

Beltz, Philipp; Link, Susanne; Ostermaier, Andreas

Working Paper

Incentives for Students: Evidence from Two Natural Experiments

ifo Working Paper, No. 133

Provided in Cooperation with:

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Beltz, Philipp; Link, Susanne; Ostermaier, Andreas (2012) : Incentives for Students: Evidence from Two Natural Experiments, ifo Working Paper, No. 133, ifo Institute - Leibniz Institute for Economic Research at the University of Munich, Munich

This Version is available at:

<https://hdl.handle.net/10419/73726>

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.

ifo Working Papers

Incentives for Students: Evidence from Two Natural Experiments

Philipp Beltz
Susanne Link
Andreas Ostermaier

Ifo Working Paper No. 133

August 2012

An electronic version of the paper may be downloaded from the Ifo website
www.cesifo-group.de.

Incentives for Students: Evidence from Two Natural Experiments*

Abstract

Incentives are widely used to increase people's effort and thus performance. While academic achievement depends heavily on effort, there is little empirical evidence on how students respond to incentives other than grades and monetary rewards. We draw on two natural experiments that occurred at a major European university and use the difference-in-differences approach to show how program and course policies affect the effort and performance of students. Our findings indicate that students perform worse (i) if their effort is rewarded belatedly, (ii) if their effort has little impact on their final grade, or (iii) if they may resit exams more often and thus less effort is required from them.

JEL Code: I21, I23, I28.

Keywords: Performance, incentives, higher education, natural experiment, difference-in-differences approach.

Philipp Beltz
Munich School of Management
University of Munich
Geschwister-Scholl-Platz 1
80539 Munich, Germany
Phone: +49(0)89/2180-3887
beltz@bwl.lmu.de

Susanne Link
Ifo Institute – Leibniz Institute for
Economic Research
at the University of Munich,
Poschingerstr. 5
81679 Munich, Germany
Phone: +49(0)89/9224-1698
link@ifo.de

Andreas Ostermaier
TUM School of Management
Technische Universität München
Arcisstr. 21
80333 Munich, Germany
Phone: +49(0)89/2892-4076
andreas.ostermaier@tum.de

* We thank Joshua Angrist, Eric Hanushek, Hans-Ulrich Küpper, Alwine Mohnen, Ludger Wößmann, seminar participants at the Ifo Institute for Economic Research, participants of the 74th Annual Meeting of the German Academic Association for Business Research and the 26th Annual Conference of the European Society for Population Economics for helpful comments. Susanne Link gratefully acknowledges financial support from the Deutsche Forschungsgemeinschaft through GRK 801.

I. INTRODUCTION

Improving the academic performance of students is a major concern of researchers and policy makers alike. There is a large body of literature on the drivers of educational achievement, which mainly focuses on the role and importance of institutional resources (Hanushek, 2003). Student characteristics, such as innate ability, were also recognized as important determinants of student performance. However, there is little research on student *effort* and the factors that influence it. This is surprising because it is common knowledge that performance heavily depends on effort and there is no reason to believe that academic performance is an exception. Unlike other drivers of academic performance, including innate ability, effort is both variable and susceptible to incentives, which makes it a primary starting point for leveraging the performance of students.

The purpose of this study is to show that the policies that universities, departments, or instructors adopt when designing programs and individual courses can be used as incentives for effort.¹ Grading and tuition fees are examples of such policies, which, however, include all sorts of prerequisites, requirements, credit points, the number of resits students can take, and the like. We analyze unique data from two natural experiments which occurred at the business school of a major European university. We offer evidence that students perform worse as the time until their effort is rewarded increases; as the number of courses they have to complete increases and the leverage each course offers to their effort decreases; as they may resit exams more often and thus less effort is required from them. We also found that students respond differently to university policies depending on their ability. Like the grading system or the level of fees, the duration of a program and the number of courses and resits are policy choices that universities can and must make. It is crucial to understand for universities, students, and the whole society how these choices affect student performance.

Grades and monetary rewards, such as exemptions from, or refunds of, tuition fees, are particularly obvious examples of incentives in higher education. Both influence the effort and thus performance of students. More specifically, among university students, monetary rewards were found to increase performance, as measured in test scores and other outcomes (Angrist et al., 2002, 2006; Kremer et al., 2009). Among secondary school students, there is evidence that merit-based scholarship programs and rewards raise enrollment rates (Cornwell et al., 2006) or success rates in exit exams required for enrollment (Angrist and Lavy, 2009). There is also evidence that tuition fees might be used to prompt university students to graduate at the scheduled time (Garibaldi et al., 2007; Leuven et al., 2010). These and other studies also point to limits of monetary incentives or sometimes provide

1. We use the word “program” to refer 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, “course” refers to a lecture, tutorial, seminar, or a combination of those, for which one overall grade is reported.

mixed results (Gneezy and Rustichini, 2000; Fryer, 2011). Nevertheless, these findings suggest that students respond to incentives, such as monetary rewards (Levitt et al., 2012). By contrast, there is no empirical evidence, as far as we know, that the program and course policies considered in this study function as incentives.

Student responses to such policies are best measured in experiments. Unfortunately, suitable experiments are hard to set up in practice because policies usually change for all students at once and the same students do not normally take the same course twice. However, reforms of education systems offer rare opportunities to conduct natural experiments (Meghir and Palme, 2005). In this study, we analyze two natural experiments that resulted from the pervasive Bologna reform of higher education in Europe. The business school that we consider offers two similar programs of study, both of which became subject to reforms. While the policies for the first program were changed as early as 2005, the reform of the second program was delayed until 2010. We adopt the difference-in-differences approach to estimate the effect of the modified policies on student performance. The students of the second program serve as a control group, which allows us to pinpoint the responses of the treatment group.

The remainder of this paper unfolds as follows: first, we depict the institutional background and research setting (Section II). We go on to present our data and introduce the econometric framework we chose to analyze them (Section III). For each experiment, we develop expectations, present the main results, test for heterogeneous effects, as well as long-term effects for the second experiment (Sections IV and V). We then test and discuss the robustness of our results (Section VI). We conclude with a discussion of our findings and their implications, as well as suggestions for further research (Section VII).

II. BACKGROUND AND RESEARCH SETTING

II.A. *Student Responses to Incentives*

Academic performance has been modeled as a function of family, peer, and school inputs as well as student characteristics, such as innate ability (Akerlof and Kranton, 2002; Hanushek, 1986, 2002; Hoxby, 2000). A number of models with varying inputs have been proposed and used in empirical research. Most such models focus on institutional or school resources as a means of improving performance. In practice, this translated primarily into providing institutions with more resources and especially into reducing class sizes (Hoxby, 2000). Like for monetary incentives, evidence is mixed (Angrist and Lavy, 1999), and doubts were raised about the effectiveness or at least efficiency of this input-based approach (e.g., Hanushek, 1996, 2003). As a result, interest shifted toward incentives, both for institutions (e.g., accountability policies or competition between universities) and individual students (e.g., Bishop, 1997, 1999; Bishop and Wößmann, 2004; Hoxby, 1994;

Wößmann, 2003).

The purpose of incentives for individual students is to increase their effort and, as a result, performance. This approach relies on the intuitive assumption that performance depends critically on effort. In support of this assumption, the OECD's Programme for International Student Assessment revealed that truancy and inattention, which can be taken to reflect a lack of effort, correlate with poor reading and mathematical skills (Bishop, 2004); conversely, attendance of tutorials was found to enhance performance (Durden and Ellis, 1995). Likewise, in some studies it was possible to attribute the effects of monetary rewards to increased effort (Angrist et al., 2002, 2009). Accordingly, effort is included in the inputs of certain education production functions (e.g., Akerlof and Kranton, 2002; Bishop, 2004; Bishop and Wößmann, 2004).

Unlike other drivers of academic performance, including innate ability, effort is variable in the short run. Students can choose how much effort they make. This choice depends both on the individual and the situation, which makes the effort of students susceptible to the influence of the university. On the whole, students can be intrinsically or extrinsically motivated to learn. In the former case, they find learning itself rewarding, whereas in the latter, they regard it as a means of obtaining other rewards, such as recognition or the prospect of higher earnings. 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 students are exclusively driven by intrinsic motivation, most modifications of the policies will shift their cost–benefit ratio. We examine student responses to two reforms of program and course policies in the context of a quasi-experimental setting, which resulted from the Bologna reform of higher education in Europe.

II.B. Quasi-experimental Setting

In 1999 the education ministers of 29 European countries convened in Bologna and agreed 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 policies. However, not all universities adopted the Bologna system at the same time, and even within universities some programs were revised later than others. Thus, it happened that at some universities

the traditional and the reformed systems coexisted temporarily. These instances offer the chance to compare different types of degree programs and program policies.

This study centers on the business school of a major public university in Germany, which offers undergraduate programs in business administration and business education. Until the Bologna reform, graduates of both programs were awarded what was known as the *Diplom* degree. The *Diplom* programs in business administration and business education were divided into two periods of study, at the end of which students were awarded the *Vordiplom* and the *Diplom* certificates, respectively. While there was no *Vordiplom* degree, the *Vordiplom* certificate was a prerequisite to qualify for the second period of study. Official program policies specified which courses and exams students had to pass in order to earn their certificates and thus their degree. Both the *Vordiplom* and *Diplom* certificates reported an overall grade, which averaged the grades students had obtained in the courses required for that certificate by the program policies and which we refer to as GPA.² It is noteworthy that the *Diplom* GPA did not reflect the courses required for the *Vordiplom*, but only those taken in the second period of study.

