convenient to just reverse the scale.

2012 among business education students. Most students took the exam in their third semester at university, as they were advised (see Table 1), which is at about the age of twenty-three in business administration, and twenty-four in business education. Students of business education were on average older because many of them received vocational training before entering university.

Students who enrolled in business education in 2012 were advised to take the exam immediately in their first semester (see Table 1 again). The students of business education were therefore on average younger and took the exam in an earlier semester in 2012. However, this change in the composition of the examinees from 2010 to 2012 affects only our tests for robustness, where we still control for age and semesters.

## 5 Identification Strategy and Empirical Model

Both the 2005 and 2008 reforms of the program in business administration can be understood as natural experiments, where the reforms are treatments; the students of business administration, the treatment group; the students of business education, the control group. Performance in the exams on “production & organization” in 2008 measures how students responded to the 2005 reform, in 2010, how they responded to the 2008 reform (see Figure 1). The exam in 2012 is considered for robustness checks explained below.

We compare the difference in the performance of the treatment group before and after each reform to the difference in the performance of the control group. The difference-in-differences approach allows us to measure the effect of each reform, because the reforms applied only to one program and, at the same time, were the only major change between each couple of exams. We also compare performance in 2010 and 2006 to measure the combined effect of both reforms



(i.e., the revised Bachelor program) and pool the data from 2006, 2008, and 2010 to measure the single effects of both reforms at once. The elements for the test of our predictions are illustrated in Figure 1, those for the two tests for consistency in Figure 2.

Insert Figure 2 about here.

To estimate the effects of the reforms, we run variants of the following model:

$$y_i = \alpha + \beta_1 D_i + \beta_2 T_i + \delta T_i D_i \gamma' X_i + \varepsilon_i, \quad (1)$$

where  $y_i$  is student  $i$ 's performance (test score, rate of failure, or rate of blank submission). The dummy variables  $D_i$  and  $T_i$  indicate student  $i$ 's program (1, if business administration) and year of examination (1, if after the reform or reforms).  $X_i$  denotes a set of control variables, and  $\varepsilon_i$  an idiosyncratic error term. The coefficient  $\delta$  captures the effect of the reform;  $\beta_1$ , the time-invariant effect specific to the treatment group;  $\beta_2$ , the time effect for the control group;  $\gamma$ , the effects of the control variables, namely gender, migration, high school GPA, age, and semester.

We report estimates with and without the set of control variables. Our baseline specification compares only treatment and control groups, but does not account for any covariates. In order to correct for remaining differences between the treatment and control group, we add control variables to the regression. As a side effect, this reduces the residual variation and improves the precision of our estimates. We first account for characteristics of the student, which are determined before enrollment, namely gender, migrant status, and high school GPA, before we include age and semester. Age and semester might be endogenous and are therefore examined closely when testing for robustness.

The interpretation of  $\delta$  as a causal treatment effect rests on two identifying assumptions. First, the trends in the performance of both groups would have been parallel without treatment. Second, the treatment and control groups did not change as a result of the treatment. Irrespective of tests for robustness, the setting of the experiments gives us reason to believe that both assumptions were met.

In support of the first assumption, the course and exam were the same for students of both programs. Students attended the same classes, were taught by the same instructor, and used the same materials and textbooks. The course is general and requires both quantitative and verbal skill so that it is unlikely to favor any group of students. Neither the instructor nor the contents changed over time. All students wrote the same exam in the same room and at the same date. Table 1 shows that, except for the reforms, curricula were nearly identical for both programs until the exam, which translates into similar workloads and priorities. Even the reforms did not affect the programs apart from the specific interventions described. More generally, all students lived in the same city, shared similar social environments, and had similar career opportunities.

The main concern about the second assumption is that students might have preferred the well-known *Diplom* to the unfamiliar Bachelor program. However, admission to the *Diplom* program was closed when the Bachelor program was introduced. Even a different university was hardly an option because similar universities introduced Bachelor programs about the same time (e.g., the Humboldt University of Berlin and the Goethe University of Frankfurt introduced the Bachelor in the winters of 2004 and 2005). It is also unlikely that students preferred the old *Diplom* program in business education to the new Bachelor program in business administration. While our research benefits from the similarity of both

programs, prospective students did not generally know about it and would not easily switch to the other program.

The summary statistics give an idea of the similarity and comparability of both groups of students. In particular, selection into one program or the other is probably reflected in observable statistics. Table 2 reveals differences in observable characteristics between the programs in 2006 already. However, these differences do not pose a threat to identification as long as they remain constant over time. Between 2006 and 2008, there is hardly any difference in the differences except in age. Between 2008 and 2010, there is a remarkable increase in the spread between high school GPAs. The changes between 2010 and 2012 do not affect our proper tests, but our tests for robustness, and will be discussed below.

High school GPA is considered the most informative indicator of the overall ability to study of German students (Trapmann et al., 2007). The enrollment of probably abler students in business administration is surprising, because it is contrary to then common concerns that students might avoid Bachelor and run into *Diplom* programs. Moreover, the 2008 reform was minor and unlikely to be perceived by prospective students. It would have attracted worse rather than better students, if any. The increase in high school GPAs can therefore hardly be taken to indicate self-selection. Nonetheless, it requires attention because it might result in biased estimates.

In order to alleviate this threat, we control for high school GPA. Furthermore, we show that our results hold in groups of students that are less prone to selection. In particular, we restrict our sample to students who took the exam in their third semester and students in the ninetieth percentile of the age distribution. Furthermore, we match students along the full vector of observable characteristics to make them more comparable. As an attempt to test the common trend

assumption, we show that, conditional on age and semester, students of business education responded like their peers when the same reforms applied to them (see Figure 2).

## 6 Results

### 6.1 Effects of the 2005 Reform

#### Main Effects

For the 2005 reform, we consider the 2006 and 2008 exams, that is, before and after the reform came into effect and the Bachelor replaced the *Diplom* program in business administration. More precisely, we compare the differences between the test scores and rates of failure of students in both programs to test whether the reform caused performance to decrease. The OLS regression estimates are reported in Table 3. Estimates for the effect of the reform on the rate of blank submission are not reported; Table 2 shows that the rate was zero for both years and programs, and there is no significant effect, as expected.

Insert Table 3 about here.

From the left-hand side of Table 3 can be seen that the 2005 reform had a significant negative effect on test scores. Specifically, the negative coefficient reported for all students (Column 1) remains about the same in size and level of significance if we restrict the regression to students without missing values (Column 2) or introduce control variables. We first control for variables determined before enrollment, namely gender, migrant status, and high school GPA, which does not affect the coefficient of interest (Column 3). As we go on to control for age and the number of semesters spent at university, the coefficient decreases, because the difference in age between students of both programs increases over

time (Column 4). (The overall result does not change if age and semester are included separately.)

The right-hand side of Table 3 reports the effect of the 2005 reform on the rate of failure. As the reform causes test scores to decrease, it has a positive effect on the rate of failure, which is barely significant for all observations, including those with missing values (Column 5). If the regression is restricted to full observations (Column 6) and predetermined student characteristics are accounted for (Column 7), the standard error increases and the effect is no longer significant. Moreover, when age and the number of semesters are controlled for, the coefficient drops by about 0.04 (Column 8), because older students are more likely to fail and, on average, the students of business education who took the exam in 2006 were older than those in 2008.

In summary, Table 3 shows that the 2005 reform had a significant negative effect on test scores and a positive effect on the rate to fail of failure, which is not statistically significant, though.

### **Heterogeneous Effects**

Prior research suggests that the effect of incentives might be moderated by ability (Leuven et al., 2010). To test for heterogeneous effects, we divide the students who took the exam in 2006 and 2008 in terciles according to their high school GPA. We then estimate the effects of the 2005 reform on test scores and the rate of failure for poor, intermediate, and excellent students separately. Table 4 displays the results.<sup>8</sup>

Insert Table 4 about here.

---

<sup>8</sup>The results are very similar if the students are categorized in other ways, so as to obtain a larger group of intermediate students. In particular, we ran the same tests for the first (low ability), the second and third (intermediate ability), and fourth quartile (high ability).

The results suggest that the 2005 reform affected in particular intermediate students. It caused both intermediate and excellent students to score lower (Columns 3 and 5). The decrease in test scores translates into a significant increase in the rate of failure of intermediate students (Column 4), whereas the effect on the rate of failure is close to zero and not significant for excellent students (Column 6). Among poor students, we observe the same pattern, but the coefficients are smaller. Neither the decrease in test scores nor the increase in the rate of failure is statistically significant (Columns 1 and 2).