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.³ While the *Diplom* and Bachelor programs necessarily differ in content and scope, the first half of the Bachelor program is identical to the first period of study of the *Diplom* program. However, the completion of the first half of the Bachelor program is no longer marked by anything equivalent to the *Vordiplom*. Also, in contrast to the earlier system, now all grades count toward the final Bachelor GPA. In 2008, the new Bachelor program was revised and, as a result, the number of times students were allowed to resit an exam was increased for most courses. Before the revision students were allowed a maximum of three attempts, whereas they may now take exams as often as they want as long as they graduate on time. However, they must not repeat an exam they have passed to improve their grade.⁴

In our study, both the 2005 and 2008 reforms can be considered natural experiments. The students of business administration serve as the treatment group, those of business education, as the control group, and the reforms, as treatments. To measure the effects of

2. More precisely, German universities award grades ranging from 1.0 (“A”) to 4.0 (“D”) as well as 5.0 (“F”). 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, it is not necessary to translate them into grade points. Nevertheless, we use the common term “grade point average” for convenience.

3. 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. We do not consider the Master program in more detail because it is of no interest for our study.

4. The cost of failure is immense. Students who do not pass an exam until their final attempt or do not graduate on time are not allowed to continue on this or any related program at *any* university in Germany. For instance, a student who eventually failed to graduate in business administration in Berlin will not be admitted for the same program in Munich.

the treatments, we observe the performance of students in a specific second-year course in 2006, 2008, 2010, and 2012. This course is compulsory for both business administration and business education students, no matter whether the former were on the *Diplom*, the Bachelor, or the revised Bachelor programs. We chose this course for our study because it is characteristic of business programs in general, as it teaches basic knowledge about companies, and the exam, which consists in a one-hour written assessment, requires both analytical writing and quantitative skills. For an overview, the design of the two experiments is depicted in Figure I.

Insert Figure I about here.

To assess the effect of the reforms, we compare the outcomes of the 2006 and 2008 exams in Experiment 1 and those of the 2008 and 2010 exams in Experiment 2. For both reforms, we check the changes in the performance of business administration students against those of business education students. The latter are an ideal control group as the two programs are very similar, although the business education program offers graduates the option of teaching at vocational schools in addition to applying for management positions. The first period of study was identical for both *Diplom* programs and students of both the treatment and control groups had to take the same exam. However, the program in business education was not revised until 2010. Hence, whereas program policies changed twice for students of business administration since 2005, they remained the same for those of business education.

In 2010 the business education program was reformed in one single step, reiterating both reforms of the business administration program. Thus the 2010 reform of the business education program restored the situation before 2005, when both programs started identically and only diverged as graduation approached. However, the new policies of the business education program took effect only after the 2010 exam. We examine the data from the 2012 exam—that is, after the two experiments—in addition to those from the 2006, 2008, and 2010 exams in order to test the robustness of our results.

III. DATA AND ECONOMETRIC FRAMEWORK

We collected data on the students of business administration and those of business education who took the exam in 2006, 2008, 2010, and 2012. Our study thus covers nearly a decade in the history of the business school or four generations of students. The data were compiled from two sources within the university. The office of the university registrar collects personal data from students, such as age, gender, country of birth, or qualifications obtained before enrollment, when they apply for admission. 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. We finally matched the data from these sources in order to combine information on performance with information on personal characteristics.

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.⁵ Conversely, 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 as noted above, the newly introduced Bachelor program in business education corresponded to the revised Bachelor program in business administration. The database totals nearly 1,600 observations, which split up into three subsamples. Experiment 1 involves the 2006 and 2008 subsamples; Experiment 2, the 2008 and 2010 subsamples. The 2012 subsample helps us test whether the results of the two experiments are robust.

For both experiments, we consider the test scores, the rates of failure, and the rate of blank submission as outcome variables. In addition to data on these variables and the programs students were enrolled in, our database contains information on demographic characteristics including gender, age, and country of birth. It also includes 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 the grade of his or her secondary education certificate (“high school GPA”). The test score ranges from 0 to 120 points; the high school GPA, from 4.0 (“A”) to 1.0 (“D”).⁶ Exams scoring 10 points or less count as blank. Students happen to score slightly above zero because they *decide* to fail after they have started filling in answers sometimes do not bother cancelling (all) their answers because they are sure enough that they will not pass or just forget to cancel some of them. Blank submission, program (“business administration”), gender (“male”), and country of birth (“migrant”) are dummy variables. Table I reports the means and standard deviations or portions for each year and program as well as the differences between the students of both programs.

Insert Table I about here.

Table I shows that between 75 and 90% of the students considered for this study were enrolled in the business administration program in each year, the rest in the business

5. Some students enrolled in the *Diplom* program happened to take the exam only in 2008 (together with those in the Bachelor program) rather than 2006, some in the Bachelor program only in 2010 rather than 2008 (together with those in the revised Bachelor program). These observations were discarded.

6. Students who take the exit exams required to apply for programs at university receive overall grades ranging from 1.0 (“A”) to 4.0 (“D”) to 5.0 and 6.0 (both “F”) in Germany. Those with overall grades worse than 4.0 are not admitted to university. For ease of interpretation, we rescaled the overall grade as described in the text and refer to it as “high school GPA.”

education program. About half of the students of the former and about a third of those of the latter program were female. The proportion of students born abroad ranged from 4 to 11% in business administration, from 0 and 3% in business education. Most students took the exam in their third semester at university, which is at about the age of twenty-three in business administration, twenty-four in business education. The number of students in both programs grew considerably over time. The high school GPAs were about 3.0 (“B”), with major increases from 2008 to 2010 among students of business administration, from 2010 to 2012 in business education. The increases in both the number of students and high school GPAs went along with applications growing faster than capacity, making admission to both programs more competitive. Looking at the dependent variables, the mean test score can be seen to vary considerably between the three exams, ranging from slightly above 60 in 2010 to around 80 in 2012. The rate of blank submission rose strikingly from 0 in 2006 to 6% in 2010 in business administration and remained above zero in 2012 for both programs.

It appears that students are quite homogeneous across both programs. Among the students of business education, the proportions of male students and students from abroad were somewhat lower than among those of business administration. The students of business education were older when they took the exam because many of them receive vocational training before enrolling at university, but they too took it in their third semester. On average, the high school GPAs of students of business education were somewhat lower until 2010, but exceeded those of their fellow students slightly in 2012. Students of business administration scored higher than those of business education in the 2006 exam, but in 2010 the outcomes were reversed. The rates of blank submission were zero for both programs at the beginning. They rose among students of business administration from 2008 to 2010 and among those of business education from 2010 to 2012.

We apply the difference-in-differences approach to assess the effects of each reform. More precisely, we compare the difference between the performances of the students of business administration to that of the students of business education before and after each reform. Performance is measured in terms of test scores, rates of failures, and rates of blank submission. We estimate the following model:

$$Y_i = \alpha + \beta_1 P_i + \beta_2 Y_i + \delta P_i Y_i + \gamma' X_i + \varepsilon_i. \quad (1)$$

Y_i is student i 's performance, while the dummy variables P_i and Y_i indicate his or her program (1, if business administration) and year of examination (1, if after the respective reform). X_i denotes a set of control variables, and ε_i an idiosyncratic error term. In this model, β_1 captures time-invariant group-specific effects, β_2 , the time-effect common to students in both programs, δ , the effect of the reform, and the vector γ , the effects of the

control variables. We control for the effects of gender, migration, age, semester, and high school GPA in order to account for further differences between the groups and improve the precision of our estimates.

Gender may have an effect on test outcomes because women are usually better at solving verbal problems, whereas men cope better with mathematical tasks (Becker et al., 1990; Gneezy and Rustichini, 2004). If this effects exists, then gender can be treated as a proxy for innate ability, which appears as an input in education production functions. However, as the exam considered requires verbal as well as quantitative skills, gender effects are probably cancelled out. Moreover, both the business administration and business education programs are likely to attract generalists rather than specialists. So although we have included gender as a precaution, we do not expect it to influence test scores, failure, or blank submission.

Age can be argued to have both positive and negative effects on academic achievement; it can also be considered another indicator of ability. Students are older when taking the exam because they enrolled at a higher age or because they take it later in their studies. Students who enroll at university later in life are likely to work even harder than their peers in order to make up for their “delay.” They also benefit from greater experience, which increases their performance. On the contrary, students often take the exam later than their peers (e.g., in their fourth rather than third semester at university) because they have trouble coping with the challenges of studying in the first place and consequently perform worse at the exam. In order to capture and separate these opposite effects, we included both age and semester in our analysis.

Migration relates to the family inputs which appear in many education production functions. Immigrants have been found to do worse at university for a variety of reasons, including insufficient command of the language of instruction (e.g., Schneeweis, 2011; Schnepf, 2007). We therefore account for the country of birth (Germany vs. another country) as a proxy for migration. (It should be noted that, strictly speaking, students born to migrant families in Germany may face similar issues like those born abroad, which is not reflected in this proxy.) Migration does *not* refer to exchange students. We consider only students enrolled in the business administration or business education programs, and excluded exchange students because they are not subject to the program and course policies examined in this study.

The high school GPA is the most informative indicator of the innate ability of German students. In many countries, universities admit applicants on the basis of the grade they have achieved in high school exit exams. Examples of such exams include the German *Abitur*, the British A-levels, the French Baccalaureate, or the New York State Regents (Angrist and Lavy, 2009). The *Abitur* grade averages the grades achieved in a dozen subjects and indicates whether students have acquired competencies which are critical

for succeeding at university. Students with a good *Abitur* grade are likely to do better at university. This expectation is corroborated by ample empirical evidence (cf. Richter, 2006).