The differences between the groups are intuitive. Poor students are or feel unable to perform better or worse as they make more or less effort. Since they cannot control their performance and thus grade in the first place, it does not matter to them how much impact their grade has on their GPA. Hence, they hardly respond to the reform. By contrast, good students know that they can influence their performance. Therefore they make less effort as the reform reduces the impact of their performance. While this is true even for excellent students, the decrease in test scores is too small to make them fail.

These results are consistent with the idea that ability and effort interact and confirm the findings of [Leuven et al. \(2010\)](#). However, unlike [Leuven et al.](#), we do not only distinguish between student with low and high ability, but single out students with intermediate ability. This third category is particularly relevant because the bulk of students are intermediate. Both public policies and university policies (e.g., course and program policies) will probably focus on this group.

## 6.2 Effects of the 2008 Reform

### Main Effects

The 2008 reform allowed students of business administration to resit the exam more often if they failed. While students who took the exam in 2008 had three attempts, those in 2010 had effectively eight. As a result, we expect students to submit blank sheets. We use again a difference-in-differences model to assess the effect of the 2008 reform on the rate of blank submission. The estimates are listed in Table 5.

Insert Table 5 about here.

The results confirm that the rate of blank submission increased. The coefficient and the standard error hardly vary, no matter whether all observations are considered (Column 1), those with missing values are excluded (Column 2) or student characteristics are accounted for (Columns 3 and 4). The effect of the reform on the rate of blank submission is highly significant.

We also examine the effects of the reform on test scores and the rate of failure. We argued above that the reform would have two effects, which should be separated. We predicted that it caused performance to decrease in general, i.e. in terms of test scores and the rate of failure, and that it led the rate of blank submission to increase, as shown in Table 5. To test whether there is an effect on performance beyond that on blank submission (which necessarily comes along with a decrease in test scores), we control for blank submission. The estimates of the effects are reported in Table 6. The effects are statistically not significant, no matter the specification of the model.

Insert Table 6 about here.

In summary, the 2008 reform had the expected effect overall. It caused the rate of blank submission to rise. However, it did not produce an effect on performance over and above the increase in blank submission for the whole sample.

### **Heterogenous Effects**

In 2008, students responded differently to the 2005 reform depending on ability. Similarly, we expect heterogeneous effects of the 2008 reform. Again, we divide students into poor, intermediate, and excellent students according to their high school GPA and estimate the effects of the reform on performance for each tercile separately. Table 7 lists the estimates.

Insert Table 7 about here.

The results show that ability actually moderated the responses. The effect of the reform on the rates of blank submission is significant for intermediate and excellent students and larger for the former than for the latter (Columns 4 and 7). We observe even a significant negative effect on test scores among intermediate students after controlling for blank submission (Column 5). By contrast, none of the effects of the reform is significant for poor students.

The 2008 reform gave ill-prepared students who hoped to do better at a later attempt the chance to submit a blank sheet. The outcomes are in line with the argument that good students use this chance, possibly hoping to “improve” when they would be resitting the exam. However, they also prepared themselves worse as the cost of failing fell, which is reflected in the decrease of test scores of intermediate students. By contrast, poor students were not susceptible to the reform as they must submit always do their best and hope to pass.

In summary, we find again that intermediate students responded most strongly, excellent students less, and poor students least. This pattern of heterogeneous



effects is similar to that of the 2005 reform.

### **Long-term Effects**

The decrease in performance suggests that the 2008 reform caused the average number of attempts to pass the exam to grow. As the number of attempts is still limited by the maximum time to graduation of four years, it may even lead students to drop out only on the verge of graduation and thus waste both social and individual resources. It is therefore important to consider the long-term effects of this reform.

An increase in the number of attempts might be beneficial if resits improve students' level of training and increase their likelihood of success. Thus, it would at least help those students who finally pass and would have dropped out otherwise. However, students may also lose motivation and perform even worse as they retake the exam. Moreover, the number of exams to be taken at a time increases as students procrastinate, which makes success even more unlikely. It is therefore dubious whether students become more likely to succeed at later attempts or whether they finally drop out.

We cannot fully explore this problem because we were unable to track the students in our database and gather data on graduation and final grades. However, Figure 3 shows that students at a higher attempt scored generally lower and were more likely to fail. Obviously, there is adverse selection because students must not resit the exam once they have passed and consequently only those students who performed worst remain for the next attempt. The effects are therefore not only driven by the additional workload and strain due to procrastination and differences in motivation and effort, but also by differences in ability. Nevertheless, this figure questions whether the higher number of resits does help students succeed.

Insert Figure 3 about here.

## 7 Robustness Tests

### 7.1 Tests for Consistency

The analyses of both reforms produced mainly two results. The 2005 reform resulted in a decrease in test scores, while the 2008 reform caused the rate of blank submission to increase. To check our results for robustness, we first test whether they hold if both reforms are considered together.

We compared the exams in 2006 and 2008 to isolate the effect of the 2005 reform. Likewise, we compared those in 2008 and 2010 to assess the effect of the 2008 reform. As a first test for consistency, we compare the exams in 2006 and 2010. We thus estimate the combined effect of both reforms or, put differently, the effect of the revised Bachelor program. In our difference-in-differences approach, this effect is modeled as the interaction of the year (2010 compared to 2006) and the program. The results are reported in Table 8.

Insert Table 8 about here.

The combined effects of both reforms are consistent with the single effects. First, both reforms together cause test scores to drop significantly. The decrease is not driven by blank submission, which is controlled for (Columns 1–3). The coefficients are slightly larger than those for the 2005 reform alone (see Table 3). Second, the reforms cause the rate of blank submission to increase significantly (Columns 7–9). This effect is about the same size as that of the 2008 reform alone (see Table 5).

As another test for consistency, we pool the data for 2006, 2008, and 2010 and model interaction terms to separate the effects of both reforms in one regression.

More precisely, we let the program interact both with the year 2008 and 2010 to capture the effects of the 2005 reform and the combined effect of both reforms. The pattern of the effects reported in Table 9 is similar to that of the first consistency test.

Insert Table 9 about here.

The effects of both the interaction term labeled “2005 reform” and that labeled “both reforms” are negative and consistent (Columns 1–3). The negative effect of the second term is consistent with the combined effect of the reforms exceeding that of the 2005 reform alone (Table 8). The combined effect of both reforms on the rate of blank submission is significant and positive, whereas the 2005 reform does not have a significant effect (Columns 7–9). Again, this is consistent both with the estimates of the isolated effects and the first test for consistency. The increase in blank submission is due to the 2008 reform only (see Tables 5 and 8).

Unlike the 2005 reform, the 2008 reform is not modeled explicitly for Table 9. It is reflected, though, in the differences between the coefficients on the 2005 reform and both reforms, which nearly equal the coefficients reported in Tables 5 and 6. For example, the 2008 reform did not affect test scores and accordingly, the differences in Columns 1–3 are minimal. Conversely, the 2005 reform did not affect blank submission, and accordingly the coefficients in Columns 7–9 are close to zero, while those for both reforms equal those in Table 5.

In summary, the results of both tests are consistent with those reported above for the single reforms. We can both combine and separate the effects in pooled data.

## 7.2 Specification Tests

The summary statistics in Table 2 reveal that students of both programs differed in age, the number of semesters spent at university until the exam, and high school GPA, and that the differences changed over time. While these characteristics have significant effects when they are included as control variables, they do not seem to drive the effects of the treatments or reforms that we measure (see in Tables 3 and 5). Nonetheless, it is crucial that the treatment and control groups are comparable to produce reliable estimates.

We therefore run our regressions on restricted subsamples that are more comparable and match the treated to the most similar non-treated individuals in terms of observable characteristics. As a first restriction, we exclude students beyond the ninetieth percentile of the age distribution or aged more than 26 years, because students of business education were on average older than their peers. Second, we only consider students who took the exam in their third semester, as recommended by the business school, excluding those who procrastinated or rushed for some reason.

For matching, we estimate propensity scores based on gender, migrant status, and high school GPA, which are predetermined before students enroll at university. The results are depicted in Figure 4 for each exam. Then, we first match each treated student with their nearest neighbor in terms of propensity scores within a caliper of 0.04.<sup>9</sup>

Insert Figure 4 about here.

As we allow replacement, regression outcomes are weighted by the frequencies non-treated observations are used as matches. Second, we perform kernel

---

<sup>9</sup>Smith and Todd (2005) note that it is difficult to know a reasonable caliper width a priori. We follow Cochran and Rubin (1973) and set the caliper width to 0.2 standard deviations of the estimated propensity score.

matching, which considers more matches but gives closer matches more weight. Both times, treated individuals with a propensity score less than the minimum or more than the maximum of the controls are dropped to ensure common support.

Insert Table 10 about here.

Table 10 reports the effect of the 2005 reform on test scores and the effect of the 2008 reform on the rate of blank submission for observations restricted by semester and age as well as for with nearest neighbor and kernel matching. As can be seen from the left-hand side of the table, there is still a significant negative effect of the 2005 reform on test scores. While the effect among third-semester students in Column 1 is about one standard deviation larger, the coefficients in Columns 2 to 4 are very similar to those in Table 3.

Likewise, the right-hand side of the table shows a significant positive effect on the rate of blank submission. The size and significance of the coefficients correspond to those in Table 5. Although we report the results while controlling for student characteristics, these effects of both reforms are insensitive to the controls. As in Table 3, the effect of the 2005 reform on the rate of failure is positive but not significant.

Overall, the results presented in this table confirm our previous findings. The estimates are similar to those for the whole sample, and the effects are the same. As far as we can control for differences between the groups, as well as differences in the differences, these are unlikely to drive the effects that we find.

### 7.3 Placebo Test

Our comparison of students of different programs relies on the assumption of a common trend. We think that this assumption is reasonable because the policies were very similar for both programs except for the reforms, and both groups of

students attended the same course, were taught by the same lecturers, and took the same exam. However, since the office of the university registrar did not keep records before 2004, we cannot test for a common trend before the reforms or treatments. Instead, we benefit from a reform of the business education program in 2010 to conduct another test of consistency.

In 2010, the business education program was reformed in a single step, reiterating the 2005 and 2008 reforms of business administration. The 2010 reform thus corresponds to the combination of the 2005 and 2008 reforms considered in Subsection 7.1. It reestablished the original situation where the policies for both programs were the same. As the 2010 reform took effect before 2012, we consider the exams in 2006 and 2012. If both groups followed the same trend as they were subject to the same reforms or treatments, we should find that the interaction term of the year (2012 vs. 2006) and the program, which we label “placebo,” did not affect performance significantly.

This test comes with a caveat. The last generation of students who enrolled on business education before the exam in 2012 was advised to take the exam in their first rather than third semester. As a result, students of business education were on average younger and took the exam in a lower semester in 2012. Table 2 also reveals that the difference in age between students of both programs almost disappeared in 2012, while there was now, unlike in 2006, a difference in semesters. This shift would easily explain a break in the common trend. However, we should not observe an effect of the placebo interaction term at least if we control for age and semesters. The results of the test are reported in Table 11.

Insert Table 11 about here.

For blank submissions, the Placebo coefficient is close to zero and statistically not significant in any specification (Columns 7–9), which is consistent with both

groups of students following the same trend before and after the increase in the number of resits. For test scores, we observe a negative significant effect in Columns 1 and 2, which translates into a positive, though not significant, effect on the rate of failure (Columns 4 and 5). As age and semesters are included, the effect decreases in magnitude and significance. That is, conditional on age and semesters, we find no significant difference in students' responses to the reforms (Columns 3 and 6).

The test offers some support for the common trend assumption, as the placebo effect is not significant if age and semesters are controlled for. Still, the test is generally a weak substitute for a test on earlier data (which we do not have) and particularly because the shifts in age and semesters prevent us from clearly rejecting the placebo effect. One might argue that the decrease in test scores is only temporary and limited to the first generation of Bachelor students, who were confused and uncertain about the Bologna reform and the new Bachelor program. It should be noted, though, that most students who took the exam in 2008 were in the third semester, and thus were already familiar with the program and its policies. Moreover, the negative effect persists after the 2008 reform (Tables 8 and 9), which itself did not affect test scores (Table 6), and is robust to several specifications. It seems to be more than a restructuring effect.

In summary, the results the consistency and specification tests. It is impossible to rule out that they are due to anything but the reforms. However, they seem quite robust overall.

## 8 Discussion and Conclusions

How can the academic achievements of students be improved? This is a concern shared by students, who strive for excellence in their education; universities,

which aim to satisfy this demand; and society in general, as human capital is a driver of economic growth and wealth (Hanushek and Wößmann, 2011). While improvements remain desirable, the conventional input-based policies applied to this end are costly and, because of diminishing returns, may be maxed out. They are no longer an effective, let alone efficient means of improving student performance (Hanushek, 2003). Class size reduction is a case in point (Hoxby, 2000). Research interest has therefore turned to incentives for universities, faculties, and students. Incentives for students directly affect effort, which is an essential input in the production of education (Bishop and Wößmann, 2004).

Our analysis of two natural experiments has shown how students respond to policies for university programs and courses. More specifically, we found that the performance of students decreases as the time until they receive a certificate that rewards the effort they have put into an exam increases; as the number of courses increases and hence the perceived leverage that each course adds to their effort decreases; as they are given more chances to fail, so to speak, which leads them to feel that they are required to put less effort into each attempt. In particular, these results imply that certificates are perceived as rewards or milestones, although it should be noted that the interim certificate—the “*Vordiplom*”—in this case was very similar to a certificate that students receive when they graduate rather than to a mere transcript of records.

The empirical finding that program and course *policies* can help create incentives for students is the main contribution of this work. This is both intuitive and in keeping with results from related research. Universities and schools have long used incentives, the most evident example of which is grades, and recent research has shown that students also respond to monetary rewards. However, it did not necessarily follow from this research that program and course



policies function as incentives. Like grades and money, time, leverage, and (formal) rewards are factors that can lead students to increase or decrease their effort. In particular, students seem to respond to these incentives as one would expect rational actors to do with the benefit of hindsight. (However, their response was not so easy to predict because of the ambiguous role of the *Vordiplom*.) It is noteworthy that responses differ depending on ability, which is neglected in most empirical studies.

These findings imply that policies are not just a necessary part of program implementation, but also offer universities a means of guiding the efforts that students put into their degrees and of fostering their academic performance. For instance, credit points reflect the time and effort students are expected to spend on a course, but also the importance that the university attributes to it and that the university expects students to give it; similarly, certificates document the academic achievements of students, but are also milestones in their studies. Policies are incentives available to every university; they are *effective* because students respond to them; they are *efficient* because they are generally inexpensive and promise especially great “returns.” Most importantly, policies function as incentives irrespective of universities’ intentions. Universities should therefore use them carefully and strategically.

Our study has several limitations, mainly due to the design of the natural experiments and which future research could address. In the first experiment, the performance of students was expected to decrease for *two* reasons: the delay in rewarding student effort, and the reduced leverage of exams on the GPA. Our experimental design did not allow us to separate these two effects. Hence, it cannot be seen from the data which effect accounts for how much of the decrease in performance and whether the two effects interact. For instance, one effect

may be much more important than the other. It is equally possible that the two effects reinforce or suppress each other. This problem is both intriguing from a scientific viewpoint and practically relevant to universities.

A peculiarity—and potential limitation—of this study is that it was conducted among students of business programs, who are trained to understand and design incentives. It is possible that they respond more strongly than those of other programs, which would lead us to overestimate the effects of the reforms. At the same time, students familiar with theories of motivation may be aware that, as the leverage of a given course decreases, the effort put into that course also decreases, and that if the leverage of *all* courses is reduced, this will clearly result in a lower GPA. If they anticipate this problem and react accordingly, it is equally possible that we may have underestimated the effect of the first reform. For that reason, similar experiments should be conducted among different student populations.

Finally, future research may extend on this study by exploring longitudinal data. The negative effects that we observed may partly be due to indirect or side-effects, such as confusion or even hostile attitudes toward the reforms among students. Moreover, as argued above, more resits may prevent students from realizing early that the program does not suit them, and thus cause them to drop out too late to embark on a different program. In support of this worry, we found that students who fail one or two attempts are even less likely to succeed at their second or third attempt. However, it is not clear whether an individual student still has a better chance to graduate. More resits might have training effects and might benefit students who would drop out otherwise. Anyway, this is an important issue both for universities and students and deserves further attention.

In conclusion, program and course policies as incentives in higher education

are an exciting field of research both from a theoretical and a practical perspective. The Bologna reform offered a unique chance to explore this topic in a specific university setting. The valuable data that this reform has produced may help to shed light on many aspects of higher education.