IV. EXPERIMENT 1: THE 2005 REFORM

IV.A. *The 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 students receive upon completing their program.⁷ The implications of this reform are ambiguous, like the role of the *Vordiplom*. As there was no *Vordiplom* degree, the certificate did not qualify students for a job. It was officially recognized only by the very university that awarded it and both the grades and the GPA were supposed to inform students about their academic record. In practice, however, students used to present their *Vordiplom* certificate to companies when they applied for internships or jobs. Moreover, the *Vordiplom* drew a line under the first period of study in the sense that grades earned until then did not count toward the *Diplom*.

The expected response of students to the 2005 reform depends much on how they perceived the *Vordiplom*.⁸ They may have taken the certificate as what it was supposed to be, that is, a kind of transcript of records. In this case, the grades that they earned until the *Vordiplom* did not matter much to students in the *Diplom* program. This is true in particular for the course and exam considered in this study. Although *Diplom* students needed to pass the exam in order to enter the second period of study, their grade did not really “count,” because it affected only their *Vordiplom* GPA but not their *Diplom* GPA. It was therefore irrelevant to their degree certificate. By contrast, the grade does matter to Bachelor students because it is part of their Bachelor GPA and thus affects their degree certificate. Assuming that students did not take the *Vordiplom* seriously, we would expect that the Bachelor students who took the exam in 2008 would make more effort and, all other things equal, perform better than the *Diplom* students who took it in 2006.

However, if students regarded the *Vordiplom* as similar to a degree, although it was not a degree officially, the effect to be expected from the reform is entirely different. In this case, the *Vordiplom* is better compared to the Bachelor certificate, since the latter replaced the former as the first certificate that students of business administration receive.

7. Students can still retrieve transcripts of records at any time. However, this was also possible before the Bachelor program was introduced.

8. There are no studies on students’ perception of the *Vordiplom* to our knowledge. This is unsurprising given that there was no other alternative than the *Diplom* program before the Bologna reform. Hence, there was nothing the *Vordiplom* could have been compared to.

Like the Bachelor certificate, the *Vordiplom* certificate was then an important reward and served as a voucher or “signal” which is necessary to redeem the extrinsic benefits of studying, such as higher earnings or more prestige (Hanushek, 2002; Spence, 1973). Under this assumption, the reform had two effects. First, it doubled the time until students’ effort was rewarded, because the Bachelor program takes three years, whereas the *Vordiplom* was awarded after 1.5 years. Second, the reform halved the impact of every grade earned in an exam—including the exam considered here—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.

As rational actors, students will account for time when comparing the costs and benefits of studying for an exam. Although studying may be experienced as 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. Hence, they will discount the expected benefits to their “net present value” in order to compare them with the costs (Frederick et al., 2002). In this vein, it has been argued that monetary rewards for students might correct for high discount rates (Angrist and Lavy, 2009). Irrespective of whether the certificate is a reward in itself or a token for the rewards of studying, the time between studying for the exam and receiving the certificate shifts the cost–benefit ratio as students discount future rewards. The value of the certificate decreases if the time until students receive it increases. Accordingly, the reform would have caused them to reduce rather than augment their effort and thus performance.

The second effect of the reform bears on the impact of the course. Since some courses require more time and effort than others, depending on the complexity and amount of their contents, universities usually account for these differences when calculating the number of credit hours or points.⁹ However, the number of credit points of a course are at the same time the weight of the grade obtained in that course in the GPA. While credit points may be *intended* to merely reflect objective differences between courses, they are likely to *mean* something quite different to students. Students usually strive to obtain as good a GPA as possible, given the limitation of resources such as time. A higher number of credit points prompts them to focus on the respective course, because it offers more “leverage” to their effort or the performance which results from their effort.¹⁰ Obviously, students will make more effort for courses with more credit points.

More generally, students will not only consider the number of credit points of one

9. In the European Credit Transfer System (ECTS), “credit points” corresponds to what is more commonly known as “credit hours” in the US. As with credit hours, the number of credit points of a course is supposed to reflect the time students have to spend on it.

10. In the terms of expectancy theory, the number of credit points or “leverage” corresponds to what Vroom (1964, 2005) calls “instrumentality,” that is, performance–outcome expectancy. Accordingly, the link between effort and performance (e.g., the test score) would be described as effort–performance expectancy.

course as compared to others when they choose their level of effort, but also the weight of the course as such. Even a course that carries more credit points than others offers little perceived leverage if the number of courses factored into the GPA is large, and the weight of each course is proportionately small. As a result, students put less effort into each course as the number of courses increases. This seems irrational at first glance, because it is likely to impair the grades obtained in the single courses and, by implication, the GPA. However, this response is perfectly intuitive from a *motivational* perspective, given that low leverage corresponds to a low performance–outcome expectancy. Like the time effect, the reduced leverage of the course considered in this study is likely to decrease the effort and thus performance of students.

In summary, we expect that the 2005 reform either increased or decreased students’ effort and, as a result, their performance, depending on whether they perceived the *Vordiplom* rather as a transcript of records or as similar to a degree certificate. The change in performance, no matter whether it is positive or negative, allows us in turn to infer how students actually perceived the *Vordiplom*.

IV.B. *Effects of the Reform*

In order to pinpoint the effects of the reform, we consider the test scores in the 2006 and 2008 exams, that is before and after the reform came into effect, in Experiment 1. The students of business administration who took the exam in 2006 were enrolled in the *Diplom* program; their grade would not affect the *Diplom* certificate, but account for about 7% of the GPA in their *Vordiplom* certificate, which they regularly received just under half a year after the exam. By contrast, the students who took the exam in 2008 were already in the Bachelor program; their grade would approximate 3% of the GPA in their Bachelor certificate, which they would be awarded about 1.5 years later. Table II reports the estimates of the effects on test scores and the rate of failure according to the difference-in-differences model described above, both with and without accounting for other effects.

Insert Table II about here.

From Column (1), it can be seen that the reform has a significant negative effect on test scores. Column (2) shows that this remains true when other effects are controlled for; in fact, all effects are significant except those of gender and semester. This result confirms, first of all, the expectation that the reform would affect test scores. It thus supports the argument that students make more effort if it is rewarded immediately and has a significant impact on the GPA, because these are two major changes the 2005 reform brought about. Moreover, it suggests that students enrolled in the *Diplom* program took the *Vordiplom* seriously. Although there was no *Vordiplom* degree, it seems that the *Vordiplom* certificate

prompted students more to make an effort in 2006 than did the Bachelor certificate in 2008, but also more than the transcripts of records, which both students in the *Diplom* and the Bachelor program could retrieve continuously.

The positive effects of the year of examination and program imply that, on average, test scores improved between 2006 and 2008 *despite* the reform and that students of business administration outperformed those of business education. Unsurprisingly, male and female students performed equally well, but migrants performed significantly worse than students born in Germany. While age correlates negatively with test score, semester does not. This, however, is unsurprising because both variables catch similar effects. The positive effect of high school GPA indicates that innate ability is a driver of test scores.

According to Columns (3) and (4), the effects on the rate of failure are similar but inverse to those on test scores. As performance improved over time, the rate of failure decreased, which is reflected in the negative effect of the year of examination. Likewise, the effects of migration, age, and high school GPA on the rate of failure are inverse to those on test scores, and they show that migrants, older students, and students with a bad high school GPA are more likely to fail. By contrast, the reform did not produce an effect on the rate of failure and we actually did not have a reason to expect that it would. The effect found in Column (3) disappears when control variables are accounted for. Likewise, there is no reason why the 2005 reform would have affected the rate of blank submission and Table I showed that the rate was indeed zero both in 2006 and 2008. As a result, there are no notable effects on the rate of blank submission that we could report.

IV.C. Heterogeneous Effects

The results of Experiment 1 suggest that the *Vordiplom* was generally perceived as a reward, and its abolition produced a negative effect on student motivation: as the leverage decreased and the time until they received their certificate increased, it became less worthwhile making an effort. However, it would mean to oversimplify the complexity of human motivation if we assumed that all students responded identically to the reform. Whether someone believes that their effort is rewarded does not only depend on whether their performance earns them a reward and whether this reward is worthwhile; it also depends on whether their effort affects their performance in the first place. For students who do not believe that making more or less effort changes their outcome—because the effect of their effort on their performance is random or will not earn them a high-class certificate anyway—the leverage and time effects of the reform do not matter.

This element of motivation is different for poor and good students, although the effects of the reform on leverage and time are the same for all. Good students know that they have the ability to influence their performance. If they make more or less effort, their performance normally increases or decreases. As a result, good students expected to earn a

high-class *Vordiplom* certificate that would help them apply for scholarships, programs at other universities, both at home and abroad, or coveted internships or jobs. Poor students do not have this ability; they anticipate that their performance will be low or depend on how “lucky” they are. (The exam may or may not focus on the contents of the course that a student happens to master, and if there are fewer contents that he or she masters, the probability that these happen to be tested is lower.) Consequently, they do not expect to earn a high-class certificate and the *Vordiplom* certificate did not promise to be rewarding.

Hence, the 2005 reform did not change much for poor students. As the value of their *Vordiplom* certificate was low, there was no sensible decrease when it was replaced with the Bachelor certificate. By contrast, the reform did make a difference for good students. While their effort would still earn them a better test performance, it would hardly earn them a high-class certificate because their performance had much less impact on their certificate than before. We expect therefore that any effect of the reform can only be observed among good students, whereas poor students did not respond to the reform. In order to test this expectation, we divided the students into three groups according to their high school GPA, which we used as a proxy for ability: students having low, middle, and high ability. We calculated the effects on test scores and rate of failure for each group separately, using the difference-in-differences model introduced above. The estimates are depicted in Table III.

Insert Table III about here.

The results suggest that the 2005 reform caused average and excellent students to score lower, whereas it did not affect the test scores of poor students. In fact, there are hardly any significant effects to be found among poor students. This result is intuitive because students who have characteristics that are typically linked to low, average, or high performance are selected into the respective group. (For example, most students born abroad have low high school GPAs and thus fall in the first group.) By contrast, the pattern of effects on test scores among average and excellent students is similar to that for the whole sample reported above in Table II.

Interestingly, the reform turns out to increase the rate of failure among average but not excellent students. Even though the reform may have reduced the motivation of excellent students and caused them to score lower, they were obviously able and had enough motivation to make the effort that was required for them to pass. The effect on average students is noteworthy because it still concerned the bulk of the students who took the exam.

V. EXPERIMENT 2: THE 2008 REFORM

V.A. The Revision of the Bachelor Program

The minor reform of the Bachelor program in 2008 increased the number of times a student could resit the exam, shifting again the cost–benefit ratio. Before the reform, students 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 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, because those who fail have to leave their program and, as noted above, may not continue on the same program even at a different university. As the number of attempts increases, the cost of studying remains the same, but the benefit of succeeding decreases for all attempts except the final one. This may have two effects, both of which compromise overall test performance.

On the one hand, students may generally make less effort and prepare worse for their exam as a result of the reform and their shifted cost–benefit ratio. 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 rather than submit an exam that would probably earn them a bad grade had to make sure that they would not happen to achieve a score that would make them pass. To this end, they either had not to answer at all or they had to cancel their answers before submitting their exam, thus scoring (close to) zero. Like an overall decrease in effort and performance, this implies that students scored lower, failed more often, and took more attempts to pass the exam. However, unlike the general decrease in effort, which is not necessarily 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 effort and thus performance to decrease and the rate of blank submission to increase. The decrease in performance will result in lower test scores and a higher rate of failure. Both effects will be driven by the increase in blank submissions, but they may exceed this effect and thus persist when blank submission is accounted for.

V.B. Effects of the Reform

We examine the outcomes of the exams in 2008 and 2010, that is before and after the 2008 reform of the Bachelor program of business administration came into effect. Students of business administration who took the exam in 2008 could only resit it twice, whereas those who took it in 2010 had the chance to resit it as often as they wanted during their time at university according to the revised Bachelor program policies. We compare the

difference between the test scores obtained by the students of business administration in 2008 and 2010, who were affected by the reform, to the difference between the test scores of the students of business education, who were not, because the program policies for the latter were still the same as in 2006. Table IV reports the effects on test scores as well as the rates of failure and blank submission, again both disregarding and accounting for effects other than those of reform, program, and time (i.e., year of examination).

Insert Table IV about here.

The effects on the rate of blank submission reported in Columns (7) and (8) are as predicted. In particular, the estimates clearly show that the rise in the rate of blank submission, which could already be seen from Table I, can be attributed to the reform. There is no other significant effect in the full model; the minimal trend reported in Column (7) disappears in the presence of control variables. Results are mixed for the effects on test scores and the rate of failure. Like the 2005 reform, the 2008 reform caused test scores to fall. The negative effect of the reform is lower both in magnitude and significance than in Experiment 1, and it can only be observed when controlling for other effects, as can be seen from Columns (1) and (2). By contrast, Columns (4) and (5) reveal that unexpectedly the reform had no effect on the rate of failure.

While the 2008 reform produced a negative effect on test scores, it seems that this effect was driven by the increase in the rate of blank submission. The estimates listed in Column (3) show that the reform had no significant effect on test scores when abstracting from the students who submitted blank exams. In turn, the estimates of the other effects listed in Column (3) are quite similar to those in Column (2). Moreover, the reform did not affect the rate of failure, no matter whether the students who turned in blank exams are considered or not. However, the students who failed in 2010 scored lower than those in 2008, because many students who failed in 2010 submitted a blank sheet. While the reform did not cause more students to fail, the proportion of students who failed *deliberately* increased among those students who failed. Hence, the decline of performance caused by the 2008 reform results from students' choice.

Generally, the estimates have a similar pattern as in Experiment 1, although the program and trend effects appear only when the control variables are included. Table I above shows that test scores decreased on average for both programs from 2008 to 2010, but the decrease was slightly higher among the students of business administration than those of business education. This is reflected in the negative effects of year and program. Like in Experiment 1, the latter effect is driven by the test outcomes in 2010, because in 2008 students of both programs performed equally well, as can be seen from Table I. The effects on the rate of failure and—as far as there are effects—on the rate of blank submission are intuitively inverse to those on test scores. In summary, the reform caused more students to turn in blank exams, but did not cause more students to fail.

V.C. Heterogenous Effects

Like in Experiment 1, we expect that students responded differently depending on their ability. More precisely, we predict that the performance of average students proved more susceptible to an increase in the number of resits than that of both excellent and poor students. Excellent students have the ability to pass the exam and do very well at their first attempt. Hence, they will not normally fail in the first place and need no further attempt to improve their result either. By contrast, whether poor students fail or pass (and probably even do well) depends, once again, on how “lucky” they are. They do not expect that they can score high because they lack the ability and their motivation is therefore categorically low. They tend to answer the exam questions rather than submit a blank paper deliberately, hoping that they may have just passed or done sufficiently well.

By contrast, average students can and will condition their effort on the number of allowed resits and are therefore most likely to respond. Unlike poor students, they are able, and thus motivated, to make an effort and do well at the exam. However, unlike excellent students, they have limited resources, which they must manage carefully. Excellent students can prepare themselves perfectly for several exams scheduled around the same time, whereas average students must prioritize. If there are many exams to take in a short space of time, they will focus on those that carry the greatest weight in the GPA and learn selectively. They will submit a blank paper if they expect to do significantly better at a resit, but if they do not, they will be equally prepared to submit a bad paper and risk a poor test score.

We therefore expect that the 2008 reform caused effort and thus performance to decrease and the rate of blank submission to increase among average students. Conversely, it should not have had any effect on high-performers and low-performers who either did not want or were not able to respond to this intervention. Like in Experiment 1, we divided the students into three groups according to their GPA and estimated our model separately for each of them. The estimates are reported in Table V.

Insert Table V about here.

The responses of students to the revision of the Bachelor program in 2008 were as predicted. The 2008 reform caused neither poor nor excellent students to score lower, fail, or submit blank exam sheets. By contrast, it both decreased test scores and increased the rate of blank submission among average students. (It did not affect the rate of failure among average students, which corresponds to the results for the whole sample presented above.) The reform took pressure from those students because failing does no longer imply the risk of imminent expulsion from university and they now can choose to take the exam again. However, this choice requires students to turn in blank papers because they may not resit the exam once they have passed. By contrast, blank submission was not an

option for poor or excellent students. The former must hope to be lucky enough to pass; the latter do not want to just pass but to perform well and they do not need to take the exam twice to this end.

The reform produced two effects on average students. On the one hand, it caused them to prepare themselves worse, which explains the decrease in test scores. On the other hand, it provided them with the option of turning in a blank sheet. However, this option did *not* protect them from scoring worse, because it does not always make sense to use it. Students who were unsure whether they would do better later or who faced other restrictions (such as further exams they wanted to take at the end of the following semester) had a good reason to prefer submitting a “bad” exam to the trouble of retaking it. That the reform did not cause more average students to fail suggests that they were good at assessing their performance: on the balance, if they were afraid that they would fail anyway, they handed in a blank sheet, otherwise they eventually did not fail.

V.D. Long-term Effects

From a motivational viewpoint, the higher number of resits is rather critical at first glance because it caused performance to decrease. However, it should be noted that the effects of the reform were not as detrimental as predicted. On the one hand, test scores decreased mainly because more students turned in blank exams, whereas we expected all students—including those who did *not* submit blank exams—to prepare worse and thus score lower. Moreover, the reform did not cause more students to fail, despite the increase in blank submissions. On the other hand, it can still be argued that there is a decrease in performance, no matter whether it is only driven by the increase in blank submission. (This argument implies that it is better for a student to fail with a “high” than with a low score.) More importantly, we cannot refute that the reform caused test scores of average students to decrease, who form the major group in our analysis.

Apart from these short-term effects, the increased number of resits may have long-term effects that are even more critical. The main reason why the number of resits was limited until 2008 was to screen out students who were not suitable for the program of their choice. While students used to resent this constraint, it helped them realize early whether they had chosen the right program and possibly decide to embark on a different one. In addition, it prevented them from resitting courses and overusing university capacities. From an economic viewpoint, it saved both university and student resources.¹¹ Strictly speaking, the reform increased the number of resits but did not allow students to take the exam as often as they wanted because they should not take more than three and must not take more than four years to graduate. It is therefore possible that students do not pass

11. Most German universities charge no or at best comparatively low tuition fees. As a result, students do not normally account for the entire costs of studying.

until their fourth year at university and drop out only then.