## References

- Akerlof, G. A. and Kranton, R. E. (2002). Identity and schooling: Some lessons for the economics of education. *Journal of Economic Literature*, 40(4):1167–1201.
- Angrist, J., Bettinger, E., Bloom, E., King, E., and Kremer, M. (2002). Vouchers for private schooling in Colombia: Evidence from a randomized natural experiment. *American Economic Review*, 92(5):1535–1558.
- Angrist, J., Bettinger, E., and Kremer, M. (2006). Long-term educational consequences of secondary school vouchers: Evidence from administrative records in Colombia. *American Economic Review*, 96(3):847–862.
- Angrist, J., Lang, D., and Oreopoulos, P. (2009). Incentives and services for college achievement: Evidence from a randomized trial. *American Economic Journal: Applied Economics*, 1(1):136–163.
- Angrist, J. and Lavy, V. (2009). The effects of high stakes high school achievement awards: Evidence from a randomized trial. *American Economic Review*, 99(4):1384–1414.
- Beltz, P., Link, S., and Ostermaier, A. (2012). Incentives for students: Evidence from two natural experiments. *Ifo Working Paper*, 133.
- Besley, T. and Ghatak, M. (2006). Sorting with motivated agents: Implications for school competition and teacher incentives. *Journal of the European Economic Association*, 4(2/3):401–414.

- Bishop, J. H. (1997). The effect of national standards and curriculum-based exams on achievement. *American Economic Review*, 87(2):260–264.
- Bishop, J. H. (2004). Drinking from the fountain of knowledge: Student incentive to study and learn—Externalities, information problems and peer pressure. *CAHRS Working Paper*, 19.
- Bishop, J. H. and Wößmann, L. (2004). Institutional effects in a simple model of educational production. *Education Economics*, 12(1):17–38.
- Bjorvatn, K. and Tungodden, B. (2010). Teaching business in Tanzania: Evaluation participation and performance. *Journal of the European Economic Association*, 8(2/3):561–570.
- Cochran, W. G. and Rubin, D. B. (1973). Controlling bias in observational studies: A review. *Sankhyā: The Indian Journal of Statistics, Series A*, 35(4):417–446.
- Cornwell, C., Mustard, D. B., and Sridhar, D. J. (2006). The enrollment effects of merit-based financial aid: Evidence from Georgia’s HOPE program. *Journal of Labor Economics*, 24(4):761–786.
- De Giorgi, G., Pellizzari, M., and Woolston, W. G. (2012). Class size and class heterogeneity. *Journal of the European Economic Association*, 10(4):795 – 830.
- Durden, G. C. and Ellis, L. V. (1995). The effects of attendance on student learning in principles of economics. *American Economic Review*, 85(2):343–346.
- Fryer, R. G. (2011). Financial incentives and student achievement: Evidence from randomized trials. *Quarterly Journal of Economics*, 126(4):1755–1798.
- Garibaldi, P., Giavazzi, F., Ichino, A., and Rettore, E. (2012). College cost and time to complete a degree: Evidence from tuition discontinuities. *The Review of Economics and Statistics*, 94(3):699–711.

- Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature*, 24(3):1141–1177.
- Hanushek, E. A. (1996). Rationalizing school spending: Efficiency, externalities, and equity, and their connection to rising costs. In Fuchs, V. E., editor, *Individual and Social Responsibility: Child Care, Education, Medical Care, and Long-Term Care in America*, pages 59–106. University of Chicago Press, Chicago, IL.
- Hanushek, E. A. (2002). Publicly provided education. In Auerbach, A. J. and Feldstein, M., editors, *Handbook of Public Economics*, volume 4, chapter 30, pages 2045–2141. North-Holland, Amsterdam.
- Hanushek, E. A. (2003). The failure of input-based schooling policies. *Economic Journal*, 113(485):F64–F98.
- Hanushek, E. A. and Wößmann, L. (2011). How much do educational outcomes matter in OECD countries? *Economic Policy*, 26(67):427–491.
- Hoxby, C. M. (2000). The effects of class size on student achievement: New evidence from population variation. *Quarterly Journal of Economics*, 115(4):1239–1285.
- Jürges, H., Schneider, K., and Büchel, F. (2005). The effect of central exit examinations on student achievement: Quasi-experimental evidence from TIMSS Germany. *Journal of the European Economic Association*, 3(5):1134–1155.
- Kremer, M., Miguel, E., and Thornton, R. (2009). Incentives to learn. *Review of Economics and Statistics*, 91(3):437–456.
- Leuven, E., Oosterbeek, H., and van der Klaauw, B. (2010). The effect of financial

rewards on students' achievement: Evidence from a randomized experiment.

*Journal of the European Economic Association*, 8(6):1243–1265.

Levitt, S. D., List, J. A., Neckermann, S., and Sadoff, S. (2012). The behavioralist goes to school: Leveraging behavioral economics to improve educational performance. *NBER Working Paper*, w18165.

Smith, J. A. and Todd, P. E. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics*, 125(1–2):305–353.

Trapmann, S., Hell, B., Weigand, S., and Schuler, H. (2007). Die Validität von Schulnoten zur Vorhersage des Studienerfolgs – eine Metaanalyse. *Zeitschrift für Pädagogische Psychologie*, 21(1):11–27.

Wößmann, L. (2003). Schooling resources, educational institutions and student performance: The international evidence. *Oxford Bulletin of Economics & Statistics*, 65(2):117–170.

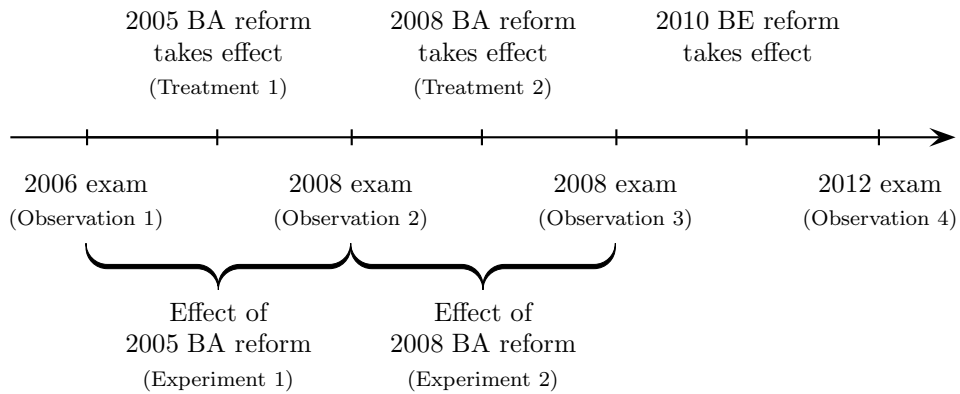


FIGURE 1. Timeline of the Reforms. In the Bologna Process, the business school of the University of Munich reformed the policies for its programs in business administration (BA) and business education (BE). Business administration was reformed first in two steps (2005, 2008), business education with a delay in one step (2010).

The 2005 reform replaced the traditional *Diplom* with a Bachelor program. The 2008 reform revised this newly introduced Bachelor program. The 2010 reform of the program in business education reestablished the original situation, where the same policies applied to both programs.

We interpret the 2005 and 2008 reforms as natural experiments, where the students of business administration are the treatment group, those of business education, the control group. Our experimental interpretations of the events appear in parentheses. The 2010 reform is considered for robustness tests.

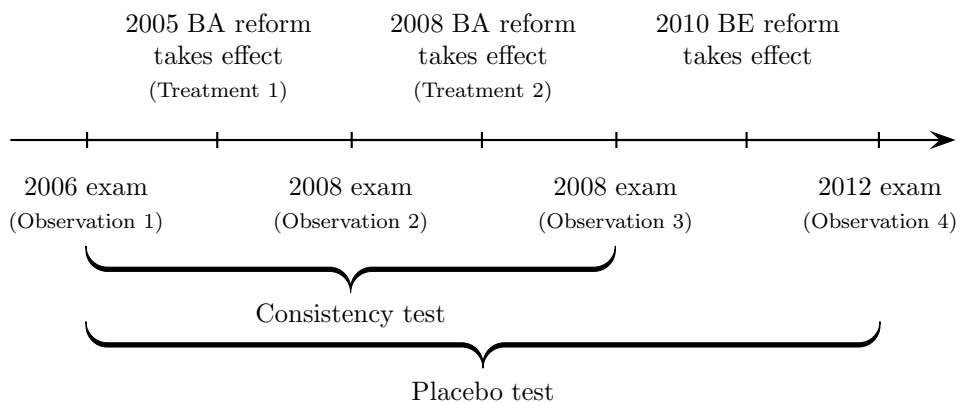


FIGURE 2. Robustness Checks. The students of business administration (BA) who took the exam in 2006 were enrolled in the *Diplom* program; in 2008, in the Bachelor program; in 2010, in the revised Bachelor program. Those of business education (BE) who took these exams were in the *Diplom* program. In 2012, all students were in Bachelor programs.

Compared to the students who took the exam in 2006, those in 2010 responded as if subject to a combination of the 2005 and 2008 reforms. When the data from 2006, 2008, and 2010 are pooled, the effects of the 2005 and 2008 reforms, which were found separately, can also be found in the pooled data. These results support consistency.