It might be argued that the reform is beneficial if resits improve students' level of training and increase their likelihood of success. While the costs of the reform may still outweigh its benefits, it would at least help those students who finally pass and would have dropped out otherwise. However, resits may as well decrease motivation, which leads students to perform even worse as they retake the exam more often. More importantly, 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 who fail their first three attempts are more likely to succeed at their fourth or fifth attempts. In order to approach this question, we pooled the data and regressed test scores and the failure rate on the number of attempts needed to pass. (We did not count blank submission as failure because we are only interested in true failure. Students who make no effort to pass and fail deliberately will obviously improve when they do make an effort.) The main results are reported in Table VI.

Insert Table VI about here.

The estimates show that the number of resits influences both test scores and the rate of failure, while the other effects measured are as usual. Students who took more attempts scored generally lower and were more likely to fail than those who succeeded their first attempt. In particular, students do not seem to benefit from their additional experience or learning effects, as one might suppose. These findings suggest that an increase in the number of resits does not really help students while it causes costs to universities. It should be noted that there is adverse selection because students must not resit the exam once they have passed and because blank submission is not considered failure. Thus only those students with low ability (i.e., low high school GPA) and 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. However, this effect is not a bias but has to be taken into account when assessing the reform.

As a limitation, the outcome of interest is not so much the performance in the single course or exam that we consider in our study, but whether students graduate and how well they perform overall. More resits possibly help students perform better in other courses and eventually complete their program. Unfortunately, we were unable to gather data on graduation and final grade of the students in our database. However, as noted above, we chose a course which is fairly characteristic of business programs and general in scope. It is therefore unlikely that the indirect effects of the increased number of resits are entirely different from the direct effects that we examine. By contrast, it is reasonable to assume that performance in this specific exam is a good indicator of the overall performance of students in both programs.

VI. ROBUSTNESS

Our study exploits the fact that the two reforms of the business administration program resemble experiments, where the students of business administration are the treatment group, while those of business education serve as the control group. However, the setting differs from a laboratory experiment in that we do not compare the responses of the same students before and after the treatment, but the responses of different groups of students: those who took the exam before and those who took it after the reforms.

For the difference-in-differences estimates to be accurate, it is critical that everything else except for the treatment was identical for both the treatment and control groups. While it is virtually impossible to account for all conceivable effects which might distort our estimates, the design of the experiments gives us reason to believe that this assumption holds. Students of both programs attended the same course, were taught by the same instructors, used the same textbooks and teaching materials, and their curricula were nearly identical until the exam (cf. Carrell et al., 2010). They lived in the same city, shared similar social environments, and had similar career opportunities. There were no shocks external to the experiment (such as a downturn of the job market for graduates of one but not the other program), which might affect the value of the graduation certificate of either group. Despite minor differences, the descriptive statistics presented above corroborate the argument that the students in both programs are rather similar.

Another critical assumption is that treatment and control groups did not change as a result of the treatment as certain students self-selected into one program rather than the other. In particular, they might have preferred the well-known *Diplom* to the yet unfamiliar Bachelor program. However, it was not possible for students of business administration to avoid the Bachelor program, because admission to the *Diplom* program was closed when the Bachelor program was introduced. While those who had then enrolled the former program were allowed to continue on it, new students had no choice but to enroll the latter. Students could not even go to a different university since the *Diplom* programs in business administration were replaced with Bachelor programs at most universities in Germany at about the same time.

By contrast, students could enroll business education rather than business administration, which implies that they would have chosen a different program to avoid the Bachelor. It seems unlikely, though, that they went to such lengths. While our research benefits from the similarity between both programs, prospective students did not generally know about it and thus would not easily switch to the other program. Applications increased faster than admissions over the period considered. As a result, the number of students as well as their high school GPAs increased in both programs and apparently there is no self-selection toward either program. The distribution of the students to the programs

described in Table I did not change dramatically either, oscillating around 80% for business administration.

The estimates reported above are robust to the inclusion of a number of control variables. Nevertheless, we consider these statistics to account for a variety of differences that might distort our results. The only major differences between the groups that are not controlled for are the reforms, which therefore explain why test performances evolve differently between the students of the two programs. However, test performances should only diverge as long as different policies apply to either program. When the business education program was revised in 2010, the same policies were introduced for students of business education as five and two years before for those of business administration. If the divergence of test performances can be attributed to the reform, it must disappear after the 2010 reform because both programs are roughly identical again. To test this argument, we conducted a placebo test and compared the students who took the exam in 2012 to those who took it in 2006.¹² Table VII displays the estimates.

Insert Table VII about here.

The placebo effect, which interacts the effects of the business administration program and time (2012 v. 2006), is insignificant. This result supports our argument. The test performance of the business administration students decreased from 2006 to 2010 relative to that of the business education students (which even increased during that period) after the (two-step) reform of their program. Accordingly, the test performance of business education students decreased between 2010 and 2012 relative to that of the business administration students, as the same reform applied to their program. It should be noted that high school GPAs increased for both programs while they were reformed, as mentioned above. The rise in high school GPAs is likely to counteract the decrease in performance caused by the reforms. As a result, we may even underestimate the effects of the reforms.

Another characteristic where the groups slightly differ is age, because students of business education often enroll university after taking a vocational training. The effect of age is also especially susceptible to cause bias because of outliers in the group of business education students. To ensure that our results are not biased by outliers, we estimated the effects of both reforms among the students in different percentiles of the age distribution. The effects of the 2005 reform on test scores and of the 2008 reform on the rate of blank submission are depicted in Table VIII.

Insert Table VIII about here.

12. A more conventional placebo test would compare students of both programs who took the exam *before* the 2005 reform, i.e. in 2004. Unfortunately, we were unable to obtain sufficient data from 2004 because the office of the university registrar did not keep records at that time. (The data were collected to decide whether applicants would be admitted, but not saved to a database.) We therefore could not conduct this test.

For the 2005 reform, the estimates in Columns (1)–(3) show that all effects except that of age exhibit the same pattern and even increase in magnitude compared to those for all students in Column (2) of Table II. In particular, this is true for the effect of the reform. By contrast, the effect of age is seen to decrease both in magnitude and significance. Likewise, the estimates for the 2008 reform in Columns (4)–(6) are similar to those listed in Column (8) of Table IV: the only significant effect is that of the reform. These results suggest that the effects reported above are not driven by outliers.

These tests suggest that the effects on test performance can be attributed to the reforms. One might still wonder, however, whether the reforms affected student *effort* and, as a result, performance. Another explanation is that the Bachelor reform produced *confusion* and uncertainty among students and thus caused them to perform worse. While we cannot rule out this explanation, the placebo test shows that the 2010 reform had the same effects on the students of business education as the 2005 and 2008 reforms on those of business administration. If the Bologna process did unsettle students in 2005, it is not so evident that this effect still occurred in 2012 when most programs other than business education had been reformed for years. Similarly, the argument challenges the effect of the 2005 rather than the 2008 reform because even the 2008 reform applied to the third generation of students who enrolled Bachelor programs. Finally, it should be noted that we consider an exam that students used to take in their third semester, when they were already quite familiar with the program and its policies.

In summary, we cannot easily dismiss alternative explanations for the decrease in test performance. However, we find that the setting and the additional tests support our argument that the reforms of the program policies were the cause. We also find that the time lags suggest that this effect was driven by a direct negative effect of the reforms on student effort, even though the confusion that the reforms produced may also have contributed.

VII. 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, 1996, 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.

Each university has to adapt the policies governing its programs to its specific needs and goals. Nevertheless, our results suggest some guidelines applicable to all universities. First, universities could offer students imminent goals, marked by milestones, and certificates seem to serve this purpose very well. After the abolition of the *Vordiplom*, the performance of students decreased even though they had access to their records and could retrieve

transcripts at any time. It seems, therefore, that a transcript of records is not a substitute for a “formal” acknowledgment of student effort, such as a certificate. Of course, the *Vordiplom* also gave students the chance to “start over” after the first half of their program, which is probably motivating in itself, but hardly compatible with the idea of a Bachelor program.

Second, students should be offered a few powerful rather than many weak levers that they can use to improve their GPA. As the number of courses increases, the leverage of each course decreases and, as we have seen, the performance of students drops. It is probably better to set up programs that comprise a small number of important courses. It would be worthwhile to test whether it is sufficient to group courses and thus organize programs in a visibly structured manner. For instance, if a program includes several courses in the subject of calculus, these could easily be grouped into a calculus module. In that way, each course would serve as an important lever for obtaining a good grade for the entire module and in turn the module grade would be perceived as an important lever for achieving a good GPA. Also, if students are awarded a certificate for passing a module, this might be perceived as a milestone.

Third, students should not be given too many chances to resit exams, as it is far from obvious that this is in their own or the university’s interest. The analysis of the second experiment suggests that student performance decreases as the number of possible resits increases. This means that repeating an exam and possibly the same course several times is a waste of capacity. Lack of capacity may preclude prospective students from enrolling on a program or oblige peer students to choose less sought-after courses. Furthermore, it could lead students to stay longer at university rather than enter the job market or finally drop out after repeated failures, but too late to enroll on a different program that might be more suitable to their abilities.