As a placebo test, 2012 is compared to 2006. The 2010 reform of business education combined the 2005 and 2008 reforms of business administration and restored the original situation, where both programs were nearly identical until the third semester. There is no longer a difference in the differences when accounting for age and semester.

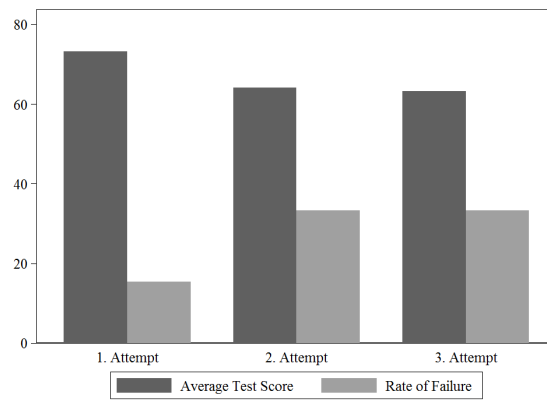


FIGURE 3. Test scores and rate of failure by number of attempts. Performance is lower for students who took the exam more than once (test scores are lower on average, the rate of failure higher). Students who submitted blank exams are not considered. They probably performed worse than they could, and their results in the exam fail to reveal their true potential.



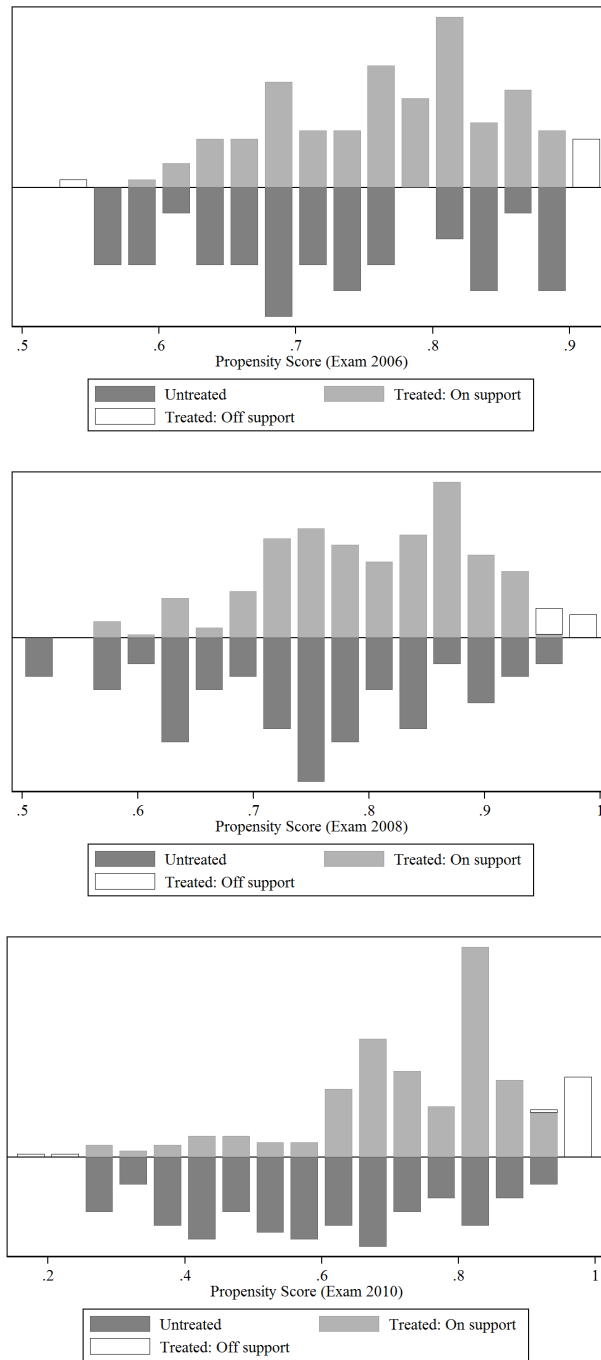


FIGURE 4. Estimated propensity scores by year of examination after kernel matching, imposing common support. Students are matched by gender, migrant status, and high school GPA to estimate the propensity score.

TABLE 1. Overview of Courses and Credit Hours.

Business administration			Business education	
<i>Diplom</i> (before 2005 reform)	Bachelor (after 2005 reform)	Revised Bachelor (after 2008 reform)	<i>Diplom</i> (before 2010 reform)	
<b>1<sup>st</sup> semester</b>				
<i>Principles of bus. admin.</i>	4	4	4	4
<i>Methods of accounting</i>	2	2	2	2
<i>Information systems</i>	4	8	8	4
<i>Mathematics</i>	2	2	2	2
<i>Statistics 1</i>	4	4	4	4
<i>Private law</i>	3	3	3	3
<b>2<sup>nd</sup> semester</b>				
<i>Marketing</i>	4	4	4	4
<i>Microeconomics</i>	8	8	8	8
<i>Statistics 2</i>	4	4	4	4
<i>Corporate law</i>	2	2	2	2
<i>Public law</i>	3	2	4	3
<b>3<sup>rd</sup> semester</b>				
<i>Accounting</i>	4	4	4	4
<i>Finance</i>	4	4	4	4
<i>Production &amp; Organization</i>	4	4	4	4
<i>Macroeconomics</i>	8	4	4	4
Empirical economics	4	4	4	4
		2	4	4
<b>4<sup>th</sup> semester</b>				
	64	65	65	68
			<i>Information systems</i>	4
			<i>Macroeconomics</i>	8
			Student teaching	—

Notes: Recommendation of the business school when to take courses. The courses common to all programs—business administration as well as business education and *Diplom* as well as Bachelor—are in italics. Until the exam in “production & organization,” all these programs share most courses and are very similar. Even the reforms did not cause the programs in business administration and business education to diverge considerably. No differences arose apart from those discussed in the text.

TABLE 2. Descriptive Statistics.

	2006			2008			2010			2012		
	BA (1)	BE (2)	Diff. (3)	BA (4)	BE (5)	Diff. (6)	BA (7)	BE (8)	Diff. (9)	BA (10)	BE (11)	Diff. (12)
Test score	72.41 (20.16)	62.59 (14.77)	9.82 (2.95)	79.77 (18.39)	79.70 (14.42)	0.07 (2.00)	58.53 (21.99)	60.36 (17.72)	-1.83 (2.06)	79.87 (23.07)	78.86 (22.18)	1.01 (2.50)
Blank subm.	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.06)	0.00 (0.00)	0.00 (0.00)	0.06 (0.23)	0.00 (0.00)	0.06 (0.01)	0.03 (0.17)	0.03 (0.17)	-0.00 (0.02)
Failure	0.23 (0.42)	0.33 (0.48)	-0.10 (0.08)	0.13 (0.33)	0.08 (0.28)	0.04 (0.04)	0.19 (0.40)	0.17 (0.38)	0.02 (0.04)	0.14 (0.35)	0.14 (0.35)	-0.01 (0.04)
Male	0.53 (0.50)	0.33 (0.48)	0.19 (0.09)	0.54 (0.50)	0.36 (0.48)	0.18 (0.06)	0.47 (0.50)	0.30 (0.46)	0.17 (0.05)	0.53 (0.50)	0.32 (0.47)	0.21 (0.05)
Foreign	0.04 (0.19)	0.00 (0.00)	0.04 (0.02)	0.09 (0.28)	0.03 (0.16)	0.06 (0.03)	0.11 (0.32)	0.02 (0.16)	0.09 (0.02)	0.08 (0.27)	0.02 (0.14)	0.06 (0.02)
HS GPA	2.85 (0.55)	2.67 (0.67)	0.17 (0.12)	2.97 (0.45)	2.78 (0.53)	0.18 (0.07)	3.11 (0.50)	2.71 (0.51)	0.40 (0.05)	3.02 (0.44)	2.92 (0.45)	0.10 (0.05)
Age	22.74 (1.43)	25.51 (3.72)	-2.78 (0.60)	22.84 (2.82)	24.36 (2.38)	-1.52 (0.32)	22.69 (2.09)	24.39 (2.60)	-1.70 (0.27)	22.96 (2.45)	22.64 (2.56)	0.32 (0.28)
Semester	2.95 (0.63)	3.26 (0.99)	-0.31 (0.17)	3.01 (0.18)	3.14 (0.38)	-0.13 (0.05)	2.98 (0.24)	3.11 (0.64)	-0.13 (0.06)	3.22 (0.88)	2.20 (0.98)	1.02 (0.11)
Observations	183 [129]	54 [39]	237 [168]	326 [288]	80 [73]	406 [361]	313 [285]	123 [122]	436 [407]	452 [428]	99 [98]	551 [526]

Notes: Summary statistics for observable characteristics of the four cohorts of students who took the exam on a compulsory third-semester course in 2006, 2008, 2010, and 2012. The table reports means and standard deviations (in parentheses) for students of business administration (BA) and business education (BE) as well as the differences between them (Diff.). “HS GPA” abbreviates high school GPA. The numbers of full observations appear in brackets below those of all observations (including observations with missing values).