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, George A., and Rachel E. Kranton, “Identity and Schooling: Some Lessons for the Economics of Education.” *Journal of Economic Literature*, 40 (2002), 1167–1201.
- Angrist, Joshua, Eric Bettinger, Erik Bloom, Elizabeth King, and Michael Kremer, “Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment.” *American Economic Review*, 92 (2002), 1535–1558.
- , Eric Bettinger, and Michael Kremer, “Long-Term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia.” *American Economic Review*, 96 (2006), 847–862.
- , Daniel Lang, and Philip Oreopoulos, “Incentives and Services for College Achievement: Evidence from a Randomized Trial.” *American Economic Journal: Applied Economics*, 1 (2009), 136–163.
- , and Victor Lavy, “The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial.” *American Economic Review*, 99 (2009), 1384–1414.
- , and Victor Lavy, “Using Maimonides’ Rule to Estimate the Effect of Class Size on Scholastic Achievement.” *Quarterly Journal of Economics*, 114 (1999), 533–575.
- Becker, William, William Greene, and Sherwin Rosen, “Research on High School Economic Education.” *American Economic Review*, 80 (1990), 14–22.
- Bishop, John H., “The Effect of National Standards and Curriculum-Based Exams on Achievement.” *American Economic Review*, 87 (1997), 260–264.
- , “Are National Exit Examinations Important for Educational Efficiency?” *Swedish Economic Policy Review*, 6 (1999), 349–398.

- , “Drinking from the Fountain of Knowledge: Student Incentive to Study and Learn—Externalities, Information Problems and Peer Pressure.” Cornell University CAHRS Working Paper 04-15, 2004.
- Bishop, John H., and Ludger Wößmann, “Institutional Effects in a Simple Model of Educational Production.” *Education Economics*, 12 (2004), 17–38.
- Carrell, Scott E., Marianne E. Page, and James E. West, “Sex and Science: How Professor Gender Perpetuates the Gender Gap.” *Quarterly Journal of Economics*, 125 (2010), 1101–1144.
- Cornwell, Christopher, David B. Mustard, and Deepa J. Sridhar, “The Enrollment Effects of Merit-Based Financial Aid: Evidence from Georgia’s HOPE Program.” *Journal of Labor Economics*, 24 (2006), 761–786.
- Durden, Garey C., and Larry V. Ellis, “The Effects of Attendance on Student Learning in Principles of Economics.” *American Economic Review*, 85 (1995), 343–346.
- Frederick, Shane, George Loewenstein, and Ted O’Donoghue, “Time Discounting and Time Preference: A Critical Review.” *Journal of Economic Literature*, 40 (2002), 351–401.
- Fryer, Roland G., “Financial Incentives and Student Achievement: Evidence from Randomized Trials.” *Quarterly Journal of Economics*, 126 (2011), 1755–1798.
- Garibaldi, Pietro, Francesco Giavazzi, Andrea Ichino, and Enrico Rettore, “College Cost and Time to Complete a Degree. Evidence from Tuition Discontinuities.” NBER Working Paper w12863, 2007.
- Gneezy, Uri, and Aldo Rustichini, “Pay Enough or Don’t Pay at All.” *Quarterly Journal of Economics*, 115 (2000), 791–810.
- , “Gender and Competition at a Young Age.” *American Economic Review*, 94 (2004), 377–381.
- Hanushek, Eric A., “The Economics of Schooling: Production and Efficiency in Public Schools.” *Journal of Economic Literature*, 24 (1986), 1141–1177.
- , “Rationalizing School Spending: Efficiency, Externalities, and Equity, and Their Connection to Rising Costs.” In *Individual and Social Responsibility: Child Care, Education, Medical Care, and Long-Term Care in America*, Victor E. Fuchs, ed. (Chicago, IL: University of Chicago Press, 1996).
- , “Publicly Provided Education.” In *Handbook of Public Economics*, Vol. 4, Cha. 30, Alan J. Auerbach and Martin Feldstein, eds. (Amsterdam: North-Holland, 2002).
- , “The Failure of Input-Based Schooling Policies.” *Economic Journal*, 113 (2003), F64–F98.
- , and Ludger Wößmann, “How Much Do Educational Outcomes Matter in OECD Countries?” *Economic Policy*, 26 (2011), 427 – 491.
- Hoxby, Caroline M., “Do Private Schools Provide Competition for Public Schools?” NBER Working Paper w4978, 1994.
- , “The Effects of Class Size on Student Achievement: New Evidence from Population Variation.” *Quarterly Journal of Economics*, 115 (2000), 1239–1285.
- Kremer, Michael, Edward Miguel, and Rebecca Thornton, “Incentives to Learn.” *Review of Economics and Statistics*, 91 (2009), 437–456.
- Leuven, Edwin, Hessel Oosterbeek, and Bas van der Klaauw, “The Effect of Financial Rewards on Students’ Achievement: Evidence from a Randomized Experiment.” *Journal of the European Economic Association*, 8 (2010), 1243–1265.
- Levitt, Steven D., John A. List, Susanne Neckermann, and Sally Sadoff, “The Behaviorist Goes to School: Leveraging Behavioral Economics to Improve Educational Performance.” ZEW Working Paper 12-038, 2012.
- Meghir, Costas, and Mårten Palme, “Educational Reform, Ability, and Family Background.” *American Economic Review*, 95 (2005), 414–424.
- Richter, Ansgar, “Intertemporal Consistency of Predictors of Student Performance: Evidence from a Business Administration Program.” *Journal of Education for Business*, 82 (2006), 88–93.
- Schneeweis, Nicole, “Educational Institutions and the Integration of Migrants.” *Journal of Population Economics*, 24 (2011), 1281–1308.
- Schnepf, Sylke V., “Immigrants’ Educational Disadvantage: An Examination Across Ten Countries and Three Surveys.” *Journal of Population Economics*, 20 (2007), 527–545.
- Spence, Michael, “Job Market Signaling.” *Quarterly Journal of Economics*, 87 (1973), 355–374.
- Vroom, Victor H., *Work and Motivation*. (New York: Wiley, 1964).
- , *On the Origins of Expectancy Theory*, pages 239–258. (Oxford: Oxford University Press, 2005).
- Wößmann, Ludger. “Schooling Resources, Educational Institutions and Student Performance: The International Evidence.” *Oxford Bulletin of Economics & Statistics*, 65 (2003), 117–170.

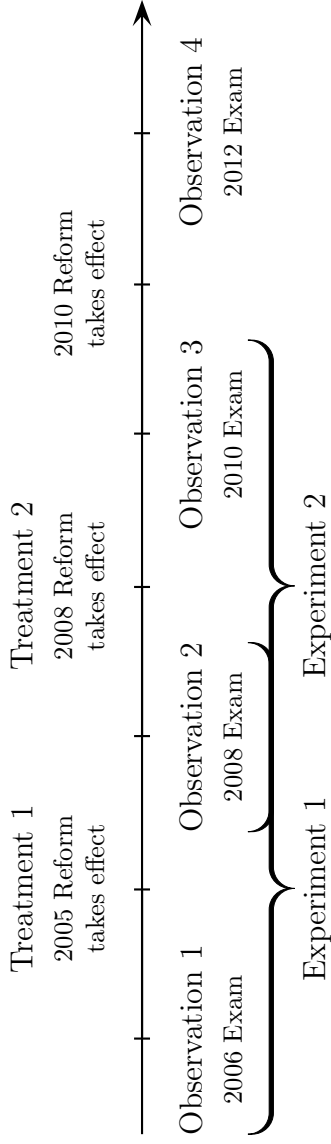


FIGURE I
Timeline of the Natural Experiments

The Bologna reform of higher education in Europe caused the business school of a major European university to reform its programs in business administration and education in 2005–2010. These reforms can be considered as natural experiments.

The 2005 and 2008 reforms applied to the business administration program and are the treatments of the two experiments. The 2005 reform replaced the traditional *Diplom* with a Bachelor program, the 2008 reform revised this Bachelor program. The business education program was not reformed until 2010; the students of business education serve as the control group.

The 2010 reform applied to the business education program, reiterating the 2005 and 2008 reforms of the business administration program in a single step. It thus reestablished the original situation, where the same policies applied to both programs. The 2010 reform is not part of the experiments considered but is used to test for robustness.

TABLE I
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 (3.48)	79.77 (18.39)	79.70 (14.42)	0.07 (2.32)	58.53 (21.99)	60.36 (17.72)	-1.83 (2.25)	79.87 (23.07)	81.62 (22.42)	-1.75 (3.08)
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)	0.12 (0.32)	0.02 (0.04)
Blank submission	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.06)	0.00 (0.00)	0.00 (0.01)	0.06 (0.23)	0.00 (0.00)	0.06 (0.02)	0.03 (0.17)	0.03 (0.18)	-0.00 (0.02)
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)	0.37 (0.49)	0.16 (0.07)
Migrant	0.04 (0.19)	0.00 (0.00)	0.04 (0.03)	0.09 (0.28)	0.03 (0.16)	0.06 (0.01)	0.11 (0.32)	0.02 (0.16)	0.09 (0.30)	0.08 (0.27)	0.02 (0.13)	0.06 (0.02)
Age	22.74 (1.43)	25.51 (3.72)	-2.77 (0.40)	22.84 (2.82)	24.36 (2.38)	-1.52 (0.36)	22.69 (2.09)	24.39 (2.60)	-1.70 (0.24)	22.96 (2.45)	23.58 (2.70)	-0.62 (0.37)
Semester	2.95 (0.63)	3.26 (0.99)	-0.13 (0.13)	3.01 (0.18)	3.14 (0.38)	-0.12 (0.03)	2.98 (0.24)	3.11 (0.64)	-0.13 (0.04)	3.22 (0.88)	3.00 (0.00)	0.22 (0.04)
High school GPA	2.85 (0.55)	2.67 (0.67)	0.17 (0.11)	2.97 (0.45)	2.78 (0.53)	0.18 (0.06)	3.11 (0.50)	2.71 (0.51)	0.39 (0.05)	3.02 (0.44)	3.13 (0.99)	-0.10 (0.13)
Observations	183 [129]	54 [39]	237 [168]	326 [288]	80 [73]	406 [361]	313 [285]	123 [122]	436 [407]	452 [428]	60 [60]	512 [488]

Notes. The table reports data on students of business administration (BA) and business education (BE) from four different cohorts who took an exam on a third-semester introductory course in 2006, 2008, 2010, and 2012 respectively. The first three rows describe means and, in parentheses, standard deviations for the dependent variables (i.e., measures of students' performance in this exam), the next five rows, for the independent variables. Columns (1), (4), (7), and (10) pertain to the students of business administration, Columns (2), (5), (8), and (11) to the students of business education. Columns (3), (6), (9), and (12) indicate the differences between the students of both programs. We report the number of observations with and without missing values because we use both. The number of observations for which complete data were available appears in brackets below the number of all observations, including those with missing values.