TABLE 3. Effects of the 2005 Reform.

	Test Score				Rate of Failure			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
2005 reform	-8.947*** (3.118)	-9.751*** (3.557)	-9.578*** (3.132)	-8.436*** (3.191)	0.144* (0.083)	0.144 (0.093)	0.140 (0.087)	0.105 (0.086)
2008	16.642*** (2.548)	17.109*** (2.885)	15.745*** (2.542)	14.723*** (2.619)	-0.270*** (0.074)	-0.251*** (0.082)	-0.232*** (0.077)	-0.205*** (0.077)
Bus. admin.	7.526*** (2.441)	9.825*** (2.940)	7.491*** (2.675)	4.955* (2.852)	-0.093 (0.074)	-0.101 (0.084)	-0.063 (0.079)	0.006 (0.081)
Male			0.700 (1.379)	0.926 (1.364)			-0.038 (0.030)	-0.045 (0.029)
Foreign			-16.141*** (3.703)	-15.598*** (3.736)			0.247*** (0.083)	0.233*** (0.084)
HS GPA			16.186*** (1.583)	15.056*** (1.546)			-0.228*** (0.032)	-0.193*** (0.031)
Age				-1.138*** (0.355)				0.022*** (0.007)
Semester				1.587 (1.741)				0.040 (0.042)
Constant	62.870*** (1.968)	62.590*** (2.343)	19.070*** (4.963)	45.890*** (11.861)	0.370*** (0.066)	0.333*** (0.076)	0.957*** (0.122)	0.175 (0.262)
Observations	644	529	529	529	644	529	529	529
R <sup>2</sup>	0.070	0.073	0.289	0.314	0.042	0.037	0.153	0.176

Notes: OLS regression estimates of the effects of the 2005 reform on test scores and the rate of failure. "2008" is 1 for students who took the exam in 2008, 0 for those who took it in 2006. "Business administration" is 1 for students of business administration, 0 for those of business education. "2005 reform" is 2008 × business administration. Columns 1–4 report the effect on test scores, Columns 5–6, on the rate of failure. Columns 1 and 5 list the estimates for all observations, including those with missing values. The estimates in Columns 2–4 and 6–8 are based on full observations only. The control variables are included successively. Robust standard errors in parentheses. Significance level: \*  $p < 0.1$ ; \*\*\*  $p < 0.01$ .

TABLE 4. Effects of the 2005 Reform by High School GPA.

	High School GPA					
	Low		Intermediate		High	
	Test Score (1)	Failure (2)	Test Score (3)	Failure (4)	Test Score (5)	Failure (6)
2005 reform	-4.968 (5.196)	0.125 (0.154)	-15.456** (7.135)	0.328* (0.183)	-12.962** (5.673)	0.008 (0.103)
2008	10.732** (4.217)	-0.221* (0.131)	22.168*** (5.766)	-0.366** (0.166)	16.808*** (4.323)	-0.064 (0.093)
Bus. admin.	0.490 (4.449)	0.061 (0.134)	12.155* (7.023)	-0.228 (0.187)	10.273** (5.074)	-0.006 (0.110)
Male	2.727 (2.696)	-0.093 (0.066)	1.152 (2.290)	-0.016 (0.051)	-1.746 (2.167)	-0.010 (0.031)
Foreign	-5.734 (7.321)	0.295 (0.230)	-23.406*** (6.046)	0.335** (0.134)	-15.678*** (5.816)	0.192* (0.114)
HS GPA	6.462 (5.269)	-0.136 (0.124)	24.016** (9.793)	-0.300 (0.203)	18.033*** (5.624)	-0.040 (0.074)
Age	-1.433*** (0.500)	0.028*** (0.008)	-1.149** (0.523)	0.020 (0.015)	-0.602 (0.634)	-0.003 (0.011)
Semester	-2.728 (2.208)	0.131** (0.057)	11.123** (4.857)	-0.240** (0.108)	5.578 (3.916)	-0.089 (0.067)
Constant	89.160*** (23.822)	-0.395 (0.502)	-15.375 (36.726)	1.506** (0.752)	9.953 (26.967)	0.548 (0.438)
Observations	187	187	172	172	170	170
R <sup>2</sup>	0.154	0.148	0.282	0.174	0.224	0.096

Notes: OLS estimates of the effects of the 2005 reform on test scores and the rate of failure depending on whether high school GPA is low, intermediate, or high. High school GPA is low (intermediate, high) if it falls in the first (second, third) tercile. The terciles vary in size because high school GPA is discrete. High school GPAs of 2.7 and below fall in the first, those higher than 3.1, in the third tercile. "2008" is 1 for students who took the exam in 2008, 0 for those who took it in 2006. "Business administration" is 1 for students of business administration, 0 for those of business education. "2005 reform" is 2008 × business administration. Robust standard errors in parentheses. Significance level: \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

TABLE 5. Effect of the 2008 Reform on the Rate of Blank Submission.

	(1)	(2)	(3)	(4)
2008 reform	0.045*** (0.013)	0.053*** (0.014)	0.055*** (0.014)	0.055*** (0.014)
2010	-0.000*** (0.000)	0.000*** (0.000)	-0.001 (0.002)	-0.001 (0.002)
Bus. admin.	0.006 (0.004)	0.003 (0.003)	0.004 (0.005)	0.003 (0.006)
Male			-0.000 (0.011)	0.000 (0.011)
Foreign			0.040 (0.032)	0.041 (0.032)
HS GPA			-0.014 (0.013)	-0.015 (0.014)
Age				-0.001 (0.001)
Semester				0.002 (0.003)
Constant	0.000*** (0.000)	-0.000*** (0.000)	0.039 (0.036)	0.048 (0.064)
Observations	842	768	768	768
$R^2$	0.025	0.032	0.039	0.039

Notes: OLS regression estimates of the effects of the 2008 reform on the rate of blank submission. “2010” is 1 for students who took the exam in 2010, 0 for those who took it in 2008. “Business administration” is 1 for students of business administration, 0 for those of business education. “2008 reform” is  $2010 \times$  business administration. Column 1 lists the estimates for all observations, including those with missing values. The estimates in Columns 2–4 are based on full observations only. The control variables are included successively. Robust standard errors in parentheses. Significance level: \*\*\* $p < 0.01$ .

TABLE 6. Effects of the 2008 Reform on Test Scores and the Rate of Failure.

	Test Score				Rate of Failure			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
2008 reform	1.212 (2.697)	1.288 (2.766)	-1.505 (2.397)	-1.343 (2.395)	-0.041 (0.056)	-0.067 (0.055)	-0.033 (0.053)	-0.037 (0.052)
2010	-19.212*** (2.269)	-19.338*** (2.323)	-18.263*** (1.951)	-18.336*** (1.946)	0.071 (0.048)	0.090* (0.047)	0.076* (0.045)	0.079* (0.043)
Bus. admin.	-1.053 (1.916)	0.284 (1.986)	-1.970 (1.669)	-3.172* (1.691)	0.045 (0.039)	0.040 (0.038)	0.068* (0.035)	0.102*** (0.035)
Blank subm.	-59.932*** (1.544)	-60.641*** (1.436)	-57.500*** (2.720)	-57.623*** (2.586)	0.828*** (0.020)	0.856*** (0.020)	0.809*** (0.041)	0.812*** (0.038)
Male			1.435 (1.174)	1.725 (1.170)			-0.031 (0.024)	-0.037 (0.024)
Foreign			-9.614*** (2.706)	-9.312*** (2.713)			0.177*** (0.056)	0.167*** (0.056)
HS GPA			13.811*** (1.370)	12.250*** (1.411)			-0.175*** (0.028)	-0.136*** (0.029)
Age			-0.914*** (0.341)	-0.914*** (0.341)				0.022*** (0.006)
Semester			-0.241 (2.172)	-0.241 (2.172)				0.056 (0.049)
Constant	79.513*** (1.619)	79.699*** (1.682)	40.989*** (4.212)	68.253*** (10.800)	0.100*** (0.034)	0.082** (0.032)	0.577*** (0.092)	-0.241 (0.231)
Observations	842	768	768	768	842	768	768	768
R <sup>2</sup>	0.361	0.382	0.481	0.491	0.108	0.129	0.196	0.220

Notes: OLS regression estimates of the effects of the 2008 reform on test scores and the rate of failure. "2010" is 1 for students who took the exam in 2010, 0 for those who took it in 2008. "Business administration" is 1 for students of business administration, 0 for those of business education. "2008 reform" is 2010 × business administration. Columns 1–4 report the effect on test scores, Columns 5–6, on the rate of failure. Columns 1 and 5 list the estimates for all observations, including those with missing values. The estimates in Columns 2–4 and 6–8 are based on full observations only. The control variables are included successively. Robust standard errors in parentheses. Significance level: \*  $p < 0.1$ ; \*\*\*  $p < 0.01$ .