TABLE II
EFFECTS OF THE 2005 REFORM

	Test Performance			
	Test score		Failure	
	(1)	(2)	(3)	(4)
Reform	-8.947*** (3.118)	-8.436*** (3.191)	0.144* (0.083)	0.105 (0.086)
2008	16.642*** (2.548)	14.723*** (2.619)	-0.270*** (0.074)	-0.205*** (0.077)
Bus. admin.	7.526*** (2.441)	4.955* (2.852)	-0.093 (0.074)	0.006 (0.081)
Male		0.926 (1.364)		-0.045 (0.029)
Migrant		-15.598*** (3.736)		0.233*** (0.084)
Age		-1.138*** (0.355)		0.022*** (0.007)
Semester		1.587 (1.741)		0.040 (0.042)
High school GPA		15.056*** (1.546)		-0.193*** (0.031)
Constant	62.870*** (1.968)	121.172*** (10.263)	0.370*** (0.066)	-0.788*** (0.221)
Observations	644	529	644	529
R^2	0.070	0.314	0.042	0.176

Notes. Columns (1) and (2) depict OLS estimates of the effects on test scores, Columns (3) and (4) on the rate of failure. The estimates in Columns (1) and (3) reflect the basic model, those in Columns (2) and (4) account for the control variables. The row labeled “2008” captures the trend; the years compared are 2008 and 2006. We let the effect of the program and the year interact to measure the effect of the reform. The effects on the rate of blank submission are not reported because as expected there are none. Robust standard errors appear in parentheses. * $p < 0.1$. ** $p < 0.05$. *** $p < 0.01$.

TABLE III
EFFECTS OF THE 2005 REFORM BY HIGH SCHOOL GPA

	High School GPA					
	Low		Average		High	
	Test Score (1)	Failure (2)	Test Score (3)	Failure (4)	Test Score (5)	Failure (6)
Reform	-0.913 (6.432)	-0.115 (0.180)	-10.573** (4.708)	0.245* (0.140)	-12.320** (4.870)	0.037 (0.040)
2008	7.122 (5.274)	-0.095 (0.157)	17.196*** (3.801)	-0.296** (0.127)	19.385*** (4.158)	-0.030 (0.034)
Bus. admin.	-5.728 (5.658)	0.246 (0.162)	7.038 (4.330)	-0.122 (0.137)	17.764*** (4.762)	-0.039 (0.041)
Male	4.711 (3.358)	-0.145* (0.076)	1.569 (1.877)	-0.038 (0.042)	-3.969* (2.307)	-0.023 (0.031)
Age	-1.602*** (0.512)	0.034*** (0.008)	-1.101** (0.456)	0.016 (0.013)	-0.369 (0.946)	-0.002 (0.012)
Migrant	-9.281 (8.334)	0.381 (0.258)	-18.547*** (5.273)	0.212* (0.119)	-15.458** (6.273)	0.224* (0.124)
Semester	-4.279 (2.694)	0.146** (0.070)	4.668 (2.944)	-0.052 (0.075)	13.417** (6.070)	-0.166 (0.101)
High school GPA	8.348 (7.080)	-0.272* (0.149)	20.052*** (5.220)	-0.321*** (0.121)	12.102** (6.080)	-0.072 (0.061)
Constant	97.736*** (29.111)	-0.386 (0.581)	18.026 (20.564)	1.049** (0.525)	-0.479 (37.210)	0.329 (0.380)
Observations	138	138	265	265	126	126
R ²	0.185	0.226	0.223	0.100	0.280	0.189

Notes. Columns (1) and (2) report OLS estimates of the effects on test performance for students with are low, Columns (3) and (4), with average, and Columns (5) and (6), with high high school GPAs. GPAs in the first quartile are considered low; in the second and third quartiles, average; above the fourth quartile, high. The numbers of students in the quartiles differ slightly from quarters because grades are discontinuous. Columns (1), (3), and (5) depict the effects on test scores; Columns (2), (4), and (6), on the rate of failure. Effects on blank submission are not reported because there are none. Robust standard errors appear in parentheses. * $p < 0.1$. ** $p < 0.05$. *** $p < 0.01$.

TABLE IV
EFFECTS OF THE 2008 REFORM

	Test Performance							
	Test Score				Failure			
	With Blank Submission (1)	W/o B. S. (3)	With Blank Submission (4)	W/o B. S. (6)	With Blank Submission (5)	W/o B. S. (7)	With Blank Submission (8)	Blank Submission
Reform	-1.484 (2.795)	-1.411 (2.394)	-0.004 (0.057)	-0.037 (0.052)	0.007 (0.052)	0.045*** (0.013)	0.055*** (0.014)	
2010	-19.210*** (2.268)	-18.293*** (1.941)	0.071 (0.048)	0.079* (0.043)	0.078* (0.043)	-0.000*** (0.000)	-0.001 (0.002)	
Bus. admin.	-1.421 (1.939)	-3.180* (1.692)	0.050 (0.039)	0.103*** (0.035)	0.105*** (0.035)	0.006 (0.004)	0.003 (0.006)	
Male		1.723 (1.331)		-0.037 (0.026)			0.000 (0.011)	
Migrant		-11.648*** (3.031)		0.200*** (0.059)			0.041 (0.032)	
Age		-0.883** (0.348)		0.022*** (0.006)			-0.001 (0.001)	
Semester		-0.371 (2.155)		0.057 (0.048)			0.002 (0.003)	
High school GPA		13.134*** (1.580)		-0.148*** (0.030)			-0.015 (0.014)	
Constant	79.513*** (1.618)	131.160*** (9.181)	0.100*** (0.034)	-0.926*** (0.193)	-0.943*** (0.192)	0.000*** (0.000)	-0.029 (0.031)	
Observations	842	751	842	751	768	842	768	
R ²	0.210	0.363	0.010	0.112	0.115	0.025	0.039	

Notes. Columns (1)–(3) depict OLS estimates of the effects on test scores; Columns (4)–(6), on the failure rate; Columns (7) and (8), on blank submission. The effects on test scores and the rate of failure reported in Columns (1)–(2) and (4)–(5) are with exams submitted blank, those in Columns (3) and (6) without. The estimates in Columns (1), (4), and (7) reflect the basic model, those in Columns (2), (5), and (8) account for the control variables. The row labeled “2010” captures the trend; the years compared are 2010 and 2008. We let the effect of the program and the year interact to measure the effect of the reform. Robust standard errors appear in parentheses. * $p < 0.1$. ** $p < 0.05$. *** $p < 0.01$.

TABLE V
EFFECTS OF THE 2008 REFORM BY HIGH SCHOOL GPA

	High School GPA											
	Low			Average			High					
	Test score (1)	Failure (2)	Blank (3)	Test score (4)	Failure (5)	Blank (6)	Test score (7)	Failure (8)	Blank (9)			
Reform	2.549 (5.385)	-0.080 (0.127)	0.021 (0.039)	-9.012*** (3.417)	0.060 (0.055)	0.076*** (0.022)	-1.986 (5.808)	-0.022 (0.078)	0.017 (0.018)			
2010	-19.253*** (3.046)	0.098 (0.082)	0.004 (0.010)	-16.081*** (2.798)	0.058 (0.043)	-0.002 (0.004)	-21.388*** (5.013)	0.077 (0.066)	0.008 (0.011)			
Bus. admin.	-6.402* (3.654)	0.157* (0.081)	0.016 (0.021)	-3.457* (2.015)	0.125*** (0.033)	-0.000 (0.006)	-6.102 (5.139)	0.061 (0.061)	0.026 (0.024)			
Male	5.745** (2.841)	-0.125* (0.067)	0.002 (0.009)	3.386* (1.777)	-0.042 (0.033)	-0.003 (0.017)	-6.614** (2.981)	0.037 (0.046)	0.006 (0.020)			
Age	-0.822 (0.621)	0.016 (0.010)	-0.002 (0.002)	-1.232*** (0.432)	0.026** (0.012)	0.000 (0.003)	-0.551 (0.965)	0.007 (0.011)	0.005 (0.004)			
Migrant	-13.951 (10.086)	0.379* (0.195)	0.275 (0.172)	-10.868** (4.570)	0.196** (0.083)	0.042 (0.048)	-13.918*** (4.457)	0.168* (0.091)	-0.026 (0.019)			
Semester	-5.718** (2.837)	0.208*** (0.062)	0.005 (0.004)	3.868 (3.614)	-0.087 (0.072)	0.014 (0.009)	6.187** (2.809)	-0.019 (0.025)	-0.008 (0.008)			
High school GPA	4.025 (5.545)	-0.145 (0.125)	0.001 (0.057)	21.414*** (4.790)	-0.329*** (0.097)	-0.033 (0.048)	2.203 (9.464)	0.043 (0.134)	0.054 (0.080)			
Constant	99.681*** (21.373)	-0.535 (0.426)	0.025 (0.140)	35.204 (21.549)	0.652 (0.471)	0.047 (0.170)	85.539* (45.487)	-0.308 (0.629)	-0.298 (0.373)			
Observations	188	188	188	424	424	424	156	156	156			
R ²	0.278	0.144	0.195	0.346	0.113	0.049	0.366	0.074	0.031			

Notes. Columns (1)–(3) report OLS estimates of the effects on test performance for students with low, Columns (4)–(6), with average, and Columns (7)–(9), with high high school GPAs. GPAs in the first quartile are considered low; in the second and third quartiles, average; in the fourth quartile, high. The numbers of students in the quartiles differ slightly from quarters because grades are discontinuous. Columns (1), (4), and (7) depict the effects on test scores; Columns (2), (5), and (8), on the rate of failure; Columns (3), (6), and (9), on the rate of blank submission. Robust standard errors appear in parentheses. * $p < 0.1$. ** $p < 0.05$. *** $p < 0.01$.