TABLE 7. Effects of the 2008 Reform by High School GPA.

	High School GPA											
	Low				Intermediate				High			
	Blank subm. (1)	Test Score (2)	Failure (3)	Blank subm. (4)	Test Score (5)	Failure (6)	Blank subm. (7)	Test Score (8)	Failure (9)			
2008 reform	0.029 (0.029)	1.870 (4.227)	-0.048 (0.108)	0.079*** (0.027)	-7.465* (4.201)	0.046 (0.055)	0.036* (0.019)	-1.600 (4.641)	-0.014 (0.064)			
2010	0.010	-18.826***	0.113	-0.005	-14.521***	0.004	0.003	-18.533***	0.059			
Bus. admin.	0.015 (0.011)	-4.792* (2.763)	0.152** (0.070)	-0.002 (0.006)	-2.361 (3.768)	0.084** (0.043)	0.013 (0.004)	-2.912 (3.970)	0.051 (0.053)			
Male	0.016 (0.016)	3.981* (2.720)	-0.046 (0.063)	0.005 (0.012)	4.560** (2.482)	-0.088** (0.033)	0.006 (0.011)	-3.402 (3.322)	0.029 (0.038)			
Age	-0.001 (0.007)	(2.190)	(0.055)	(0.021)	(1.813)	(0.034)	(0.019)	(2.164)	(0.034)			
	-0.002	-1.004**	0.018**	0.003	-0.930	0.023*	0.003	-0.472	0.010			
	(0.001)	(0.476)	(0.009)	(0.005)	(0.608)	(0.014)	(0.003)	(0.633)	(0.009)			
Foreign	0.288** (0.140)	1.799 (5.973)	0.147 (0.177)	0.023 (0.051)	-14.188*** (4.846)	0.274*** (0.097)	-0.030* (0.017)	-12.134*** (3.872)	0.142* (0.075)			
Semester	0.008 (0.006)	-5.207** (2.595)	0.188*** (0.056)	0.015 (0.012)	6.823* (3.677)	-0.147* (0.080)	-0.008 (0.007)	6.057** (3.013)	-0.020 (0.023)			
HS GPA	-0.001 (0.037)	6.286** (3.184)	-0.069 (0.086)	-0.045 (0.087)	17.216** (8.590)	-0.193 (0.156)	0.009 (0.064)	5.013 (5.983)	0.018 (0.082)			
Blank subm.		-52.537*** (5.331)	0.590*** (0.142)		-58.413*** (1.918)	0.866*** (0.038)		-67.732*** (2.411)	0.925*** (0.033)			
Constant	0.021 (0.101)	98.079*** (15.030)	-0.729** (0.357)	0.027 (0.243)	31.170 (31.114)	0.528 (0.668)	-0.083 (0.253)	68.753** (27.350)	-0.264 (0.355)			
Observations	264	264	264	288	288	288	216	216	216			
R <sup>2</sup>	0.236	0.388	0.136	0.044	0.577	0.354	0.027	0.503	0.257			

Notes: OLS estimates of the effects of the 2008 reform on the rate of blank submission, test scores, and the rate of failure, depending on whether high school GPA is low, intermediate, and high high school GPA. High school GPA is low (intermediate, high) if it falls in the first (second, third) tercile. The terciles vary in size because high school GPA is discrete. High school GPAs of 2.8 and below fall in the first, those higher than 3.2, in the first, those higher than 3.2, in the third tercile. "2010" is 1 for students who took the exam in 2010, 0 for those who took it in 2008. "Business administration" is 1 for students of business administration, 0 for those of business education. "2008 reform" is 2010 × business administration. Robust standard errors appear in parentheses. Significance level: \*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$ .



TABLE 8. Combined Effects of the 2005 and 2008 Reforms.

	Test Score			Rate of Failure			Rate of Blank Submission		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Both reforms	-11.658*** (3.593)	-14.082*** (3.271)	-10.462*** (3.251)	0.122 (0.094)	0.155* (0.089)	0.070 (0.089)	0.056*** (0.014)	0.055*** (0.014)	0.056*** (0.014)
2010	-2.229 (2.838)	-2.411 (2.533)	-2.808 (2.584)	-0.161* (0.083)	-0.159** (0.079)	-0.130* (0.078)	-0.000*** (0.000)	-0.001 (0.002)	-0.002 (0.003)
Bus. admin.	9.825*** (2.939)	7.612*** (2.658)	6.780** (2.825)	-0.101 (0.084)	-0.072 (0.080)	-0.006 (0.082)	-0.000*** (0.000)	-0.000 (0.004)	-0.001 (0.007)
Male		1.438 (1.658)	1.396 (1.431)		-0.015 (0.034)	-0.022 (0.031)		-0.003 (0.014)	-0.003 (0.014)
Foreign		-10.723*** (4.009)	-7.223** (3.447)		0.191** (0.074)	0.142** (0.069)		0.063 (0.049)	0.063 (0.049)
HS GPA		13.466*** (1.625)	12.495*** (1.564)		-0.194*** (0.033)	-0.151*** (0.033)		-0.010 (0.013)	-0.010 (0.014)
Age			-0.347 (0.331)			0.018** (0.009)			-0.000 (0.002)
Semester			0.287 (1.330)			0.064* (0.034)			0.001 (0.003)
Blank subm.			-57.998*** (2.597)			0.824*** (0.038)			
Constant	62.590*** (2.342)	26.098*** (5.039)	36.625*** (11.773)	0.333*** (0.076)	0.856*** (0.122)	0.074 (0.293)	0.000*** (0.000)	0.027 (0.037)	0.037 (0.079)
Observations	575	575	575	575	575	575	575	575	575
R <sup>2</sup>	0.070	0.188	0.388	0.010	0.079	0.202	0.029	0.039	0.039

Notes: OLS regression estimates of the combined effects of the 2005 and 2008 reforms. "2010" is 1 for students who took the exam in 2010, 0 for those who took it in 2006. "Business administration" is 1 for students of business administration, 0 for those of business education. "both reforms" is 2010 × business administration. Columns 1–3 report the effect on test scores; Columns 4–6, on the rate of failure; Columns 7–9, on the rate of blank submission. Columns 1, 4, and 7 list the estimates for all observations, including those with missing values. The estimates in Columns 2–3, 5–6, and 8–9 are based on full observations only. The control variables are included successively. Robust standard errors in parentheses. Significance level: \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

TABLE 9. Effects of both the 2005 and 2008 Reforms.

	Test Score			Rate of Failure			Rate of Blank Submission		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Both reforms	-8.253** (3.516)	-11.011*** (3.180)	-9.923*** (3.226)	0.074 (0.094)	0.109 (0.089)	0.068 (0.088)	0.056*** (0.014)	0.057*** (0.014)	0.057*** (0.014)
2005 reform	-9.541*** (3.548)	-9.467*** (3.095)	-8.476*** (3.112)	0.141 (0.092)	0.139 (0.087)	0.102 (0.086)	0.003 (0.003)	0.003 (0.004)	0.003 (0.004)
2010	-2.229 (2.839)	-2.444 (2.534)	-3.273 (2.581)	-0.161* (0.083)	-0.160** (0.078)	-0.129* (0.078)	0.000*** (0.000)	-0.001 (0.002)	-0.001 (0.002)
2008	17.109*** (2.885)	15.814*** (2.495)	15.009*** (2.523)	-0.251*** (0.082)	-0.234*** (0.077)	-0.207*** (0.076)	0.000*** (0.000)	0.000 (0.002)	-0.000 (0.002)
Bus. admin.	9.825*** (2.939)	7.541*** (2.640)	5.479** (2.760)	-0.101 (0.084)	-0.069 (0.080)	0.002 (0.080)	0.000 (0.003)	0.001 (0.003)	-0.000 (0.004)
Blank subm.	-60.641*** (1.436)	-57.346*** (2.807)	-57.416*** (2.718)	0.856*** (0.020)	0.806*** (0.043)	0.808*** (0.040)	.	.	.
Male	1.225 (1.074)	1.225 (1.074)	1.505 (1.071)	-0.028 (0.023)	-0.028 (0.023)	-0.036 (0.023)	-0.000 (0.009)	-0.000 (0.009)	-0.000 (0.009)
Foreign	-10.527*** (2.627)	-10.527*** (2.627)	-10.425*** (2.639)	0.185*** (0.055)	0.185*** (0.055)	0.177*** (0.055)	0.037 (0.029)	0.037 (0.029)	0.037 (0.029)
HS GPA	14.068*** (1.186)	14.068*** (1.186)	12.903*** (1.212)	-0.192*** (0.025)	-0.192*** (0.025)	-0.154*** (0.025)	-0.011 (0.010)	-0.011 (0.010)	-0.012 (0.011)
Semester	0.501 (1.356)	0.501 (1.356)	0.501 (1.356)	0.057* (0.033)	0.057* (0.033)	0.057* (0.033)	0.000 (0.002)	0.000 (0.002)	0.000 (0.002)
Age	-0.851*** (0.297)	-0.851*** (0.297)	-0.851*** (0.297)	0.021*** (0.006)	0.021*** (0.006)	0.021*** (0.006)	0.000 (0.001)	0.000 (0.001)	0.000 (0.001)
Constant	62.590*** (2.343)	24.557*** (3.960)	47.657*** (9.869)	0.333*** (0.076)	0.857*** (0.104)	0.039 (0.216)	-0.000*** (0.000)	0.030 (0.027)	0.039 (0.052)
Observations	936	936	936	936	936	936	936	936	936
R <sup>2</sup>	0.334	0.448	0.456	0.110	0.184	0.205	0.036	0.042	0.042