TABLE VI
EFFECTS OF RESITS ON PERFORMANCE

	Test Performance	
	Test score (1)	Failure (2)
Number of Resits	-9.208*** (2.571)	0.156*** (0.044)
Business administration	-0.650 (1.325)	0.008 (0.022)
Male	-0.525 (1.087)	-0.008 (0.018)
Migrant	-10.907*** (2.072)	0.208*** (0.035)
High school GPA	10.621*** (1.066)	-0.114*** (0.018)
Observations	1,327	1,327
R^2	0.228	0.082

Notes. Column (1) reports OLS estimates of the effects on test scores, Column (2), on the rate of failure. Blank submission does not count as failure because we are interested in “true” failure only but not deliberate failure. We pooled the data from 2006 through 2012. The effects of the year dummies are omitted for convenience. Age and semester are not used as control variables because students necessarily increase in age and semester as they resit the exam. Robust standard errors appear in parentheses. * $p < 0.1$. ** $p < 0.05$. *** $p < 0.01$.

TABLE VII
PLACEBO TEST

	Test Performance	
	Test Score (1)	Blank Submission (2)
Placebo	-5.151 (3.905)	-0.013 (0.021)
2012	12.456*** (3.392)	0.038* (0.021)
Business administration	3.001 (3.026)	0.024* (0.013)
Male	-1.328 (1.602)	0.001 (0.011)
Migrant	-16.150*** (4.123)	-0.002 (0.026)
Age	-1.682*** (0.364)	0.004 (0.004)
Semester	-2.754** (1.176)	0.022* (0.013)
High school GPA	12.304*** (1.611)	-0.007 (0.011)
Constant	82.051*** (11.794)	-0.173 (0.114)
Observations	644	644
R^2	0.225	0.033

Notes. Column (1) depicts OLS estimates of the effects on test scores, Column (2), on the rate of blank submission. The row labeled “2012” captures the trend; the years compared are 2012 and 2006, because the business education program was reformed in 2010. We let the effect of the program and the year interact to model the placebo effect. Robust standard errors appear in parentheses. * $p < 0.1$. ** $p < 0.05$. *** $p < 0.01$.

TABLE VIII
EFFECTS OF THE REFORMS BY AGE

	Effect of 2005 Reform on Test Scores			Effect of 2008 Reform on the Rate of Blank Submission		
	90 th (1)	95 th (2)	99 th (3)	90 th (4)	95 th (5)	99 th (6)
Reform	-9.479*** (3.340)	-9.062*** (3.220)	-9.721*** (3.155)	0.060*** (0.015)	0.056*** (0.015)	0.055*** (0.014)
2008/2010	15.951*** (2.789)	15.654*** (2.650)	16.057*** (2.578)	-0.003 (0.004)	-0.001 (0.003)	-0.001 (0.002)
Bus. admin.	6.936** (3.003)	6.377** (2.905)	6.431** (2.904)	0.012 (0.010)	0.007 (0.009)	0.003 (0.007)
Male	1.580 (1.385)	0.851 (1.360)	0.843 (1.372)	-0.003 (0.012)	-0.001 (0.011)	0.000 (0.011)
Migrant	-17.305*** (4.037)	-16.067*** (3.794)	-15.898*** (3.711)	0.049 (0.036)	0.042 (0.033)	0.041 (0.032)
Age	-1.108* (0.584)	-0.807 (0.494)	-0.860* (0.480)	0.006 (0.005)	0.002 (0.003)	0.000 (0.002)
Semester	1.534 (1.653)	1.556 (1.621)	2.325 (1.791)	0.003 (0.006)	0.002 (0.005)	0.003 (0.004)
High school GPA	15.640*** (1.621)	15.130*** (1.543)	15.263*** (1.540)	-0.019 (0.016)	-0.016 (0.015)	-0.015 (0.014)
Constant	119.642*** (14.345)	112.531*** (12.602)	111.693*** (12.377)	-0.087 (0.126)	-0.004 (0.102)	0.046 (0.080)
Observations	493	514	524	712	743	761
R ²	0.306	0.299	0.296	0.046	0.041	0.039

Notes. Columns (1)–(3) report OLS estimates of the effects of the 2005 reform on test scores, Columns (4)–(6), those of the 2008 reform on the rate of blank submission. Columns (1) and (4) account for students whose age is in the 90th percentile; Columns (2) and (5), for students in the 95th percentile; Columns (3) and (6), for students in the 99th percentile. The second row (“2008/2010”) reports the effect of time, which is the effect of 2008 (as compared to 2006) for Columns (1)–(3), and of 2010 (as compared to 2008) for Columns (4)–(6). Robust standard errors appear in parentheses. * $p < 0.1$. ** $p < 0.05$. *** $p < 0.01$.

Ifo Working Papers

- No. 132 Felbermayr, G.J. and I. Reczkowski, International Student Mobility and High-Skilled Migration: The Evidence, July 2012.
- No. 131 Sinn, H.-W., Die Europäische Fiskalunion – Gedanken zur Entwicklung der Eurozone, Juli 2012.
- No. 130 Felbermayr, G.J., A. Hauptmann and H.-J. Schmerer, International Trade and Collective Bargaining Outcomes. Evidence from German Employer-Employee Data, March 2012.
- No. 129 Triebs, T.P. and S.C. Kumbhakar, Management Practice in Production, March 2012.
- No. 128 Arent, S., Expectations and Saving Behavior: An Empirical Analysis, March, 2012.
- No. 127 Hornung, E., Railroads and Micro-regional Growth in Prussia, March, 2012.
- No. 126 Seiler, C., On the Robustness of the Balance Statistics with respect to Nonresponse, March 2012.
- No. 125 Arent, S., A. Eck, M. Kloss and O. Krohmer, Income Risk, Saving and Taxation: Will Precautionary Saving Survive?, February 2012.
- No. 124 Kluge, J. and R. Lehmann, Marshall or Jacobs? Answers to an Unsuitable Question from an Interaction Model, February 2012.
- No. 123 Strobel, T., ICT Intermediates, Growth and Productivity Spillovers: Evidence from Comparison of Growth Effects in German and US Manufacturing Sectors, February 2012.
- No. 122 Lehwald, S., Has the Euro Changed Business Cycle Synchronization? Evidence from the Core and the Periphery, January 2012.
- No. 121 Piopiunik, M. and M. Schlotter, Identifying the Incidence of “Grading on a Curve”: A Within-Student Across-Subject Approach, January 2012.
- No. 120 Kauppinen, I. and P. Poutvaara, Preferences for Redistribution among Emigrants from a Welfare State, January 2012.

- No. 119 Aichele, R. and G.J. Felbermayr, Estimating the Effects of Kyoto on Bilateral Trade Flows Using Matching Econometrics, December 2011.
- No. 118 Heid, B., J. Langer and M. Larch, Income and Democracy: Evidence from System GMM Estimates, December 2011.
- No. 117 Felbermayr, G.J. and J. Gröschl, Within US Trade and Long Shadow of the American Secession, December 2011.
- No. 116 Felbermayr, G.J. and E. Yalcin, Export Credit Guarantees and Export Performance: An Empirical Analysis for Germany, December 2011.
- No. 115 Heid, B. and M. Larch, Migration, Trade and Unemployment, November 2011.
- No. 114 Hornung, E., Immigration and the Diffusion of Technology: The Huguenot Diaspora in Prussia, November 2011.
- No. 113 Riener, G. and S. Wiederhold, Costs of Control in Groups, November 2011.
- No. 112 Schlotter, M., Age at Preschool Entrance and Noncognitive Skills before School – An Instrumental Variable Approach, November 2011.
- No. 111 Grimme, C., S. Henzel and E. Wieland, Inflation Uncertainty Revisited: Do Different Measures Disagree?, October 2011.
- No. 110 Friedrich, S., Policy Persistence and Rent Extraction, October 2011.
- No. 109 Kipar, S., The Effect of Restrictive Bank Lending on Innovation: Evidence from a Financial Crisis, August 2011.
- No. 108 Felbermayr, G.J., M. Larch and W. Lechthaler, Endogenous Labor Market Institutions in an Open Economy, August 2011.
- No. 107 Piopiunik, M., Intergenerational Transmission of Education and Mediating Channels: Evidence from Compulsory Schooling Reforms in Germany, August 2011.
- No. 106 Schlotter, M., The Effect of Preschool Attendance on Secondary School Track Choice in Germany, July 2011.