Notes: OLS regression estimates of the effects of both the 2005 and 2008 reforms. “2010” and “2008” are 1 for students who took the exam in 2010 and 2008, otherwise 0. “Business administration” is 1 for students of business administration, 0 for those of business education. “2005 reform” is 2008 × business administration; accordingly, “both reforms” is 2010 × business administration. Columns 1–3 report the effect on test scores; Columns 4–6, on the rate of failure; Columns 7–9, on the rate of blank submission. Columns 1, 4, and 7 list the estimates for all observations, including those with missing values. The estimates in Columns 2–3, 5–6, and 8–9 are based on full observations only. The control variables are included successively. Robust standard errors in parentheses. Significance level: \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

TABLE 10. Effects of the 2005 and 2008 Reforms for Subsamples and Matched Observations.

	Effect of 2005 Reform on Test Scores			Effect of 2008 Reform on Rate of Blank Submission				
	3 <sup>rd</sup> semester (1)	Age $\leq$ 26 yrs (2)	Nearest neighb. (3)	Kernel (4)	3 <sup>rd</sup> semester (5)	Age $\leq$ 26 yrs (6)	Near. neighb. (7)	Kernel (8)
Reform	-14.535*** (3.915)	-9.485*** (3.374)	-7.784*** (2.233)	-11.132*** (3.196)	0.056*** (0.014)	0.061*** (0.015)	0.049*** (0.014)	0.052*** (0.015)
Year	17.851*** (3.291)	16.346*** (2.828)	13.957*** (1.264)	16.730*** (2.553)	-0.001 (0.002)	-0.003 (0.003)	0.004 (0.002)	0.000 (0.003)
Bus. admin.	10.773*** (3.751)	8.380*** (2.988)	3.531* (1.978)	5.927** (2.866)	0.003 (0.006)	0.005 (0.006)	0.003 (0.004)	0.003 (0.005)
Male	0.269 (1.424)	0.977 (1.389)	0.093 (1.023)	2.277 (1.385)	0.001 (0.011)	-0.000 (0.012)	-0.000 (0.007)	-0.001 (0.007)
Foreign	-15.193*** (3.752)	-17.217*** (4.013)	-17.485*** (3.779)	-8.944 (6.395)	0.042 (0.033)	0.048 (0.036)	0.075 (0.047)	0.056 (0.038)
HS GPA	14.350*** (1.730)	16.242*** (1.631)	13.768*** (1.158)	15.228*** (1.670)	-0.016 (0.015)	-0.024 (0.016)	-0.005 (0.007)	-0.008 (0.007)
Age	-1.142*** (0.391)		-1.573*** (0.275)	-1.313*** (0.347)	-0.001 (0.002)		-0.000 (0.001)	-0.001 (0.001)
Semester		1.366 (1.643)	1.718 (1.115)	1.034 (1.500)		0.003 (0.006)	0.001 (0.002)	0.001 (0.003)
Constant	50.623*** (12.303)	13.895* (7.695)	61.173*** (7.691)	49.672*** (11.376)	0.060 (0.067)	0.058 (0.057)	0.019 (0.032)	0.034 (0.035)
Observations	444	493	794	502	736	712	1,056	723
R <sup>2</sup>	0.274	0.300	0.389	0.385	0.039	0.044	0.051	0.049

Notes: OLS estimates of the effects of the 2005 and 2008 reforms for restricted samples and matched observations. Columns 1–4 list the effects of the 2005 reform on test scores, Columns 5–8, of the 2008 reform on the rate of blank submission. The estimates in Columns 1 and 5 are for students who took the exam in their third semester only, those in Columns 2 and 6, for students aged 26 years or less. Columns 3, 4, 7, and 8 report estimates when treated are matched with non-treated observations according to propensity scores. For Columns 3 and 7, treated observations are matched with their nearest neighbors within a caliper of 0.04, imposing common support. Non-treated observations are weighted by the frequency with which they are used as a match. For Columns 4 and 8, non-treated observations are weighted in proportion to their proximity to treated observations, imposing common support. For Columns 1–4, year is 1 for students who took the exam in 2008, 0 for those who took it in 2006. For Columns 5–8, year is 1 for students who took it in 2010, 0 for those who took it in 2008. Business administration is 1 for students of business administration, 0 for those of business education. Reform is year  $\times$  business administration. Robust standard errors in parentheses. Significance level: \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

TABLE 11. Comparison of Reform of Business Education to the Reforms Business Administration.

	Test Score			Failure			Blank Submission		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Placebo	-8.819** (3.856)	-7.373** (3.477)	-2.092 (3.669)	0.096 (0.093)	0.076 (0.088)	-0.008 (0.089)	-0.000 (0.019)	-0.002 (0.020)	-0.028 (0.022)
2012	16.273*** (3.237)	12.663*** (2.937)	8.313*** (3.146)	-0.190** (0.084)	-0.145* (0.079)	-0.077 (0.080)	0.031* (0.017)	0.035* (0.018)	0.056** (0.022)
Bus. admin.	9.825*** (2.937)	8.008*** (2.651)	4.249 (2.827)	-0.101 (0.084)	-0.077 (0.079)	-0.019 (0.082)	0.000*** (0.000)	0.002 (0.003)	0.014 (0.011)
Male		-1.851 (1.597)	-1.434 (1.584)		0.005 (0.028)	-0.002 (0.028)		0.005 (0.012)	0.003 (0.011)
Foreign		-15.833*** (4.288)	-15.410*** (4.158)		0.296*** (0.074)	0.289*** (0.071)		0.002 (0.026)	-0.001 (0.025)
HS GPA		15.986*** (1.518)	13.799*** (1.546)		-0.209*** (0.031)	-0.175*** (0.032)		-0.019 (0.012)	-0.012 (0.011)
Age			-1.349*** (0.339)			0.021*** (0.007)			0.003 (0.003)
Semester			-0.962 (0.934)			0.017 (0.018)			0.014 (0.010)
Constant	62.590*** (2.341)	20.454*** (4.802)	63.704*** (11.144)	0.333*** (0.076)	0.891*** (0.118)	0.224 (0.225)	-0.000*** (0.000)	0.049 (0.033)	-0.085 (0.098)
Observations	694	694	694	694	694	694	694	694	694
R <sup>2</sup>	0.042	0.180	0.202	0.021	0.120	0.140	0.008	0.012	0.021

Notes: OLS regression estimates for the placebo test. "2012" is 1 for students who took the exam in 2012, 0 for those who took it in 2006. "Business administration" is 1 for students of business administration, 0 for those of business education. "Placebo" is 2012  $\times$  business administration and thus captures the combined effect of the 2005 and 2008 reforms of the business administration program, which should be cancelled out by the effect of the 2010 reform of the business education program. Columns 1-3 report the effect on test scores; Columns 4-6, on the rate of failure; Columns 7-9, on the rate of blank submission. Columns 1, 4, and 7 list the estimates for all observations, including those with missing values. The estimates in Columns 2-3, 5-6, and 8-9 are based on full observations only. The control variables are included successively. Robust standard errors in parentheses. Significance level: \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .