RUHR

ECONOMIC PAPERS

Friederike Hertweck

Student Performance in Large Cohorts: Evidence from Unexpected Enrollment Shocks

#984

Imprint

Ruhr Economic Papers

Published by

RWI – Leibniz-Institut für Wirtschaftsforschung
Hohenzollernstr. 1-3, 45128 Essen, Germany

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Editors

Editorial Office

Ruhr Economic Papers #984

Responsible Editor: Ronald Bachmann

ISSN 1864-4872 (online) – ISBN 978-3-96973-149-9

Ruhr Economic Papers #984

Friederike Hertweck

Student Performance in Large Cohorts: Evidence from Unexpected Enrollment Shocks

Bibliografische Informationen der Deutschen Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at **http://dnb.dnb.de**

RWI is funded by the Federal Government and the federal state of North Rhine-Westphalia.

http://dx.doi.org/10.4419/96973149
ISSN 1864-4872 (online)
ISBN 978-3-96973-149-9

Friederike Hertweck[1]

Student Performance in Large Cohorts: Evidence from Unexpected Enrollment Shocks

Abstract

This article investigates the effects of large groups of first-year students on individual college performance. The study is based on administrative micro-level data from the universe of higher education institutions in Germany. The empirical strategy exploits shocks in undergraduate enrollment induced by policymakers' underestimation of expected demand for college. The results show that students in large cohorts of freshmen have a lower probability to complete their degree on-time, partly because students take longer to attain their college degree and partly because more students are forced to de-register. The results are relevant for policymakers to promote educational expansion without jeopardizing student performance.

JEL-Code: I20, I23, J08

Keywords: Higher education; cohort size; student performance; schooling reforms; policy evaluation

November 2022

1 Friederike Hertweck, RWI. – I thank Mirco Tonin for discussions throughout the development of this article and the staff at the Research Data Centers in Munich and Fürth for their support and hospitality during on-site data analysis. I am grateful for helpful comments from Silke Anger, Ronald Bachmann, David Card, Bernardo Fanfani, Mathias Huebener, Jan Marcus, Alexander Moradi, Nikolas Schöll, Steven Stillman, Bjarne Strøm, Stephan Thomsen, Amelie Wuppermann as well as workshop and seminar participants at Collegio Carlo Alberto, DIW Berlin, DZHW Hannover, the EEA Virtual, IAB Nürnberg, ifo München, Freie Universität Bozen, RWI – Leibniz Institute for Economic Research Essen, Università Bocconi, Universität Bamberg, NIFU Oslo, and the VfS Annual Conference. I gratefully acknowledge the financial support that I received from Freie Universität Bozen to access the data. - All correspondence to: Friederike Hertweck, Hohenzollernstr. 1-3, 45128 Essen, Germany, e-mail: friederike.hertweck@rwi-essen.de

1 Introduction

Many countries expand access to higher education in order to sustain high levels of human capital and economic growth. Yet, policy-makers need to solve the trade-off between providing more study places, assuring high teaching quality, and complying with tight financial budgets. While most countries experience an increasing number of entrants to higher education, they do not adjust their spending on education proportionally. But if larger groups of first-year entrants had detrimental effects on students, colleges might be better off adjusting their supply-side activities with increased demand.

This paper adds to this debate by showing that larger cohorts of first-year students negatively affect individual student performance. The empirical strategy exploits massive demand-side shocks induced by a series of policy reforms in Germany between 2011 and 2013 in two-way fixed effects settings. In the aftermath of these reforms, policy-makers underestimated the number of freshmen by 15 %. The results show that students who enroll as part of a larger cohort of freshmen have a lower probability of completing an undergraduate degree on time. The unexpected shock in demand in 2011 led to an average reduction in the on-time completion rate by 6.3 percentage points. This drop translates into a reduction by 17.3 %. The effect is stronger for institutions that experienced a larger increase (more than 25 %) in first-year entrants. The drop in on-time completion persists over the years of the demand-side shock.

The reduction in on-time completion is partly due to a higher share of students ultimately failing their course and partly because students extended their study duration. More specifically, students, who finally completed an undergraduate degree, needed on average almost a month longer until graduation. Such an additional month of undergraduate studies may have hindered these graduates from enrolling to a master's degree at the beginning of a term and probably led to a delayed labor market entry. Moreover, the number of students who started their course in 2012 but ultimately failed increased by 2.5 percentage points (or 36.2 %). One interpretation of this deferred response is that institutions tried to adjust student numbers back to a desired level.

In order to interpret the results of the empirical specification as the causal effect of larger entry cohorts on student performance, I rely on parallel trends in outcomes in pre-reform periods and absence of any anticipation behavior. I show that parallel pre-trends are satisfied and possible anticipation did not occur to an extent that threatens the results. In addition, I provide a series of robustness checks to underline the validity of my results. These robustness checks account for possible changes in students' enrollment behavior, spatial spillovers on neighboring states, and temporal spillovers on other groups. I also test the validity of my results with placebo treatments and alternative sample selections. Overall, I can show that the presented results are conservative estimates that are rather downward biased. All effects can be interpreted as lower bounds of the true effect of increased enrollment on student performance.

To estimate the effect of increased enrollment on student performance, I leverage three types of administrative data on the universe of German universities: Student-level data, which include detailed information on all graduates and forced drop-outs for the years 2003 to 2017, the registry of personnel that includes individual-level data on every single employee at a German institution between 2006 to 2016, and each institution's detailed financial accounts from 2006 to 2013. The present study is the first that combines these administrative datasets. By combining colleges' demand- and supply-side data, this study can have a closer look at potential underlying channels.

I can show that those institutions that employed more teaching staff – and thereby had only slight increases in the student-staff-ratio – experienced less detrimental effects on their students than those institutions that could not or were not willing to adjust the number of teaching personnel. However, the number of entrants per academic staff does not fully explain the drop in on-time completion. An additional channel – which I can not control for – can be congestion effects outside college such as competition for cheap housing or student jobs.

This paper contributes to several debates in the literature. First and foremost, it adds to the literature on class and cohort sizes in higher education. Research to date often concentrates on the effects of *class size* on student performance. This strand of the literature agrees that performance decreases in larger classes (e.g., Machado & Vera-Hernandez, 2009; Bandiera, Larcinese, & Rasul, 2010; Monks & Schmidt, 2011; Bettinger & Long, 2018). Previous literature also looked at how the number of study places – and not the number of peers – affect college completion rates (Bound & Turner, 2007; Bound, Lovenheim, & Turner, 2010). Compared to the setting in this paper, the number of study places in Bound and Turner (2007) and Bound et al. (2010) were rather fixed so that colleges could not fully meet the increased demand coming from larger (birth) cohorts.

Limited possibilities for supervised learning, less contacts to faculty, and scarce library resources may lead to more competition among students if more students started college. In turn, Levin (2001) and Dobbelsteen, Levin, and Oosterbeek (2002) argue that students could also benefit from larger groups if they had a larger pool of peers out of which to select a study buddy. Despite policy-makers' and colleges' ridge walks between higher education expansion, quality, and funding, sound evidence on the effects of large groups of first-year students – in contrast to classes – is scarce. The results in this paper provide evidence that larger cohorts of freshmen rather harm student performance.

This paper also contributes to our understanding of educational reforms. It exploits a series of German schooling reforms as a natural experiment that aimed at reducing high school by one year to allow for an earlier labor market entry. Previous research on these reforms concentrates on high school student outcomes under the two different schooling regimes (e.g., Marcus & Zambre, 2019). This paper is different by looking at the longer-term educational outcomes of those students who experienced more competition and congestion during college education.

The results of this paper are relevant for researchers and policy-makers alike. Most research concentrates on class sizes – partly because colleges have the possibility to easily adjust class sizes and partly due to data availability. This paper shows that the size of an entry cohort matters beyond the student-staff-ratio. More research is needed to understand additional channels how cohort sizes affect student performance to allow policymakers to adjust relevant levers for educational expansion without jeopardizing student performance.

The remainder of this paper is organized as follows. Section 2 describes the policy reforms that led to the shock in demand for college education. Section 3 introduces the different administrative data sets that have been linked for the purpose of this study. Section 4 presents the empirical strategy. It discusses the identifying assumptions and addresses the direction of potential bias. The main results and robustness checks are presented and discussed in section 5 followed by a discussion of the underlying channels and policy implications in section 6. Section 7 concludes.

2 Institutional Setting

In Germany, the *Standing Conference of the Ministers of Education and Cultural Affairs of the Länder* (Kultusministerkonferenz or KMK) regularly estimates the number of freshmen to provide institutions with planning criteria. While the KMK accurately forecasted expected demand of study places until 2009, they failed to account for the massively increasing demand in the aftermath of two major policy reforms. These reforms were the shortening of high school by one year and the abolition of compulsory military service. Both reforms are explained in the following.

The German federal states (*Länder*) are autonomous in undertaking educational reforms. Until 2007, German high school students in 14 of the 16 states had to complete a total of nine years of academic high school until they sat the final examinations (G9 schooling regime). From 2007 onwards, some German states undertook the so-called *G8 reform* that reduced the duration of academic high school by one year so that adolescents received their university entrance qualification (UEQ) a year earlier than preceding cohorts.

By design of the reform, each affected state had a year with two cohorts of academic high school graduates: The final cohort of students graduating under the G9 regime and the first cohort graduating under the G8 regime, as illustrated in Figure 16 in Appendix E.1.

In order to avoid negative effects on the total accumulation of human capital during high school, the schooling curriculum was compressed from nine to eight years. Thereby, policy-makers avoided the usual trade-off between an earlier labor market entry and acquisition of additional human capital. In this respect, the G8 reform differs from other educational reforms that have been used as natural experiments in the economics of higher education: The legge no. 685/61 in Italy (Bianchi, 2020) allowed students from technical schools to start a STEM major at universities. Similarly, the double cohorts in the Canadian province of Ontario (Morin, 2015) resulted from the abolition of grade 13. In these two settings, students of different cohorts had different levels of human capital when they enrolled to higher education.

As for Germany, high school graduates received the same UEQ (in terms of human capital acquired) irrespective of the schooling regime. Meyer and Thomsen (2016, 2018) do not identify any significant differences between students from the G8 or G9 regimes in terms of self-reported motivation and abilities. The shock in demand on the higher education market appears to be a uniform shock along the ability distribution.[1]

In addition, an unexpected reform of compulsory military service further increased the number of eligible (male) students: In 2011, the German government decided to put conscription into abeyance. The suspense of military service was unexpected by the population and resulted in a sharp increase in male university entrants in 2011.

The aforementioned two different reforms caused a shock in student numbers in 2011 and even beyond: Figure 1 shows that the number of college entrants has been correctly predicted until 2010 but has been underestimated since 2011. In 2011, the double cohorts in two of the most populated states graduated from high school (see Figure 17 in Appendix E.1). More than 40.000 unexpected undergraduate students enrolled to German universities, leading to an unexpected average increase in cohort sizes of 15 %.

Certainly, the shock in student numbers induced by the schooling reform could have been (partly) foreseen by policy-makers: The double cohorts of high school graduates resulted from a change in the schooling regime that came into effect several years prior to these double cohorts. Yet, colleges were in fact surprised by the demand-side shock for study places and were not prepared to handle the additional numbers of college entrants (Kloepfer, 2011).

3 Data

Data have been provided by the Federal Statistical Office of Germany and cover the universe of German higher education institutions. For each institution, data on (1) students, (2) personnel, and (3) financial accounts are available. These data sets can be linked via year-, institution- and field-identifiers.

The core of the dataset is the Exam Registry (*Prüfungsstatistik*). It is a retrospective cross-section

[1] A large body of literature compares high school students of the different schooling regimes but excludes the double cohorts from their analyses. Please refer to Büttner and Thomsen (2015) for comparisons of acquired competencies, Huebener and Marcus (2017) for grade repetition rates, Quis (2018) for differences in adolescents' mental health, Dahmann and Anger (2014) for personality traits, Westermaier (2016) for drug abuse, and Marcus and Zambre (2019) and Meyer, Thomsen, and Schneider (2019) for the timing of college enrollment. Huebener and Marcus (2015) and Thomsen and Anger (2018) provide sound reviews of studies related to the G8 reforms.

that provides information on each student's demographic background and schooling history. Students enter the registry in the year of their de-registration from college. Usually, students de-register when they graduate from college. However, if a student fails a mandatory exam for several times, a college can also ban that students from continuing his studies (forced de-registration). Students in the Exam Registry can be considered as students who were keen on completing their studies.[2]

The Exam Registry comprises detailed information about a student's university entrance qualification and the academic progress (e.g., detailed subject, degree-awarding institution, previous institution, year of enrollment, type of degree, final GPA) in one of the 278 programs (see Table 6 in Appendix C). Complete individual-level data is available for roughly 90 % of the students in the Exam Registry. Missing observations result from software malfunctioning during the data preparation process at the Federal Statistical Office of Germany. These missing observations are random. Despite its high data quality, the Exam Registry is almost not used for studies at the individual level. The only exception is a recent working paper by Bietenbeck, Leibing, Marcus, and Weinhardt (2020) who identify the effect of tuition fees on educational attainment.

In addition to student-level data, this study aggregates detailed supply-side data from the Financial Statistics of Higher Education Institutions (*Hochschulfinanzstatistik*) and the Registry of Personnel (*Personal- und Stellenstatistik*). These datasets contain detailed financial accounts of every higher education institutions as well as individual-level data of every employee (see Table 1). I aggregate the financial and personnel data on the field-level and link these to the Exam Registry. Due to some inconsistent institutional identifiers not all institutions can be perfectly linked to the student-level data. Hence, the main results will be based on the Exam Registry only.

3.1 Sample Selection

The Exam Registry is a retrospective cross-section. Students enter the dataset in the year of leaving higher education. Based on the registries from all winter and summer terms between 2003 and 2017, I group all individual students on the year-institution-field cell of their initial enrollment to college. I select only those students who enrolled to a Bachelor's degree program in STEM, Social Sciences or Humanities at a public university between winter term 2007/08 and winter term 2013/14.[3] By conditioning on Bachelor's programs, I automatically exclude highly selective fields such as Medicine and Law as well as the field of Education that follows state-specific curricula. In a second step, I exclude all students who first enrolled to a program in a state with an early or late adoption of the schooling reform.[4] I also exclude mature students (i.e., above the age of 25) and those who had not received a university entrance

[2] Two additional groups of students in Germany are *ghost students* and *voluntary drop-outs*: Low tuition combined with large financial benefits for registered students (e.g., free public transport, cheap health care) lead to a considerable gap between the number of registered students and those committed to their studies. Some institutions estimate that more than 20 % of their enrolled students are *ghost students*, subject-specific shares are even higher (Jannasch & Olbrisch, 2014; Landtag Nordrhein-Westfalen, 2019). Also, subject-specific voluntary drop-out rates vary between 33 % in cultural disciplines and 8 % in medicine. Heublein, Schmelzer, and Sommer (2008) estimate an average dropout rate of 21 % over all higher education institutions for 2006.

[3] Although data are available for the graduates since 2003, I restrict the sample to college entrants who enrolled in or after 2007 for two reasons. First, some states charged a tuition fee of up to 1000 Euros per academic year since 2007. Any earlier cohorts would have studied under a different institutional setting (see Appendix E.1 for details). Second, the Bologna Process required Germany to convert their degrees into the two-tier system of Bachelor's and Master's degrees. In 2007, most institutions had adapted to the new degree system so that the majority of students enrolled to reformed undergraduate programs (see Figure 25 in Appendix G for details).

[4] More specifically, I exclude the states of Saxony-Anhalt, Mecklenburg-Vorpommern, Saarland, and Hamburg because these had a larger cohort of high school graduates between 2007 and 2010. I furthermore exclude the state of Hesse due to their step-wise implementation of the schooling reform. Please see Figure 17 for the timing of the larger cohorts across the German states.

qualification (UEQ) when enrolling to college.[5]

Bachelor's degree programs are designed to be completed within three years. An extension by another year is possible without any penalties. To allow for comparability across cohorts despite the right-censored nature of the Exam Registry, I restrict the sample furthermore to students that completed their degree within these four years.

Descriptive statistics of students in the sample

The final sample consist of $N = 152,244$ students who enrolled to an undergraduate program in one of eleven German states between 2007 and 2013 and de-registered by the end of 2017.[6] The number of students increased over the period of interest and peaked in 2011 (see Table 2) as expected from the underestimation of cohort sizes illustrated in Figure 1. The sharp drop in student numbers in 2013 is due to the sample selection procedure and the nature of the data: Students are only included in the sample if they completed their degree within four years or had to de-register. The drop in Table 2 indicates that, compared to pre-2013 freshmen, a much smaller number of students completed the program within four years. Possible explanations are an extended stay at college (i.e., > 4 years), voluntary changes of the program and voluntary dropout. Additional years of data - which are not yet available - could shed light to the reasons why less people completed their degree. In the results section, I discuss how the sample selection might affect regression results.

Table 2 also provides information on the demographic and educational background of the students: 42 % to 48 % of entrants to higher educations are female. The vast majority start college in winter term and hold the most generic university entrance qualification (UEQ), i.e., the German *Allgemeine Hochschulreife* or *Abitur*. 31 % to 47 % of the high school graduates did not start college in the year of graduating from high school. This is partly due to military service until 2011. Only 3 to 4 % of the students completed vocational training prior to enrolling.[7] Table 2 also shows some college outcomes: 33 % to 47 % completed their undergraduate degree on-time. An average of 8 % to 15 % of the students were forced to de-register, for instance as a result of failing a compulsory exam for too many times. The GPA in Germany ranges from 1 (excellent) to 4 (passed) and varies between 2.11 and 2.16 over the sample years. The mean GPA is averaged over all students who successfully graduated from college. Tables 7 and 8 in Appendix C.2 provide these descriptive statistics for the subsamples of graduates and forced de-registrants. A larger share of female than male students successfully completed their undergraduate education. Other individual characteristics are similar across the graduates and the dropouts.

Institutional characteristics over time

Figure 2 illustrates the entry cohort size, the headcount, and the expenses of colleges over the relevant years. The entry cohort sizes peaked in 2011 (panel A in Figure 2) on the level of subjects and programs. Panel A in Figure 2 also shows a drop in cohort sizes in 2013. This drop is, as discussed before, most likely caused by data availability and the construction of the sample.

Panel B in Figure 2 provides the average headcount of the institutions: The number of professors slightly increased by an overall 5 % from 2007 until 2012 while the number of research and student

[5] Highly talented secondary school students sometimes get the opportunity to take university classes besides attending high school. This is, however, rare: A total of 179 observations of early entrants were deleted during sample selection.

[6] Descriptive statistics and results remain stable even if alternative sample restrictions are employed. Please refer to Appendix D.1 for details.

[7] The smaller share of students with a gap year who enrolled to college in 2013 can be due to data availability. Students who completed higher education in 2013 and had a gap year started their college education in 2014. If they needed four years to complete their undergraduate education, they de-registered in 2018. However, when writing this article, data has only been available until 2017.

assistants increased by 25 to 35 % until 2012 (see panel B in Figure 2). Over the entire sample period, the expenses for buildings and personnel increased (panel B in Figure 2). Complete descriptive statistics of the institutions are in Appendix C.2.

4 Empirical Strategy

The relationship of interest is whether the size of a student cohort has an effect on a student's performance at college. The unexpected shock in student numbers (see section 2 and Figure 1) is deemed to be a natural experiment and allows to run an event study with three leads and three lags of the form

$$y_{ijkst} = \sum_{t=2007}^{t=2009} \beta_t DC_s + \sum_{t=2011}^{t=2013} \beta_t DC_s + \mu_t + \gamma X_i + \nu_s + \lambda_k + \epsilon_{ijkts} \tag{1}$$

where the outcome y_{ijkst} is a measure of performance of student i who enrolled in any course of field k at a college j in state s in year t. A student's study performance is operationalized by the probability of completing the degree on-time (i.e. within three years) and by the probability to be forced to leave the program because of failing (at least one) compulsory exam several times. ν_s and λ_k are state- and field-specific fixed effects. The vector X_i contains a set of individual student characteristics. ϵ_{ijkts} is an unobserved error term that is assumed to be uncorrelated with the covariates. I compare the student performance in states that had a schooling reform and thereby a much larger cohort between 2011 and 2013 ($DC_s = 1$) to students in those states only affected by the military reform ($DC_s = 0$), i.e., without a schooling reform between 2007 and 2013.

The coefficients β_t capture the causal effect of the strong population shock resulting from the reforms at high school on top of the shock induced by the military service on the aforementioned measures of student performance over time. The β_t can be interpreted as weighted averages of field- and institution-specific heterogeneous causal effects. The latter is important because the nationwide reform of the military service affected only male students and thereby mostly male-dominated fields. This is discussed in more detail in the results.

4.1 Identifying Assumptions

The ideal set-up to evaluate the effects of larger cohorts on student outcomes is a randomized experiment in which cohort sizes are randomized over institutions and fields. However, randomizing cohort sizes is concomitant with ethical problems because an experimental setting might affect economic and even social outcomes of the participants by hindering some students from studying. Relying on quasi-experimental evidence can solve the ethical dilemma. By using the schooling reform as a shock to student numbers, I take advantage of changes in cohort sizes that were not implemented by policymakers to measure student performance.

In order to interpret the results of the event study specification as the causal effect of larger entry cohorts on student performance, I rely on two key identifying assumptions. First, there must have been parallel trends in outcomes in pre-reform periods and second, neither students nor colleges must have changed their behavior in anticipation of the larger cohorts. This is discussed in more detail in the following.

Parallel trends

Under the assumption of parallel trends, the difference in student performance is the same across all colleges and all years in the absence of the reforms, i.e., $E[y_{ijkst}(DC = 1) - y_{ijkst}(DC = 0)|X] = 0$. Figure 3 illustrates the changes in outcomes in pre-reform year between the treatment and the control states. The trends appear rather parallel across states for both outcome variables. A detailed analysis of the pre-trends is performed and discussed in Appendix D.5. The results of these tests for pre-trends reveal that pre-trends are parallel.

Selection effects

Another set of threats to identification might come from selection effects: Students who graduated from high school as part of a larger cohort could have tried to avoid the run on colleges by either postponing their college entry or by starting college in another state. To understand individual responses to the reforms, several static and dynamic two-way fixed effect models are estimated that compare states with the expected schooling reform to states that were not expecting any increase in student numbers. Please see Appendix C.3 for details.

The results of this exercise suggest that students who graduated from high school as part of a double cohort were 2.5 percentage points less likely to enroll to a college outside their home state unless they lived in a border district. Also, they were up to 7.9 percentage points less likely to delay their university entry (see Table 9 in Appendix C.3). Overall, students did not react much to the increase in cohort sizes induced by the schooling reform. The remaining doubts are dispelled in a battery of robustness checks accounting for selection effects on the side of the students as well as spatial spillovers from affected to unaffected states (see Appendices D.3 and D.4).

Policy responses

Policy responses in anticipation of the schooling reform can lead to confounding factors that are correlated both with the timing of the reforms and with the outcome variable of interest. Yet, the schooling reforms were not implemented to induce large cohorts or to promote educational expansion but to allow for an earlier labor market entry by reducing the years of schooling.

In order to deal with the larger cohorts, an administrative agreement between the Federal Ministry of Education and the states was signed that allowed for additional funding of 38.5 billion Euros from 2007 to 2020 to support colleges in coping with the shock in student numbers. The so-called Pact for Higher Education (*Hochschulpakt 2020*) led to more funds to colleges.[8] However, a report by the supreme authority for federal audit matters in Germany provides that part of the funds were not used to create more study places (Bundesrechnungshof, 2020).

Spillover on preceding and subsequent cohorts

The unit of analysis are students who started their undergraduate education in year t at a college j. These students are, however, not isolated from students at the same college j who start their degree in year $t - 1$ or $t + 1$. In particular, if a student fails an exam, he or she may attend the class again in the subsequent year. This may lead to spillovers from one cohort to another. If students from the reform

[8] In addition, the Federal Ministry of Education and Research introduced the Pact for Teaching Quality (*Qualitätspakt Lehre*) in 2011 that provided funds to higher education institutions on a competitive basis to improve teaching independently from the larger cohorts. A report commissioned by BMBF (2020) shows that the funds were not used to deal with larger cohorts but to engage in innovative teaching methods.

year had a slower study progress, the subsequent cohorts may have experienced larger classes too. These spillover effects on subsequent cohorts are captured by including the lags in the event study design.

In addition, students of preceding cohorts can have experienced competition for scarce resources such supervisory time in lecturer's office hours or desks in the library. If this competition led to a slower study progress or even to failing the program, the rate of on-time completion might have dropped in pre-reform years while the rate of failing might have increased. This effect can be present for treatment and control groups because both groups experienced a shock in student numbers from the suspense of military service. Figure 3 point into that direction: On-time completion decreases in pre-reform years while the rate of failing increases. Because these trends are parallel across treatment and control group, the possibility of spillover effects on precending cohorts does not challenge the empirical identification. Still, I account for these in a robustness check (see Appendix D.3).

Sample selection

A final threat to identification lies in the nature of the data. The Exam Registry is a retrospective cross-section and is thereby right-censored: Students enter the dataset *after* their de-registration. For the main analysis, I exclude all students who had a study duration of more than four years. The main reason for excluding these students is that every student who extend an undergraduate degree by more than a year has to pay fees for being a *long-term student*.

The exclusion of students who study for more than four years in their undergraduate course potentially leads to downward biased coefficients because these students did obviously not complete their degree on-time. In a robustness check, I extend the maximum study duration to five years. The effects are similar in magnitude (see Table 10 in Appendix D.1) and indicate that four years is a reasonable threshold for including students in the sample.

4.2 Direction of potential bias

The assumption that neither prospective students nor colleges change their behavior in anticipation of the treatment, i.e. $E[y_{it}(DC = 1) - y_{it}(DC = 0)|X_i] = 0$ and $E[y_{jt}(DC = 1) - Y_{jt}(DC = 0)|X_j] = 0$ for all $t < 2011$ may not hold due to the hitherto stated selection effects and policy responses. These threats to identification, however, rather underestimate the true effects: If colleges received funds to cope with larger cohorts, financial and personnel resources per student may not change so that a null effect is expected.

Similarly, if students move away or delay their college education, this voluntary selection out of a larger cohort lowers the actual entry cohort size and may thereby not affect the performance of the remaining students as long as the average ability of students remain constant. Although I can not control for average ability because data lack any measure of performance at high school, I provide evidence from survey data that high school performance does not differ much between students who enroll to college in the year of graduating from high schoool and those who enroll later. If at all, the voluntary selection creates positive selection which would lead to an underestimation of the true effect (see Appendix E.1).

If spillover on one preceding cohort existed so that the baseline cohort of 2010 already experienced a lower rate of on-time completion and a higher rate of failing, the lagged coefficients would be downward biased. If multiple preceding cohorts were already affected by the larger cohorts, a slight trend may be visible over the β_t in the leads.

Altogether, the regression results for the coefficients of interest, i.e., the lagged β_t in Equation 1, are rather downward biased and may therefore provide lower bounds of the true effects of cohort sizes on student performance.

5 Results

The larger cohorts of college entrants in treated states in 2011 led to an average reduction in the on-time completion rate of these cohorts by 6.3 percentage points. This decrease translates into a reduction of the on-time completion rate by 17.3 %. The effects for 2012 and 2013 are similar in magnitude: The on-time completion rate for the entry cohort of 2012 decreased by 4.5 percentage points and for the entry cohort of 2013 by 7.7 percentage points, respectively (see Figure 4 and Table 3). Results for 2013 must be interpreted cautiously because data are right-censored.

As already discussed in section 4.2, the results for the baseline regression are rather lower bounds of the true effects because also the colleges in the control group may have experienced a shock in male students due to the suspense of military service. Provided that the suspense of military service affected mostly male-dominated fields of study, I run a subsample analysis only for the female-dominated fields of Humanities and Social Sciences. Figure 5 provides the event study plots for the fields of Humanities and Social Sciences that comprise the highest and second highest share of female students. The subsample analysis reveals that in the female-dominated field of Humanities the magnitude was – as expected – much larger: The larger cohorts lead to a decrease in on-time completion by 16.4 percentage points for the students who started college in 2011 and 15.3 percentage points for the cohort of 2012. Similarly, the larger cohorts in the Social Sciences caused a decrease in the on-time completion by 10.9 percentage points for the cohort of 2011 and by 6.2 percentage points for those students who enrolled to college in 2012.

A lower rate in on-time completion can be caused by more students who leave college before graduating but also by an extended study duration. The probability to fail a program did not change in 2011, but increased on average by 2.5 percentage points in 2012 (see column (2) in Table 3) for the pooled regressions. Due to the overall small fraction of students who are forced to de-register, an increase by 2.5 percentage points translates into an increase by 36.2 %. In addition, students may have voluntarily dropped out of college due to larger cohorts. Voluntary dropouts are, however, not available in the data. The field-specific subsample analysis reveals again that female-dominated fields were much more affected: Forced de-registrations in the Humanities increased by 1.6 and 3.7 percentage points for the cohorts of 2011 and 2012, respectively, but lack statistical singificance. In the Social Sciences, forced de-registrations did not change for the cohort of 2011 but increased statistically significantly by 1.8 percentage points for the cohort of 2012.

In another set of regressions, I use the average study duration as an additional outcome. The results in column (3) in Table 3 provide that the study duration of the entry cohorts of 2011 and 2012 increased by an average of three weeks (or 12.3 % of a term) and the study duration of the cohort of 2013 by more than six weeks (or 24.6 % of a term). Overall, one in eight students extended his or her undergraduate education by a semester as a result of the unexpected run on colleges in 2011 and 2012. This extension is likely to result in a postponed labor market entry or a delayed enrollment to a master's course.

Overall, the results show that larger cohorts affect students: With more students entering college, a smaller fraction can complete a degree on time.

5.1 Robustness of results

I conduct a series of robustness checks to provide supporting evidence for the validity of the presented negative impact of larger entry cohorts on student performance. All robustness checks are explained in detail in Appendix D.

First, I follow two alternative sampling procedures. For the first set of alternative sampling, I restrict the sample to those students who completed their studies within five years, thereby allowing for an additional year until graduation or de-registration. The reason for extending the sample to those students who studied up to five years is grounded in the fact that students might have needed even more than a year longer. A caveat of this approach lies in the retrospective nature of the Exam Registry: Data are only available for students who de-registered until 2017. Allowing five years of studying requires an exclusion of all students who started their undergraduate course in 2013 or later. The results from estimating Equation 1 with the adjusted sample point into the same direction as the hitherto presented baseline results (see Figure 10 in Appendix D.1).

In a similar manner, I restrict the sample to those students who completed their studies within four years but remove the cohort of 2013 due to the data issues adressed before. The results in Appendix D.1) show that the on-time completion rate significantly drops in the aftermath of the larger cohorts. The estimated coefficients are slightly larger in magnitude. The statistical significance does not change between the baseline sample and the restricted sample that exclude the cohort of 2013.

Second, I account for selection out of larger cohorts: I restrict the entire sample to those students who started college in the year and state of graduating from high school. Any student who delayed college education or migrated to another state is thereby excluded. Although there is low student mobility in Germany (see Figure 20 in Appendix E), delaying college entry is quite common. The remaining subsample comprise only of a quarter of the initial observations.

The results in Appendix D.2 show that larger cohorts still negatively affect the on-time completion rate. The results are slightly smaller in magnitude though: In the reduced sample, larger cohorts lead to a reduction in the on-time completion rate by 4.4 percentage points compared to a reduction of 6.3 percentage points in the full sample. The results for forced de-registrations remain statistically insignificant.

Third, I show that spillovers on earlier cohorts would downward bias the baseline results. Therefore, I exclude two pre-periods from the analysis and show that the baseline normalization causes a mechanic shift in estimated coefficients (see Appendix D.3). I discuss that, if temporal spillovers on pre-2011 cohorts existed, i.e., if students of pre-reform entry cohorts were affected by the spike in freshmen in 2011, the baseline results would in fact underestimate the true effects.

Fourth, I account for spatial spillovers on neighboring states (see Appendix D.4). The stylized facts in section C.3 provide suggestive evidence that students of the larger high school cohorts were more likely to go to college in another state if they completed high school in a border district. I run the event study from Equation 1 separately for each treated state with and without neighboring states in the control group. The results do not change substantially in magnitude and significance. The robustness check reveals that spatial spillover neither drive the results for on-time completion nor for forced de-registrations.

Fifth, I run a placebo check for parallel trends in which I assign a placebo treatment a year ahead, i.e., to the college entrants of 2010. The results in Appendix D.5 show a small and insignificant effect for the placebo treatment. The results of this robustness check reveal that the baseline results are not driven by any significant pre-trends in outcomes.

Finally, I use alternative empirical approaches to address the relationship of interest (see Appendix D.6). Instead of estimating Equation 1, I first follow a standard difference-in-differences approach (DiD) in which I compare states with the schooling reform to those without a schooling reform in pre- and post-reform years. The DiD estimate in the regression that employs the students' on-time completion as the outcome variable provides that the average effect in post-reform years is a reduction in the on-time completion rate by 6.4 percentage points. This is similar to the effect for 2011 from the baseline event study regression. The strategy does not provide a statistically significant effect for the probability of forced de-registration. In a second set of alternative empirical approaches, I run an instrumented difference-in-differences regression in which I use the treatment status in and around the year of birth of each student as instruments for being treated when entering higher education. The instruments are weak and the resulting coefficients, although pointing into the expected direction, are statistically not significant.

6 Channels

The results have provided a negative link from increased cohorts of first-year college entrants to their on-time college completion. This result highlights the importance of understanding congestion effects inside colleges. In this section, I discuss potential underlying channels.[9]

6.1 Entrants per academic staff

A high number of entrants (and students) per academic staff can lead to congestion effects inside colleges because students may receive less supervised learning, shorter slots in office hours, and delayed responses to e-mail inquiries. To understand whether congestion effects can explain the negative link between cohort size and student performance, I match the student-level data with data on academic staff and look at the hetereogeneity in the ratios of entrants per staff across institutions in different ways: In a first exercise, I run the baseline regression from Equation 1 seperately for each college located in a treated state. I then group these colleges into three categories depending on the direction and statistically significance of the coefficients of interests β_{2011}, β_{2012}, and β_{2013}.

Figure 7 in Appendix A.4 illustrates the mean entrants-per-staff ratios for these groups. The figure provides some interesting insights into the potential mechanism of congestion effects inside colleges: It shows that the number of college entrants per staff was much higher in 2011 than in 2012 and that institutions that experienced a negative impact on their on-time completion rate in 2011 are those colleges that had a relative small number of professors and temporary staff to deal with the larger cohorts (as indicated by large entrants-per-staff-ratios in Panel A in Figure 7). This is suggestive evidence that less faculty per student negatively affects student performance.

In a second exercise I run the baseline regression from Equation 1 not separately for each college but for groups of colleges that experienced an increase in the entrants-per-staff ratio to various extents. For each college, I calculate the increase in the ratios for the cohort of 2011 over that college's increase in entrants per staff for the years 2007 to 2010[10]:

$$\Delta_{j,2011} = \frac{\sum_{i_j=1}^{N_j} i_{j,2011}}{\sum_{l_j=1}^{L_j} l_{j,2011}} \cdot \left(\frac{1}{4} \sum_{t=2007}^{t=2010} \frac{i_{j,t}}{l_{j,t}} \right)^{-1} \tag{2}$$

where $i_{j,t} = \{1, ..., N_j\}$ are individual entrants at college j in year t and $l_{j,t} = \{1, ..., L_j\}$ are individual academic staff at college j in year t. I then run Equation 1 for colleges that experiences any increase, i.e., $\Delta_{j,2011} > 0$, an increase of at least 10%, i.e., $\Delta_{j,2011} \geq 0.10$, of at least 25%, i.e., $\Delta_{j,2011} \geq 0.25$ or of at least 50%, i.e. $\Delta_{j,2011} \geq 0.50$.

The results in Figure 8 in Appendix A.4 as well as in more detail in Tables 22, 23, 24, and 25 in Appendix F suggest that students at colleges with a lower number of academic staff per student

[9] In order to analyze potential channels, I have to merge student-level data with institutional data. Some identifiers are missing so that all results must be interpreted cautiously. See also section 3 and Appendix C for a detailed description of the data and how the datasets have been linked.

[10] The reason is that overall participation in higher education has been increasing over the years (see also figure 25 in Appendix G).

experienced a larger drop in the on-time completion rate. The strongest effect can be found for entrants at colleges in 2011 where the number of entrants per professor increased by 50 % or more compared to the average increase in the entrants-per-staff-ratio over pre-reform years. Those students experienced an average reduction in on-time completion by 10 percentage points.

It must be reminded that the sample only contains students who either completed their undergraduate degree or were forced to de-register. Students who voluntarily dropped out of their course do not appear in the sample. However, some colleges might have experienced an increase in entrants but also a larger number of voluntary drop-outs. In such a case, I would not be able to identify an increase in entrants in the data.[11]

All in all, the results suggest that on-time completion ratio is lower at colleges with fewer academic staff per entrant. However, it can not be stated with certainty, that an increase in students per academic staff *causes* a decrease in on-time completion.

6.2 Institutional adjustments

Additional factors such as competition for scarce resources (e.g., library books or quiet study rooms) can also affect student performance. Unfortunately, the data do not allow to account for all these college-specific factors. However, I can account for time-invariant college-specific characteristics such as a better equipped library by including college fixed effects in the main regression from Equation 1 instead of fixed effects for states.

The regression results in Table 26 are very similar to those of the baseline results and thereby indicate that college-specific heterogeneity does not drive the results. Yet, it can not be identified whether colleges reacted differently to the larger cohorts and, if they did, whether possibly heterogeneous adjustments systematically affected student performance. More research is needed to understand the link between a college's resources and student performance.

6.3 Congestion effects outside colleges

Students could have also experienced more competition outside colleges, e.g., on the housing market or the market for student jobs. If students needed more time to find appropriate accommodation (or a student job), they might have missed introductory classes and had to re-take some of the first-year classes.

A piece of suggestive evidence is offered by the German Student Unions' data on student dorms: The data show that the number of publicly funded student dorms increased by 6.83 % between 2008 and 2017 while the number of enrolled students increased by 34.61 % over the same period. The lack of student dorms led to a drop in the accommodation rate from 12.13 % in 2008 to 9.62 % in 2017 (DSW, 2017). Overall, student dorms are scarce in Germany (as suggested by the accomodation rate) and were not expanded in line with the increasing numbers of students.

Data on student jobs as well as additional data on student housing are not available. I leave these channels for future research.

[11] This might be the reason why the on-time-completion rate of the entry cohort of 2011 decreased also at those colleges in treated states that did not experience a stronger-than-average increase in students over staff in that year (see column (2) in Tables 22, 23, 24, and 25 in Appendix F).

7 Conclusion and Outlook

A recently growing literature examines the effects of *class sizes* on student performance in higher education. Despite an increase in higher education participation, few is known about effects of larger *entry cohorts* on student performance. Determining a causal relationship is challenging because selection issues cause severe endogeneity concerns.

I use variation in entry cohort sizes induced by a series of schooling reforms at the secondary school level to understand the link between larger cohorts and student performance in a two-way fixed effects setting. I can show that an unexpected increase in the number of entrants to college negatively affected the on-time completion. More precisely, students in the larger cohorts were 6.3 percentage points (in 2011), 4.5 percentage points (in 2012) and 7.7 percentage points (in 2013) less likely to complete their undergraduate degree on-time. The lower on-time completion is partly due to an increased number of students who were forced to drop-out and partly because students extended their study duration by – on average – three weeks. A battery of robustness checks supports the validity of the results.

I can show that one underlying channel are student-staff ratios: The on-time completion ratio is lower at colleges with fewer academic staff (professors and teaching assistants) per entrant. Beside congestion inside colleges, students may also have experienced more competition for student housing and student jobs. Those students who required more time to find appropriate accommodation and a student job might have missed introductory classes or had to re-take some of the first-year classes. More research is needed to fully understand these channels. I leave these channels for future research.

References

Adam S. Booij, E. L., & Oosterbeek, H. (2017). Ability Peer Effects in University: Evidence from a Randomized Experiment. *Journal of Labor Economics*, *84*(2), 547-578. doi: 10.1093/restud/rdw045

Autorengruppe Bildungsberichterstattung. (2018). *Bildung in Deutschland: Ein indikatorengestützter Bericht mit einer Analyse zu Bildung und Migration.* Bielefeld: wbv Media.

Bandiera, O., Larcinese, V., & Rasul, I. (2010). Heterogeneous class size effects: New evidence from a panel of university students. *The Economic Journal*, *120*(549), 1365-1398. doi: 10.1111/j.1468-0297.2010.02364.x

Bettinger, E., & Long, B. T. (2018). Mass Instruction or Higher Learning? The Impact of College Class Size on Student Retention and Graduation. *Education Finance and Policy*, *13*(1), 97-118. doi: 10.1162/EDFP_a_00221

Bianchi, N. (2020). The indirect effects of educational expansions: Evidence from a large enrollment increase in university majors. *Journal of Labor Economics*, *38*(3), 767-804. doi: 10.1086/706050

Bietenbeck, J., Leibing, A., Marcus, J., & Weinhardt, F. (2020). Tuition fees and educational attainment. *Center for Economic Performance Discussion Paper No. 1839*.

BMBF. (2020). Evaluation des Bund-Länder-Programms für bessere Studienbedingungen und mehr Qualität in der Lehre ("Qualitätspakt Lehre"). Abschlussbericht über den gesamten Förderzeitraum 2011-2020. *Study commissioned by Federal Ministry of Education and Research (BMBF), Project Team: Michelle Andersson, Christoph Besch, Susanne Heinzelmann, Uwe Schmidt and Katharina Schulze*.

Bound, J., Lovenheim, M. F., & Turner, S. (2010). Why Have College Completion Rates Declined? An Analysis of Changing Student Preparation and Collegiate Resources. *American Economic Journal: Applied Economics*, *2*(3), 129-157. doi: 10.1257/app.2.3.129

Bound, J., & Turner, S. (2007). Cohort crowding: How resources affect collegiate attainment. *Journal of Public Economics*, *91*(5), 877-899. doi: 10.1016/j.jpubeco.2006.07.006

Brändle, T., & Lengfeld, H. (2015). Erzielen Studierende ohne Abitur geringeren Studienerfolg? Befunde einer quantitativen Fallstudie. *Zeitschrift für Soziologie*, *44*(6), 447-472. doi: 10.1515/zfsoz-2015-0605

Büttner, B., & Thomsen, S. L. (2015). Are We Spending Too Many Years in School? Causal Evidence of the Impact of Shortening Secondary School Duration. *German Economic Review*, *16*(1), 65-86. doi: 10.1111/geer.12038

Bundesrechnungshof. (2020). Bericht an den Haushaltsausschuss des Deutschen Bundestages nach § 88 Abs. 2 BHO über die Prüfung der zweckentsprechenden Verwendung restlicher Hochschulpaktmittel und der Bedingungen des Zukunftsvertrags Studium und Lehre stärken. *Kapitel 3003 Titel 632 05*.

Carrell, S. E., Fullerton, R. L., & West, J. E. (2009). Does your cohort matter? measuring peer effects in college achievement. *Journal of Labor Economics*, *27*(3), 439-464. doi: 10.1086/600143

Dahmann, S., & Anger, S. (2014). The impact of Education on Personality: Evidence from a German high school reform. *IZA Discussion Paper No. 8139*.

Destatis. (2011). *Bildung und Kultur: Nichtmonetäre hochschulstatistische Kennzahlen, 1980-2009. Fachserie 11, Reihe 4.3.1.* Wiesbaden: Statistisches Bundesamt.

Destatis. (2012a). *Bildung und Kultur: Nichtmonetäre hochschulstatistische Kennzahlen, 1980-2010. Fachserie 11, Reihe 4.3.1.* Wiesbaden: Statistisches Bundesamt.

Destatis. (2012b). *Bildung und Kultur: Nichtmonetäre hochschulstatistische Kennzahlen, 1980-2011. Fachserie 11, Reihe 4.3.1.* Wiesbaden: Statistisches Bundesamt.

Destatis. (2014a). *Bildung und Kultur: Nichtmonetäre hochschulstatistische Kennzahlen, 1980-2012. Fachserie 11, Reihe 4.3.1.* Wiesbaden: Statistisches Bundesamt.

Destatis. (2014b). *Bildung und Kultur: Nichtmonetäre hochschulstatistische Kennzahlen, 1980-2013. Fachserie 11, Reihe 4.3.1.* Wiesbaden: Statistisches Bundesamt.

Destatis. (2015). *Bildung und Kultur: Nichtmonetäre hochschulstatistische Kennzahlen, 1980-2014. Fachserie 11, Reihe 4.3.1.* Wiesbaden: Statistisches Bundesamt.

Destatis. (2016). *Bildung und Kultur: Nichtmonetäre hochschulstatistische Kennzahlen, 1980-2015. Fachserie 11, Reihe 4.3.1.* Wiesbaden: Statistisches Bundesamt.

Destatis. (2019a). *Kommunale Bildungsdatenbank: D13.1: Anzahl der Klassenwiederholungen.* Wiesbaden: Statistisches Bundesamt.

Destatis. (2019b). *Kommunale Bildungsdatenbank: D15.1: Schulabgangsquote an allgemeinbildenden Schulen.* Wiesbaden: Statistisches Bundesamt.

Destatis. (2019c). *Schulabsolventinnen/-absolventen und Schulabgänger/-innen nach Art des Abschlusses, Ländern und Geschlecht. Fachserie 11, Reihen 1 und 2.* Wiesbaden: Statistisches Bundesamt.

Dobbelsteen, S., Levin, J., & Oosterbeek, H. (2002). The causal effect of class size on scholastic achievement: distinguishing the pure class size effect from the effect of changes in class composition. *Oxford Bulletin of Economics and Statistics*, *64*(1), 17-38.

DSW. (2017). Wohnraum für Studierende: Statistische Übersicht 2017. *Deutsches Studentenwerk (DSW)*.

Görlitz, K., & Gravert, C. (2018). The effects of a high school curriculum reform on university enrollment and the choice of college major. *Education Economics*, *26*(3), 321-336. doi: 10.1080/09645292.2018.1426731

Grözinger, G. (2017). Einflüsse auf die Notengebung: eine Analyse ausgewählter Fächer auf Basis der Prüfungsstatistik. In V. Müller-Benedict & G. Grözinger (Eds.), *Noten an Deutschlands Hochschulen: Analysen zur Vergleichbarkeit von Examensnoten 1960 bis 2013* (p. 79-116). Wiesbaden: Springer Fachmedien Wiesbaden. doi: 10.1007/978-3-658-15801-9_2

Hans-Peter Blossfeld, H.-G. R., & von Maurice, J. (2011). Education as a Lifelong Process – The German National Educational Panel Study (NEPS). *Zeitschrift für Erziehungswissenschaft: Sonderheft 14*.

Hübner, M. (2012). Do tuition fees affect enrollment behavior? Evidence from a 'natural experiment' in Germany. *Economics of Education Review*, *31*(6), 949-960. doi: 10.1016/j.econedurev.2012.06.006

Heublein, U., Schmelzer, R., & Sommer, D. (2008). Die Entwicklung der Studienabbruchquote an deutschen Hochschulen. *HIS Projektbericht*.

Horstschräer, J., & Sprietsma, M. (2015). The effects of the introduction of bachelor degrees on college enrollment and dropout rates. *Education Economics*, *23*(3), 296-317. doi: 10.1080/09645292.2013.823908

Huebener, M., & Marcus, J. (2015). Empirische befunde zu auswirkungen der g8-schulzeitverkürzung [DIW Roundup: Politik im Fokus]. *DIW Roundup No. 57*(57).

Huebener, M., & Marcus, J. (2017). Compressing instruction time into fewer years of schooling and the impact on student performance. *Economics of Education Review*, *58*, 1-14. doi: 10.1016/j.econedurev.2017.03.003

Jannasch, S., & Olbrisch, M. (2014). Ich bin doch nicht blöd! *UniSpiegel*, *2014*(2).

Kloepfer, I. (2011). Studenten im Kino. *Frankfurter Allgemeine Sonntagszeitung*, *15.10.2011*.

KMK. (2005). Prognose der Studienanfänger, Studierenden und Hochschulabsolventen bis 2020. *Statistische Veröffentlichungen der Kultusministerkonferenz (KMK)*, *Dokumentation Nr. 176 - Oktober 2005*.

KMK. (2012). Vorausberechnung der Studienanfängerzahlen 2012 - 2025: Erläuterung der Datenbasis und des Berechnungsverfahrens. *Statistische Veröffentlichungen der Kultusministerkonferenz (KMK)*, *Dokumentation Nr. 197 - Juli 2012*.

Landtag Nordrhein-Westfalen. (2019). Phantomstudenten an staatlichen Hochschulen in Nordrhein-Westfalen. *Drucksache 17/6385 vom 28.05.2019*.

Leibniz Institute for Educational Trajectories (LIfBi). (2018). FDZ-LIfBi Codebook NEPS Starting Cohort 5 - First-Year Students: From Higher Education to the Labor Market. *Scientific Use File Version 10.0.0.*.

Levin, J. (2001). For whom the reductions count: a quantile regression analysis of class size and peer effects on scholastic achievement. *Empirical Economics*, *26*, 221-246.

Machado, M. P., & Vera-Hernandez, M. (2009). Does class-size affect the academic performance of first year college students? *Working Paper*, 1–29.

Marcus, J., & Zambre, V. (2019). The effect of increasing education efficiency on university enrollment: Evidence from administrative data and an unusual schooling reform in germany. *Journal of Human Resources*, *54*(2), 468-502. doi: 10.3368/jhr.54.2.1016.8324R

Meyer, T., & Thomsen, S. L. (2016). How important is secondary school duration for postsecondary education decisions? Evidence from a natural experiment. *Journal of Human Capital*, *10*(1), 67–108. doi: 10.1086/684017

Meyer, T., & Thomsen, S. L. (2018). The role of high-school duration for university students' motivation, abilities and achievements. *Education Economics*, *26*(1), 24-45. doi: 10.1080/09645292.2017.1351525

Meyer, T., Thomsen, S. L., & Schneider, H. (2019). New Evidence on the Effects of the Shortened School Duration in the German States: An Evaluation of Post-secondary Education Decisions. *German Economic Review*, *20*(4), e201–e253. doi: 10.1111/geer.12162

Monks, J., & Schmidt, R. M. (2011). The impact of class size on outcomes in higher education. *B.E. Journal of Economic Analysis and Policy*, *11*(1). doi: 10.2202/1935-1682.2803

Morin, L. P. (2015). Do men and women respond differently to competition? Evidence from a major education reform. *Journal of Labor Economics*, *33*(2), 443–491. doi: 10.1086/678519

Quis, J. (2018). Does compressing high school duration affect students' stress and mental health? evidence from the national educational panel study. *Jahrbücher für Nationalökonomie und Statistik*, *238*(5), 441-476. doi: 10.1515/jbnst-2018-0004

RDC. (2019a). Research Data Centres of the Federal Statistical Office and Statistical Offices of the Federal States. *Hochschulfinanzstatistik, survey years 2006 - 2013.*

RDC. (2019b). Research Data Centres of the Federal Statistical Office and Statistical Offices of the Federal States. *Personal- und Stellenstatistik, survey years 2006 - 2016.*

RDC. (2019c). Research Data Centres of the Federal Statistical Office and Statistical Offices of the Federal States. *Statistik der Prüfungen, survey years 2003 - 2017.*

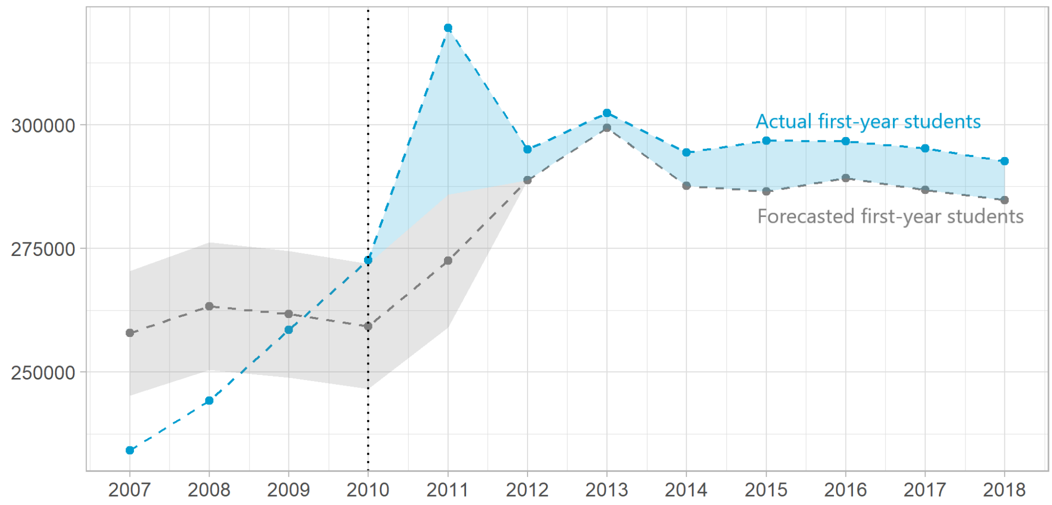
Thomsen, S. L., & Anger, S. (2018). Die notwendigkeit ökonomischer politikberatung für eine evidenzbasierte bildungspolitik: Verkürzung und verlängerung der schulzeit am gymnasium. *Perspektiven der Wirtschaftspolitik*, *19*(3), 167-184. doi: 10.1515/pwp-2018-0026

Westermaier, F. (2016). The Impact of Lengthening the School Day on Substance Abuse and Crime: Evidence from a German High School Reform. *DIW Berlin Discussion Paper No. 1616.*

A Figures

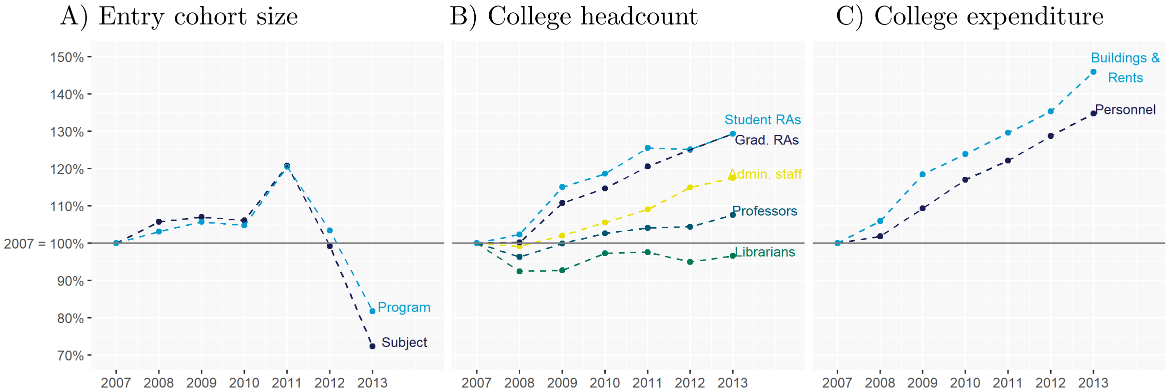
A.1 Institutional setting and sample descriptives

Figure 1: Underestimation of expected student numbers



Note: Actual (blue) and forecast (gray) number of first-year entrants to colleges. The gray-shaded area illustrates upper and lower bounds of the forecast until 2011. From 2012 onwards, the KMK no longer published upper and lower bounds. The blue-shaded area illustrates the KMK's underestimation of the number of first-year entrants (Bachelor degree) to universities in Germany. Own illustration based on data from KMK (2005); KMK (2012).

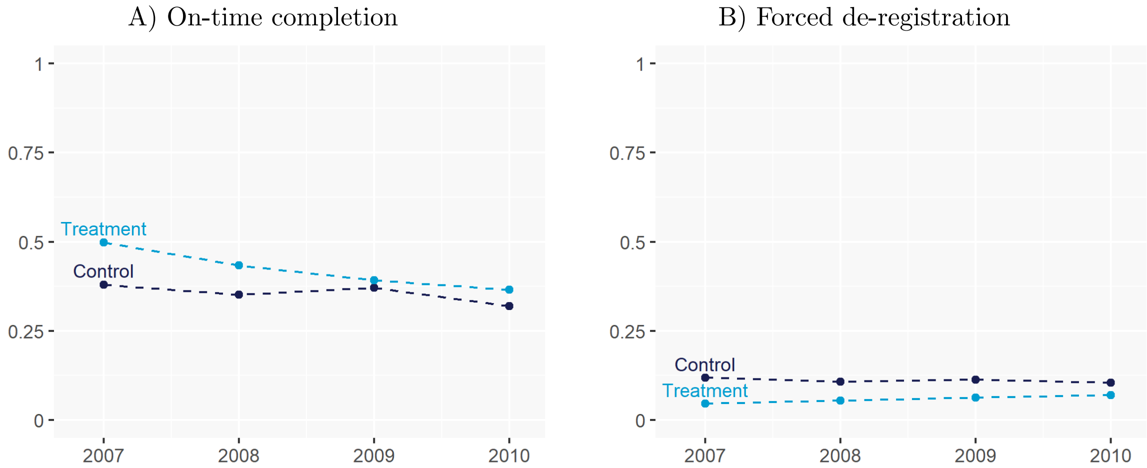
Figure 2: Mean cohort sizes, headcounts and financial budgets at colleges



All mean values are average values over all colleges in the sample. Panel A provides mean values for entry cohorts of all 278 programs and 58 subjects over time. The drop in 2013 is due to data availability and the sample selection procedures: Students are only included if they completed the program within four years. The drop indicates that students either needed more time, changed the program or voluntarily dropped-out. Panel B illustrates mean headcounts. Panel C illustrates the average expenditure for personnel and for buildings. Own illustration based on the sample derived from RDC (2019a, 2019b, 2019c) as described in section 3.1. More details on the cohort sizes are in Table 14 in Appendix F.

A.2 Parallel trends in pre-reform periods

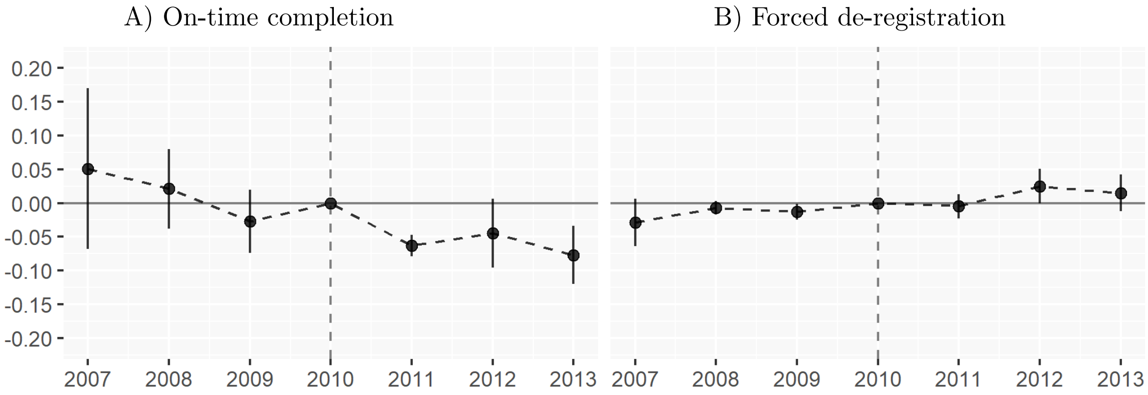
Figure 3: Parallel trends in on-time completion and failing within three years



The figure illustrates pre-trends in the outcome variables *on-time completion* (panel A) and *failing within three years* (panel B). On-time completion is the share of all students per year and state that graduate within three years after enrollment. Forced de-registration is the share of students per year and state that have to leave the college because of failing a compulsory exam several times. States in the treatment group are those that had a double cohort of high school graduates between 2011 and 2013. States in the control group are those without a schooling reform between 2007 and 2015. The illustration is based on the sample derived from RDC (2019c) as described in section 3.1. All values including year- and group-specific sample sizes are provided in Tables 7 and 8 in Appendix C.2.

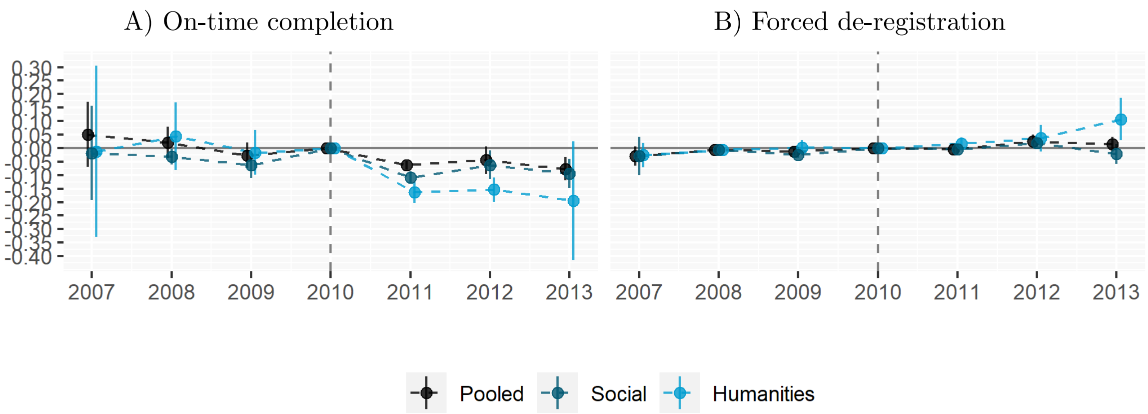
A.3 Regression results

Figure 4: Main results: On-time completion and forced de-registration



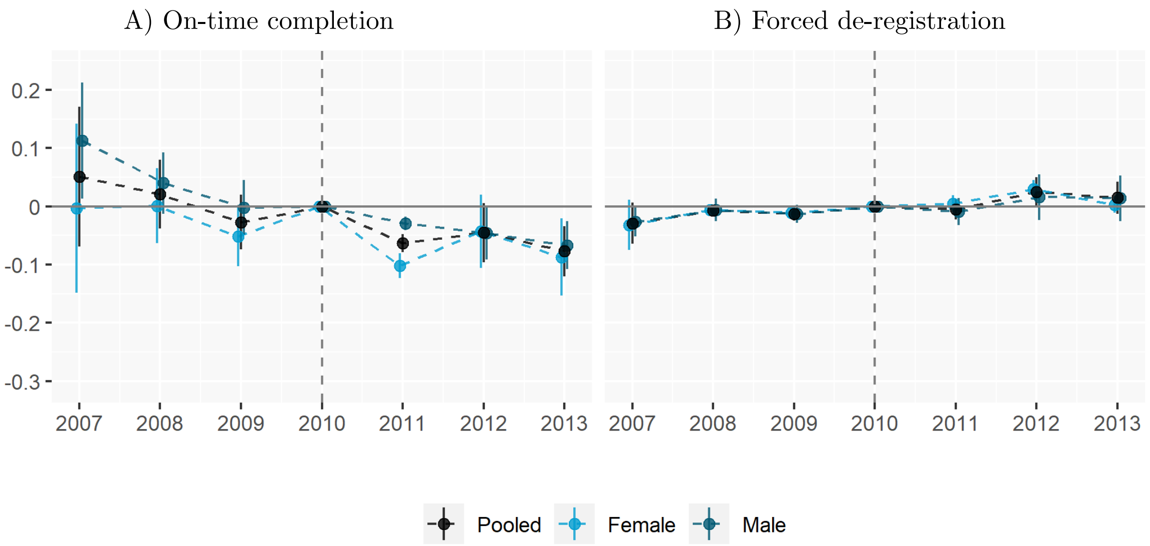
The figure visualizes the coefficients and 95-% confidence intervals of interaction between year and the treatment status, i.e. β_t from Equation 1. All regressions include fixed effects for years, states, and fields as well as a dummy for gender and a constant. The effects for 2013 may partly be driven by the sample selection procedure. Standard errors are clustered on the level of years and states. Complete regression results are available in Table 3.

Figure 5: Heterogeneity across fields: On-time completion and forced de-registration



The figure visualizes the coefficients and 95-% confidence intervals of interaction between year and the treatment status, i.e. β_t from Equation 1. The pooled regression contains all students of the sample, i.e. all undergraduate students in Engineering, Natural Sciences and Maths, Social Sciences, and Humanities. The regressions for Humanities and Social Sciences are based on subsamples containing only students of the respective fields. All regressions include fixed effects for years and states as well as a dummy for gender and a constant. The effects for 2013 may partly be driven by the sample selection procedure. The pooled baseline regression also includes fixed effects for fields. Standard errors are clustered on the level of years and states. Complete regression results are available in Table 4.

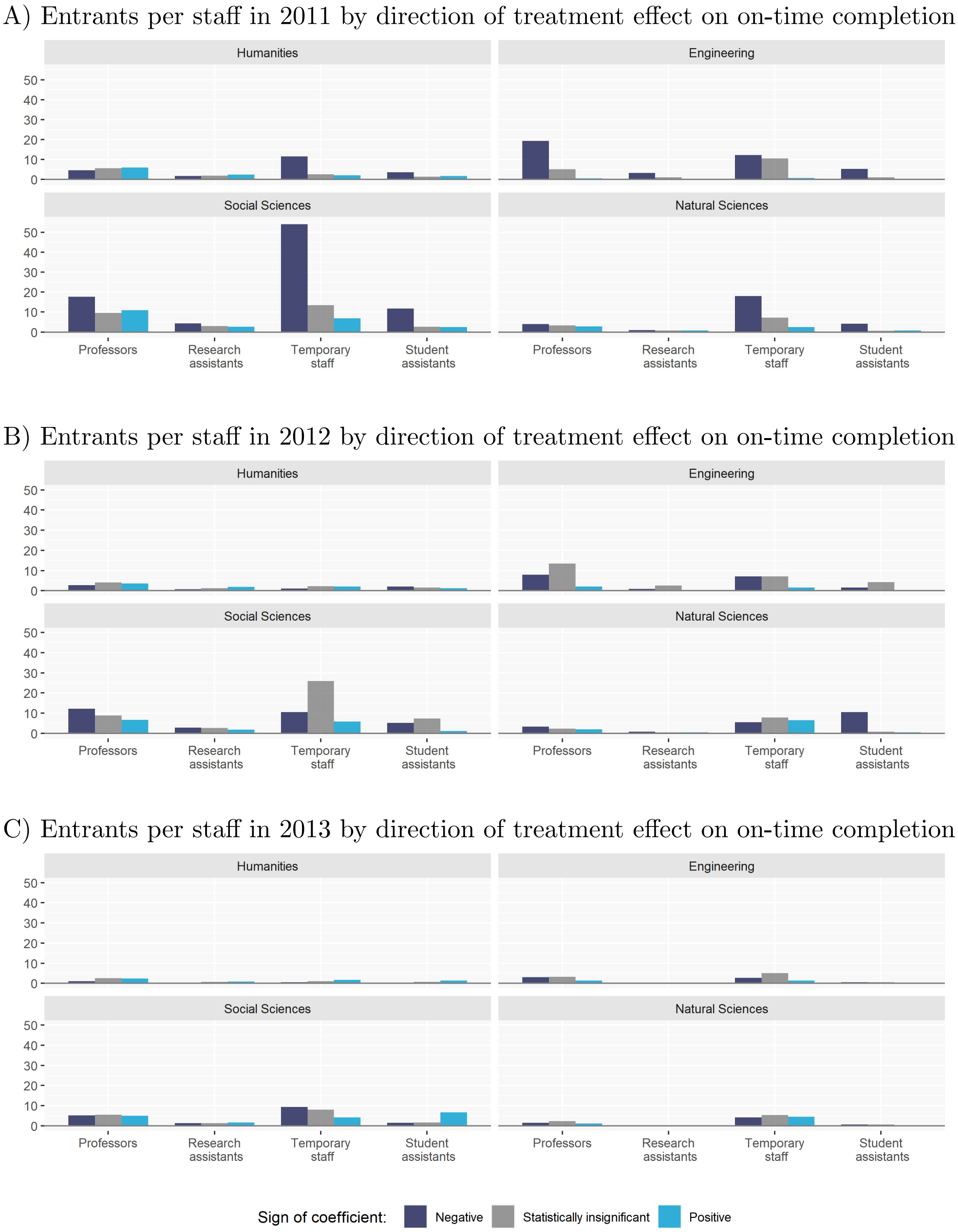
Figure 6: Heterogeneity across gender: On-time completion and forced de-registration



The figure visualizes the coefficients and 95-% confidence intervals of interaction between year and the treatment status, i.e. β_t from equation 1 separately for male and female students. All regressions include fixed effects for years, states, and field of study as well as a constant. Standard errors are clustered on the level of years and states. Complete regression results are available in Table 5 in Appendix B.

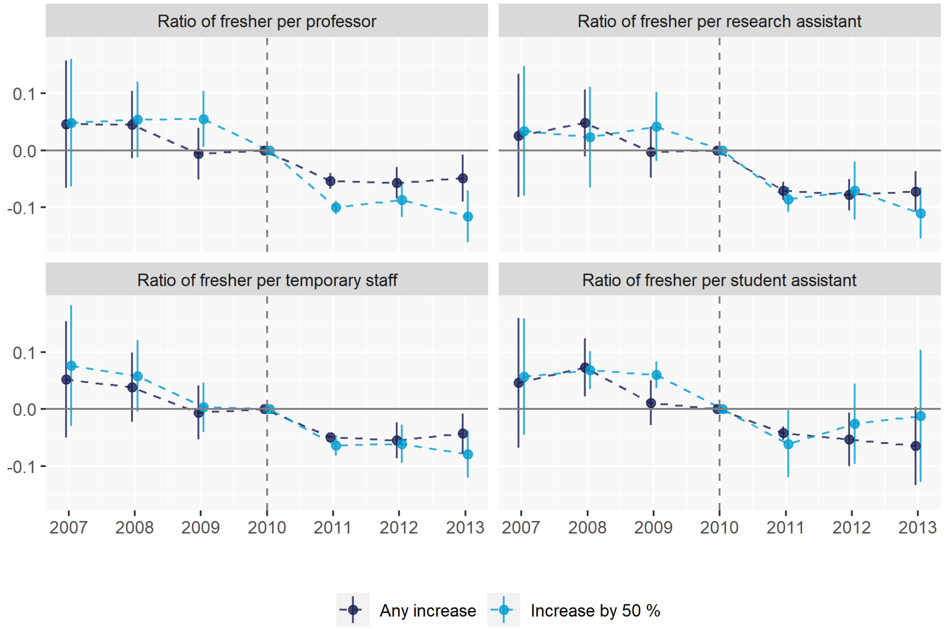
A.4 Channels

Figure 7: On-time completion and field-specific entrants per staff



Note: All panels illustrate field-specific number of entrants per staff by direction of the coefficient of interest for the years 2011 (Panel A), 2012 (Panel B), and 2013 (Panel C). More specifically, I run the baseline regression from Equation 1 seperately for all colleges in the treated states. I then provide the mean number of entrants per staff for these colleges depending whether the coefficients of interest β_{2011} (Panel A), β_{2012} (Panel B), and β_{2013} (Panel C) are statistically significant and positive (light blue – higher on-time completion rate), statistically significant and negative (dark blue – lower on-time completion rate), or statistically insignificant (gray). Panel (C) may be biased due to data availability and sample selection (see also section 3.1 for a discussion of the entry cohort of 2013).

Figure 8: On-time completion for different ratios in entrants per staff



Note: The figure visualizes the coefficients and 95-% confidence intervals of interaction between year and the treatment status, i.e., β_t from Equation 1, for two subsamples: The first subsample contains all colleges that experienced an above average increase in the ratio of freshers to staff in 2011 compared to previous years (dark blue). The second subsample contain only those colleges that experienced an above average increase of at least 50 % in 2011. All regressions include fixed effects for years, states, and fields as well as a dummy for gender and a constant. The effects for 2013 may partly be driven by the sample selection procedure. Standard errors are clustered on the level of years and states. Complete regression results are available in Tables 22, 23, 24, and 25 in Appendix F.

B Tables

B.1 Data and Descriptives

Table 1: Summary of available administrative datasets

Dataset	Type	Period	Smallest Unit	No. of obs.
Exam Registry	Cross-section	2003 - 2017	Individual student	5,506,374
Financial Statistics	Panel	2006 - 2013	Narrow field	48,687
Registry of Personnel				7,304,700
Academic	Cross-section	2006 - 2016	Individual staff	4,392,126
Administrative	Cross-section	2006 - 2016	Individual staff	2,912,574

Own calculations based on RDC (2019a, 2019b, 2019c).

Table 2: Sample descriptives

			Transition to higher education				College outcomes		
Entry cohort	No. of obs.	Female students	Winter term	Generic UEQ	Gap year	Voc. train.	On-time compl.	Forced dropout	Mean GPA
2007	15,944	0.48	1.00	0.98	0.46	0.04	0.47	0.08	2.16
2008	20,575	0.45	0.96	0.98	0.47	0.03	0.42	0.08	2.16
2009	23,342	0.46	0.97	0.98	0.47	0.04	0.39	0.09	2.16
2010	24,521	0.47	0.97	0.97	0.46	0.04	0.36	0.09	2.16
2011	30,664	0.42	0.91	0.98	0.34	0.03	0.33	0.11	2.17
2012	22,230	0.47	0.96	0.97	0.31	0.03	0.33	0.13	2.15
2013	14,968	0.47	0.96	0.97	0.35	0.03	0.36	0.15	2.10

Own calculations based on RDC (2019a, 2019b, 2019c). The mean GPA does not include forced de-registrations. State-specific descriptive statistics are Appendix C.2.

B.2 Major regression results

Table 3: On-time completion, forced de-registration, and study duration

	Dependent variable:		
	On-time completion (1)	*Forced de-registration* (2)	*Study duration* (3)
Treat x 2007	0.051	−0.029	−0.190
	(0.061)	(0.015)	(0.135)
Treat x 2008	0.021	−0.007	−0.059
	(0.030)	(0.005)	(0.053)
Treat x 2009	−0.027	−0.013**	0.060
	(0.024)	(0.006)	(0.049)
Treat x 2011	−0.063***	−0.005	0.123***
	(0.008)	(0.009)	(0.015)
Treat x 2012	−0.045*	0.025*	0.122**
	(0.026)	(0.013)	(0.059)
Treat x 2013	−0.077***	0.015	0.246***
	(0.022)	(0.014)	(0.043)
Female	0.092***	−0.045***	−0.118***
	(0.015)	(0.011)	(0.031)
Observations	152,244	152,244	136,616
Adjusted R^2	0.077	0.059	0.060

Note: All regressions include fixed effects for years, states, and fields. Standard errors are clustered on the level of states and years. The regression in column (3) is based on students who successfully graduated. Thus, the sample size is smaller than in columns (1) and (2) that also include students who were forced to de-register. Significance levels: *p<0.1; **p<0.05; ***p<0.01.

Table 4: Student performance in Humanities and Social Sciences

	Dependent variable:			
	On-time completion		*Forced de-registration*	
	Humanities (1a)	Social Sciences (1b)	Humanities (2a)	Social Sciences (2b)
Treat x 2007	−0.012	−0.018	−0.026	−0.029
	(0.162)	(0.089)	(0.023)	(0.036)
Treat x 2008	0.043	-0.032^{***}	−0.007	−0.006
	(0.064)	(0.015)	(0.011)	(0.003)
Treat x 2009	−0.016	-0.062^{**}	0.003	-0.025^{***}
	(0.042)	(0.025)	(0.013)	(0.005)
Treat x 2011	-0.164^{***}	-0.109^{***}	0.016^{***}	−0.004
	(0.020)	(0.012)	(0.012)	(0.007)
Treat x 2012	-0.153^{***}	-0.062^{***}	0.037^{***}	0.018^{***}
	(0.023)	(0.027)	(0.025)	(0.007)
Treat x 2013	-0.195^{*}	-0.094^{***}	0.107^{***}	−0.022
	(0.112)	(0.028)	(0.040)	(0.019)
Observations	9,729	69,830	9,729	69,830
Adjusted R^2	0.087	0.033	0.078	0.039

Note: All regressions include fixed effects for years and states and a coefficient for gender. Standard errors are clustered on the level of states and years. Significance levels: *p<0.1; **p<0.05; ***p<0.01.

Table 5: Heterogeneity across gender

	Dependent variable:			
	On-time completion		*Forced de-registration*	
	Male	Female	Male	Female
	(1a)	(1b)	(2a)	(2b)
Treat x 2007	0.113**	−0.003	−0.026**	−0.032
	(0.051)	(0.074)	(0.013)	(0.022)
Treat x 2008	0.040	0.001	−0.006	−0.006
	(0.027)	(0.033)	(0.010)	(0.004)
Treat x 2009	−0.002	−0.052**	−0.013	−0.011**
	(0.024)	(0.026)	(0.008)	(0.005)
Treat x 2011	−0.029***	−0.102***	−0.009	0.005
	(0.006)	(0.011)	(0.012)	(0.007)
Treat x 2012	−0.046**	−0.043	0.016	0.030**
	(0.023)	(0.032)	(0.020)	(0.008)
Treat x 2013	−0.067***	−0.087***	0.014	0.003
	(0.021)	(0.034)	(0.020)	(0.007)
Observations	82,883	69,361	82,883	69,361
Adjusted R^2	0.071	0.047	0.058	0.034

Note: All regressions include fixed effects for years and states. Standard errors are clustered on the level of states and years. Significance levels: *p<0.1; **p<0.05; ***p<0.01.

C Data Appendix

C.1 Description of the data

The German Federal Statistical Office provides a variety of administrative higher education data. Access is granted upon application and payment of a fee. Data can only be analysed in a Research Data Centre and all results are subject to statistical disclosure control.
To date, German administrative higher education data are mostly used by the statistical offices to provide descriptive statistics about higher education in Germany. In a similar manner, two German publications base their descriptive evidence on the Exam Registry (Brändle & Lengfeld, 2015; Grözinger, 2017). Student-level data from the Student Registry have been used by Bietenbeck et al. (2020), Marcus and Zambre (2019), Horstschräer and Sprietsma (2015) and Görlitz and Gravert (2018).
This paper is based on the Exam Registry (*Prüfungssstatistik*), the Registry of Personnel (*Personal- und Stellenstatistik*), and the Financial Statistics of Higher Education Institutions (*Hochschulfinanzstatistik*). It is the first paper that combines all three datasets. Details on the years and number of observations for each dataset are available in Table 1. In the following, I describe each dataset in more detail.

Exam Registry

The Exam Registry is a retrospective cross-section. It comprise student characteristics of all students who de-registered in a term. The individual data are obtained from the administrative data of all German universities and are collected twice a year (at the end of each summer and winter term). The Exam Registry is thereby a secondary statistic. Data are currently available from 1995 onwards.
The dataset include information on each student's demographic background (gender, date of birth, nationality), the university entrance qualification (year, district), the study career (field of study, information on first enrollment, desired final exam) and the outcomes of the studies (college, field of study, final GPA, reason for de-registration). Complete individual-level data is available for roughly 90 % of the students in the Exam Registry. Missing observations result from software malfunctioning during the data preparation process at the Federal Statistical Office of Germany. These missing observations are random.
It is worth emphasizing that the dataset does not include all students who are enrolled in a given term but only those who complete their degree or fail during that term. Because information on first enrollment is available, I can construct a dataset that include all college entrants in a given year and college. Because the Exam Registry does not include any voluntary dropouts or ghost students, it can be argued that all students that appear in the Exam Registry were willing to complete their degree.
A note on *ghost students* in Germany: Some institutions estimate that more than 20 % of their enrolled students are ghost students, in some subjects the share is even higher (Jannasch & Olbrisch, 2014; Landtag Nordrhein-Westfalen, 2019). A considerable number of high school graduates enroll to college without being interested in completing a degree because they experience financial benefits: Registered students get highly discounted tickets for public transport, cheap health care and tax discounts when working in student jobs.
A dataset that includes *all registered* students per term – i.e. committed students, ghost students, and voluntary drop-outs – is the Student Registry (*Studentenstatistik*). The difference in the number of observations between the Exam Registry and the Student Registry is large due to the large share of ghost students (see above) and the number of drop-outs that vary between 8 % in medicine and 33 % in cultural disciplines. Yet, it does not provide detailed information about graduation.

Registry of Personnel

The Registry of Personnel is an administrative dataset comprising individual data on all academic and non-academic staff at German universities as well as occupied and vacant positions. Data are currently available from 1998 onwards. The individual level data are obtained from the administrative data of the universities and are therefore secondary statistics.
Each individual observation relates to a person that is employed in higher education. The data contain information on the demographic background as well as the employment relationship (career group, salary or remuneration group, type of financing) and on the professional and organizational affiliation of the employees.
For the purpose of this study, I aggregated the Registry and counted the number of different types of employees per field and institution.

Financial Statistics

The Financial Statistics provide information on the teaching and research structure of the higher education institutions and, in particular, show the differences in funding between the teaching and research areas. The data contain information on the different types of income such as public and third-party funding as well as different classifications of expenditure (personnel, buildings, investments). Data has been available until the reporting year of 2013. Due to lack of scientific usage, the dataset has not been provided for the years after 2013.

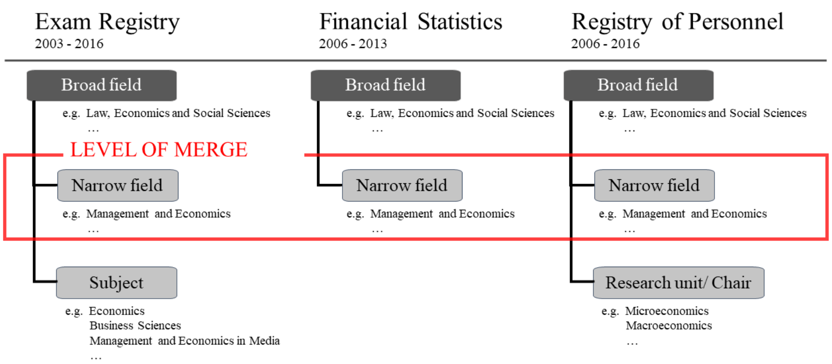
List of fields

The data distinguish 278 different programs that are organized in 59 subjects under the seven fields Humanities, Law, Education and Social Sciences, Arts, Sports, Medicine, Agriculture and Veterinary Medicine, Engineering, and Mathematics and Natural Sciences. Table 6 lists all these subjects and fields. A complete list of all 278 programs is available on request.

Data Linkage

Figure 9 illustrates the level of the linkage between the Exam Registry, the Financial Statistics and the Registry of Personnel.

Figure 9: Merge of Exam Registry, Financial Statistics and Registry of Personnel



Source: Own illustration.

Table 6: List of subjects and fields

Humanities	**Sports**
Linguistics and cultural studies	Sport, sports science
Evangelical theology	**Medicine**
Catholic theology	Health sciences
Philosophy	Human medicine
History	Dentistry
Library Science	**Agriculture and Veterinary Medicine**
Literature and linguistics	Veterinary Medicine
Classical Philology	Land maintenance, environmental design
German studies	Agriculture, food and beverage technology
English and American Studies	Forestry and timber industry
Romanic languages and literature	Nutrition and household sciences
Slavic, Baltic and Finno-Ugrian	**Engineering**
Non-European language and cultural studies	General engineering
Cultural Studies	Industrial engineering
Psychology	Mining and metallurgy
Education	Process Engineering
Special education	Electrical engineering
Law, Economics and Social Sciences	Traffic engineering incl. nautical
Economics and Management	Architecture, interior design
Regional Studies	Spatial planning
Political Science	Civil Engineering
Social sciences	Surveying and mapping
Social work	**Mathematics and Natural Sciences**
Law	General mathematics and Sciences
Administrative science	Mathematics
Management for engineering	Computer Science
Arts	Physics, astronomy
Art, Art Science in General	Chemistry
Music, Musicology	Pharmacy
Fine arts	Biology
Design	Earth Sciences
Performing Arts, Film, Theater Studies	Geography

Note: Own summary based on RDC (2019c). A complete list of all 278 programs is available on request.

C.2 Additional descriptive statistics of the sample

The Exam Registry includes all students who de-registered, i.e., all students who successfully graduated from college or who had to de-register because they failed their studies. The following tables include the descriptive statistics as in Table 2 but seperately for the subsamples of graduates (Table 7) and those who failed their studies (Table 8).

Table 7: Sample descriptives: Students who completed their undergraduate degree

Entry cohort	No. of obs.	Female students	Transition to higher education: Winter term	Generic UEQ	Gap year	Voc. train.	College outcomes: On-time completion	Years at college
2007	14,732	0.50	1.00	0.98	0.46	0.04	0.51	3.31
2008	18,944	0.46	0.96	0.98	0.48	0.04	0.45	3.37
2009	21,296	0.48	0.97	0.98	0.48	0.04	0.42	3.40
2010	22,295	0.49	0.97	0.97	0.46	0.04	0.39	3.43
2011	27,388	0.44	0.91	0.98	0.35	0.03	0.36	3.45
2012	19,298	0.51	0.96	0.97	0.32	0.03	0.38	3.46
2013	12,663	0.51	0.96	0.97	0.37	0.03	0.43	3.39

Own calculations based on RDC (2019a, 2019b, 2019c).

Table 8: Sample descriptives: Students who failed their undergraduate degree

Entry cohort	No. of obs.	Female students	Transition to higher education: Winter term	Generic UEQ	Gap year	Voc. train.	College outcomes: Fail within 3 years	Years at college
2007	1,212	0.34	1.00	0.99	0.46	0.03	0.83	2.17
2008	1,631	0.29	0.96	0.98	0.43	0.02	0.83	2.22
2009	2,046	0.27	0.96	0.96	0.43	0.03	0.82	2.23
2010	2,226	0.27	0.96	0.97	0.44	0.03	0.83	2.23
2011	3,276	0.22	0.90	0.96	0.30	0.03	0.85	2.13
2012	2,932	0.24	0.96	0.97	0.24	0.02	0.89	1.94
2013	2,305	0.25	0.96	0.97	0.27	0.02	0.94	1.77

Own calculations based on RDC (2019a, 2019b, 2019c).

C.3 Stylized facts

To understand individual responses to the reforms, a two-way fixed effects model of the form

$$y_i = \beta DC_{ist} + \gamma X_i + \nu_s + \mu_t + \lambda_k + \sum \tau_{ijts} + \epsilon_{ijkts} \tag{3}$$

is estimated with the sample data. The outcome y_i is an indicator that equals 1 if a student i moved to another state for taking up higher education (column 1), if a student delays university entry by more than a year (column 2) or if a student switched to another institution during the course of study (column 3). Completing military service is not considered as a delay for men prior to 2011. The variable DC_{ist} is an indicator that equals 1 if student i graduated as part of a larger cohort in state s in year t. X_i includes a student's gender and ν_s, μ_t, and λ_k are fixed effects for state, time, and field, respectively. The regression in column (1) also includes an indicator whether a student received the UEQ in a district that has a border with a neighboring state as well as an interaction term. Column (3) includes an indicator that equals 1 if the student enrolled to college in a state and year that had a larger cohort of high school graduates. These additional co-variates are captured by $\sum \tau_{ijts}$ in equation 3.

Table 9: Individual behavioural changes following larger cohorts at school

	Dependent variable:		
	State migration (1)	*Delay entry* (2)	*Switch institution* (3)
DC at school	-0.025^{**}	-0.079^{***}	0.004
	(0.012)	(0.023)	(0.008)
Border district	0.081^{***}	-	-
	(0.012)		
Border district x DC at school	0.009	-	-
	(0.006)		
DC at college	-	-	-0.019^{***}
			(0.006)
Constant & covariates	Yes	Yes	Yes
Fixed effects	Yes	Yes	Yes
No. of observations	152,244	152,244	152,244
Adjusted. R^2	0.223	0.101	0.028

Note: Columns (1), (2), and (3) provide regression results of two-way fixed effects models that regress student migration (column 1), delay of college entry by more than a year (column 2) and switching to another institution during the course of study (column 3) on graduating from high school as part of a double cohort (*DC at school*). The indicator *DC at college* equals 1 if the student enrolled to college in a state and year that had a larger cohort of high school graduates. *Border* is an indicator that equals 1 if a student received the UEQ in a district that has a border with a neighboring state. All regressions also include a constant, an indicator for gender as well as fixed effects for time, state, and study field. Standard errors are clustered on the level of states and years. Significance levels: *p<0.1; **p<0.05; ***p<0.01.

D Robustness checks

In order to provide supporting evidence for the validity of the results presented in section 5, I conduct a series of robustness checks. This section describes all robustness checks.

D.1 Alternative sampling procedures

D.1.1 Extended study duration of up to 5 years

The main sample has been restricted to all undergraduate students aiming for a Bachelor's degree at public universities who completed their degree within four years after enrolling. The rationale for this restriction is that students have to pay a penalty fee after four years.

In the first robustness check, I follow an alternative sampling procedure: I exclude students only if they needed more than five years to complete their degree (11 semesters or more). Thereby I allow for an additional year until graduation or de-registration. A caveat of this approach lies, however, in the retrospective nature of the Exam Registry: Data is only available for students who de-registered until 2017. Allowing for an additional year of studying requires an exclusion of all students who started their undergraduate course in 2013 or later. The alternative sample is thereby smaller and includes $N = 137,321$ students.

Table 10: Sample descriptives of alternative sample: Up to five years of studying

			Transition to higher education				College outcomes		
Entry cohort	No. of obs.	Female students	Winter term	Generic UEQ	Gap year	Voc. train.	On-time compl.	Forced dropout	Mean GPA
2007	15,876	0.46	1.00	0.98	0.46	0.03	0.41	0.07	2.41
2008	20,452	0.42	0.96	0.98	0.48	0.03	0.36	0.08	2.43
2009	23,510	0.43	0.97	0.98	0.48	0.04	0.33	0.08	2.45
2010	24,943	0.44	0.97	0.97	0.47	0.04	0.31	0.09	2.45
2011	30,849	0.40	0.90	0.98	0.34	0.03	0.28	0.10	2.51
2012	21,691	0.45	0.95	0.97	0.32	0.03	0.29	0.13	2.55

Own calculations based on RDC (2019a, 2019b, 2019c). The mean GPA does not include forced de-registrations. State-specific descriptive statistics are Appendix C.2.

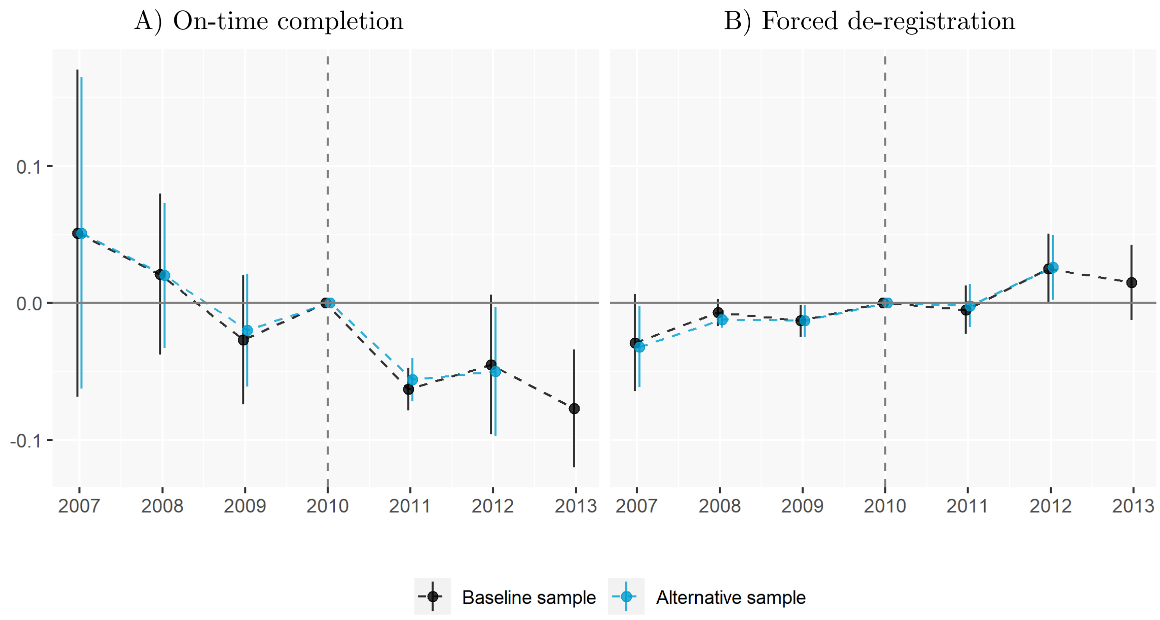
The descriptive statistics of the alternative sample show that the share of female students is smaller than in the baseline sample which indicates that male students are often those who need more than four years to complete a degree. Table 10 reveals no differences in the type of the university entrance qualification, the share of students who completed vocational training prior to studying or the timing when they started their degree across the alternative sampling method. By definition of the sample, the share in students who completed a degree on time is smaller. Interestingely, the mean GPA upon graduation is worse in the alternative sample while there are no changes in the share of forced dropouts. The descriptive statistics suggest that students who need longer to complete their degree perform worse in terms of GPA. Forced de-registrations, however, appear to happen during the first years of studying.

I follow the same dynamic two-way fixed effects regression as described with Equation 1. I keep the same states in the control group but have to exclude one state from the treatment group.[12]The regression

[12] In the state of North Rhine-Westfalia, two cohorts of high school students graduated from high school in 2013. Data is only available until 2017 so that I have to exclude all students who started in the year of 2013. Consequently, I have to drop the state of North Rhine-Westfalia in this sampling procedure due to data availability.

results are presented in Table 16. They are similar in magnitude as the baseline results. The slight differences in effect sizes (visualized in Figure 10) may be driven by the exclusion of the state rather than including those students who needed much longer to complete their degrees.

Figure 10: Alternative sampling procedure: On-time completion and forced de-registration



The figure visualizes the coefficients and 95-% confidence intervals of interaction between year and the treatment status, i.e. β_t from Equation 1 for the baseline sample and the alternative sample. All regressions include fixed effects for years, states, and fields as well as a dummy for gender and a constant. Standard errors are clustered on the level of years and states. Regression results for the alternative sample are in Table 16.

D.1.2 Exclusion of the cohort of 2013

In a second alternative sampling strategy, I only remove the cohort of 2013 due to the data issues adressed before. The descriptive statistics can be found in table 17.

The results in Table 2) show that the on-time completion rate significantly drops in the aftermath of the larger cohorts – as in the baseline results. More specifically, Table 11 provides that the larger cohorts of 2011 led to a drop in the on-time completion rate for students who started their undergraduate degree in 2011 by 7 percentage points (compared to 6.3 percentage points in the baseline results) and 4.8 percentage points for the cohort of 2012 (compared to 4.5 percentage points in the baseline results). The statistical significance does not change between the baseline sample and the restricted sample that exclude the cohort of 2013.

Table 11: Excluding the cohort of 2013 from the baseline regression

	Dependent variable:	
	On-time completion	*Forced de-registration*
	(1)	(2)
Treat x 2007	0.049	-0.032^{*}
	(0.064)	(0.019)
Treat x 2008	0.024	-0.008^{**}
	(0.031)	(0.004)
Treat x 2009	-0.027	-0.014^{*}
	(0.024)	(0.007)
Treat x 2011	-0.070^{***}	-0.003
	(0.008)	(0.009)
Treat x 2012	-0.048^{*}	0.034^{***}
	(0.028)	(0.013)
Observations	119,782	119,782
Adjusted R^2	0.080	0.055

Note: All regressions include fixed effects for years, states, and fields and a coefficient for gender. Standard errors are clustered on the level of states and years. Significance levels: *p<0.1; **p<0.05; ***p<0.01.

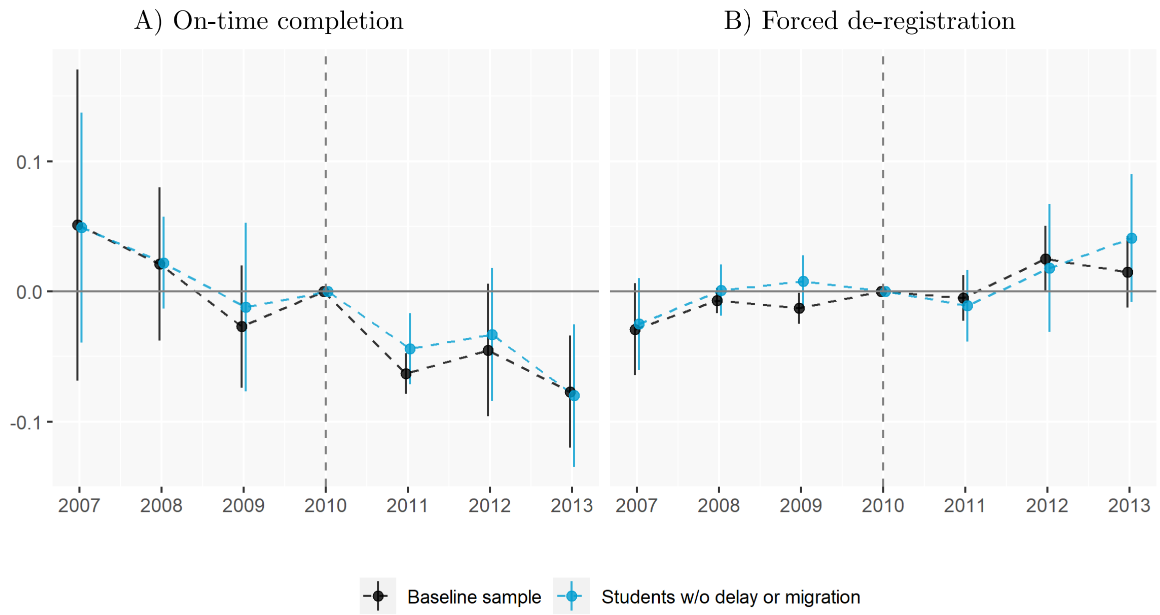
D.2 Students' selection out of larger cohorts

In a second robustness check, I account for students' selection out of larger cohorts. Of course, individuals and institutions can respond to reforms. On the institutional level, it appears that colleges have not been, on average, more selective than in the years before the demand shock (see the kink in panel A in Figure 2). However, as suggested by the results in section C.3, high school graduates appear to have slightly changed their enrollment behavior in the year of the larger cohorts induced by the schooling reforms.

To understand whether this behavior affects the baseline results, I run the dymamic two-way fixed effects regression on a subsample that only include those students that neither delayed their college education nor migrated to another state to enroll in higher education.

The results in Figure 11 show that the results still point in the same direction as in the baseline regression: Larger cohorts negatively affect the on-time completion rate of students. Yet, the results are slightly smaller in magnitude: In the reduced sample, larger cohorts lead to a reduction in the on-time completion rate by 4.4 percentage points compared to a reduction of 6.3 percentage points on the full sample. The results for forced de-registrations remain statistically insignificant.

Figure 11: Subsample of non-moving and non-delaying students



The figure visualizes the coefficients and 95-% confidence intervals of interaction between year and the treatment status, i.e. β_t from Equation 1 for the baseline sample and the sample that includes only those students who enrolled at a college in their home state in the year of graduating from high school. All regressions include fixed effects for years, states, and fields as well as a dummy for gender and a constant. Standard errors are clustered on the level of years and states. Detailed results for the subsample regression are in Table 17.

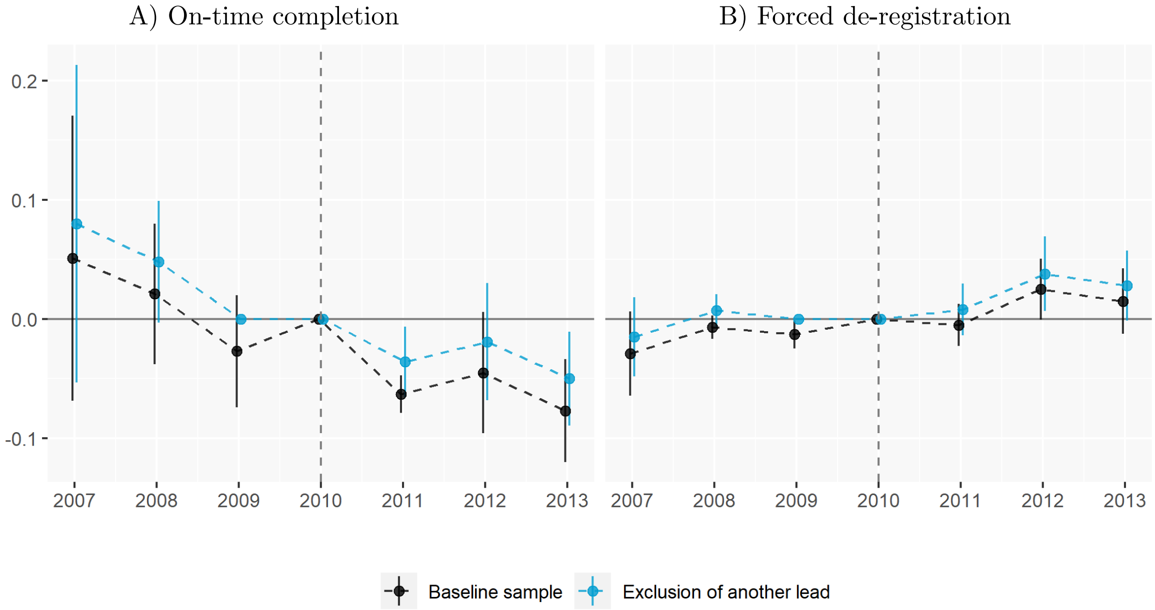
D.3 Spillovers on incumbent cohorts

Students of the pre-DC cohorts could have been affected by the larger cohorts that started their college education in 2011 and beyond. As a result, the baseline results would underestimate the true effect. While it is not possible to determine which incumbent cohorts were affected by the spike in freshmen in 2011, I run a robustness check and show that normalizing the coefficients to two years (2009 and 2010 instead of 2010 as in the baseline regression in Equation 1) leads to a mechanical shift in coefficients.
I run the dynamic two-way fixed effects regression

$$y_{ijkst} = \sum_{t=2007}^{t=2008} \beta_t DC_s + \sum_{t=2011}^{t=2013} \beta_t DC_s + \mu_t + \gamma X_i + \nu_s + \lambda_k + \kappa_j + \epsilon_{ijkts} \tag{4}$$

on the baseline sample where the outcome y_{ijkst} is again a measure of performance of student i. Compared to the baseline regression in Equation 1, I include only two leads but keep three lags.
The results in Figure 12 show that the exclusion of two pre-2011 cohorts leads to a mechanic upward shift in coefficients. If the cohorts of 2007 and 2008 were the last unaffected cohorts, a normalization of the coefficients to the level of 2008 would then shift the curves downwards, leading to effects larger in magnitude for the pre-reform years. The effects on forced de-registration remain the same as in the baseline specification. That means that, if temporal spillovers on pre-2011 cohorts existed, the baseline results would in fact underestimate the true effects.

Figure 12: Spillovers on pre-reform cohorts



The figure visualizes the coefficients and 95-% confidence intervals of interaction between year and the treatment status, i.e. β_t from Equation 1 for the baseline regression and a regression that excludes another lead from the regression. All regressions include fixed effects for years, states, and fields as well as a dummy for gender and a constant. Standard errors are clustered on the level of years and states. Complete regression results are available in Table 18.

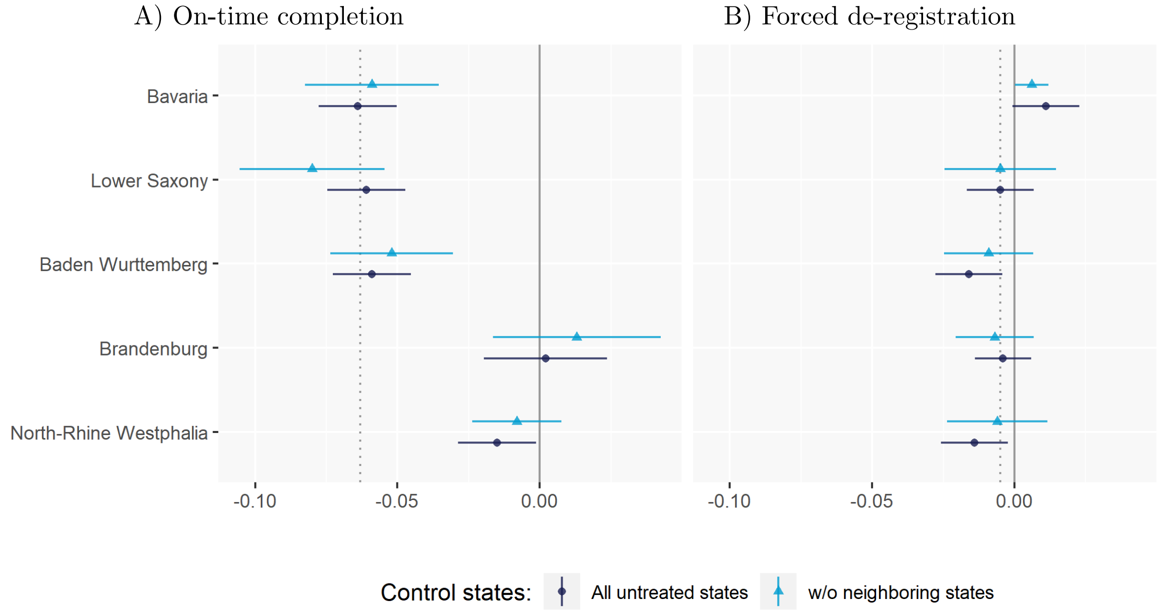
D.4 Spatial spillovers

Student mobility is low in Germany. If students move to another state, they usually start college in a neighboring state (see Figure 20 in Appendix E.1). Still, the descriptive evidence in section 3.1 suggests that students of the larger cohorts of high school graduates were more likely to start college outside their home state, in particular if they completed high school in a border district.

To understand whether this behavior drives the results, I run the event study from Equation 1 separately for each treated state with and without neighboring states in the control group. Figure 13 provides the point estimates and 95 %-confidence intervals for all β_{2011} of these regressions. The complete regression results for on-time completion and forced de-registration are in Tables 19 and 20 in Appendix F.

Figure 13 shows that there are only slight differences in the point estimates if neighboring states are excluded from the control group. The results thereby suggest that spatial spillovers do not drive the results.

Figure 13: Control groups with and without neighboring states



The figure visualizes the point estimates and 95-% confidence intervals of interaction between year and the treatment status for the year 2011 only, i.e. β_{2011} from Equation 1, separately for each treated state with and without neighboring states in the control group. The point estimate from the baseline regression pooled over all states is indicated by the dotted line. The states of Berlin and Bremen are excluded because these states share common borders only with treated states. Detailed regression results are in Tables 19 (on-time completion) and 20 (forced de-registration) in Appendix F.

D.5 Placebo checks

In another robustness check, I run a placebo check for parallel trends. More specifically, I assign a placebo treatment period to 2010 and run an adapted event study for the main outcomes, i.e. on-time completion and forced de-registration:

$$y_{ijkst} = \sum_{t=2007}^{t=2008} \beta_t DC_s + \sum_{t=2010}^{t=2010} \beta_t DC_s + \mu_t + \gamma X_i + \nu_s + \lambda_k + \kappa_j + \epsilon_{ijkts} \tag{5}$$

Figure 14 visualizes the coefficients for the two main outcome variables. The small and insignificant effect for the placebo treatment indicates that the results are not driven by any pre-trends.

Figure 14: Regression results with a placebo treatment in 2010

The figure visualizes the coefficients and 95-% confidence intervals of interaction between year and the treatment status, i.e. β_t from Equation 5. Instead of the real treatment in 2011, I assume a placebo treatment period for 2010. All regressions include fixed effects for years, states, and fields as well as a dummy for gender and a constant. Standard errors are clustered on the level of years and states. The complete regression table is in Table 21 in Appendix F.

D.6 (Instrumented) difference-in-differences estimation

The last set of robustness checks uses alternative specifications to address the relationship of interest. Instead of following a dynamic two-way fixed effects approach to estimate the effect of larger cohorts on student performance, I first use a standard difference-in-differences approach. I estimate

$$y_{ijkst} = \beta Treat_s + \mu_t + \gamma X_i + \nu_s + \lambda_k + \kappa_j + \epsilon_{ijkts} \tag{6}$$

where $Treat_s$ equals 1 if a student enrolled at a college that is located in a treated state in 2011 or later. Treated states are, again, those that had a schooling reform on top of the nationwide suspense of the military service. Student performance y_{ijkst} is again measured as the on-time completion rate and the rate of forced de-registrations.

Table 12 provides the regression results. Those students who started their college education in a reform state in or after the year of the larger cohorts, had a lower probability to graduate on-time. The average effect of a reduction by 6.4 percentage points is similar in magnitude as in the event study specification. The result thereby suggests that the reduction in on-time completion was persistent in post-reform years. While the coefficient for the effect of the larger cohorts on forced de-registrations point into the same direction as the event study results, it remains statistically not signifcant.

I also run placebo treatments for the difference-in-differences approach. In doing so, I run 1,500 iterations of the same regression as in Equation 6 but randomly add a placebo treatment period ($Placebo - Treat_s$) to the data instead of using the real treatment period ($Treat_s$). The distribution of the 1,500 estimated coefficients of $Placebo - Treat_s$, each for the two outcomes of interest, is plotted in Figure 15. The difference-in-differences approach seems to provide robust results.

In addition, I also instrument the treatment status at the time when enrolling at college by the treatment status around the year of birth. The treatment status at birth is exogenous to the schooling reforms because the political decision-making process about the schooling reforms started in the late 1990s when affected individuals were already born. The first instrument is an indicator that equals 1 if a student should have been treated according to his year of birth even if he had one year of grade retention. The second instrument is an indicator that equals 1 if a student should have been treated according to his

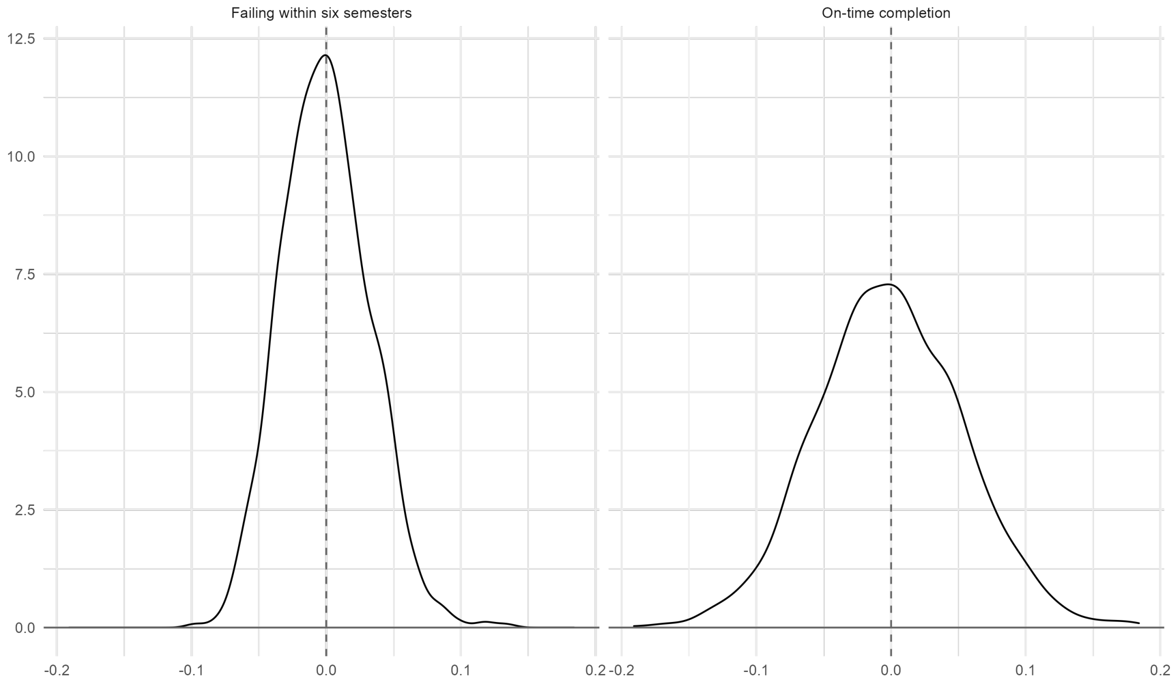
year of birth allowing for up two two years of grade retention.
Although the treatment assignment at birth is theoretically exogeneous, the first-stage regressions show that the instruments are not correlated with the endogenous regressors (see table 12). The weakness of the instrument may be grounded in grade retentions at secondary school or earlier and later entry to primary school. Figure 23 in Section G shows that 10 to 25 % of primary school entrants start school earlier or later than determined by their year and month of birth.
Still, the IV estimates 12 for on-time completion (columns 2a and 2b) point into the same direction as the baseline regressions and the difference-in-differences results (column 1). However, these are not statistically significant. The coefficients for forced de-registration turn the sign and are negative in columns 4a and 4b. It seems that using the instruments adds additional bias coming from earlier or later school entry as well as grade retention at high school.

Table 12: (Instrumented) Difference-in-differences results

	Dependent variables:					
	On-time completion			*Forced de-registration*		
	OLS	IV		OLS	IV	
	(1)	(2a)	(2b)	(3)	(4a)	(4b)
Treat	-0.069^{**}	-0.042	-0.037	0.020	-0.051	-0.099
	(0.028)	(0.097)	(0.045)	(0.014)	(0.156)	(0.129)
Female	0.092^{***}	0.092^{***}	0.092^{***}	-0.045^{***}	-0.045^{***}	-0.045^{***}
	(0.015)	(0.013)	(0.013)	(0.011)	(0.011)	(0.011)
First Stage IV:	-	0.081	0.104^{*}	-	0.081	0.104^{*}
	-	(0.055)	(0.063)	-	(0.055)	(0.063)
Observations	152,244	152,244	152,244	152,244	152,244	152,244
Adjusted R^2	0.077	0.077	0.077	0.059	0.056	0.052

Note: The table displays the coefficients for *Treat* in the standard difference-in-differences setting (column 1 for on-time completion, column 3 for forced de-registration) and in the instrumented difference-in-differences setting (columns 2a and 2b for on-time completion, columns 4a and 4b for forced de-registration). The instruments are the treatment status at birth plus an allowance of one year of grade retention (columns 2a and 4a) and treatment status at birth plus an allowance of two years of grade retention (columns 2b and 4b). All regressions include fixed effects for states, years, and fields. Standard errors are clustered on the level of states and years. Significance levels: *p<0.1; **p<0.05; ***p<0.01.

Figure 15: Distribution of $Treat_s$ coefficient in placebo regressions



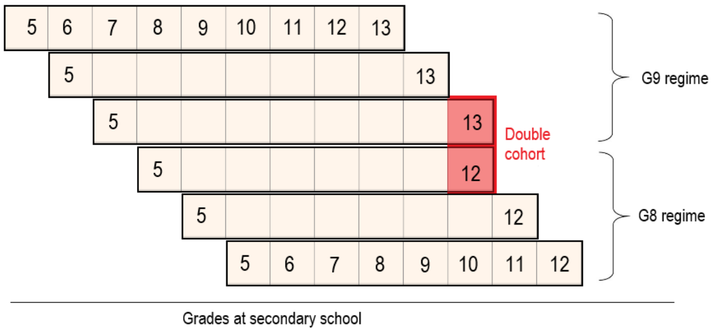
Note: The figure visualizes the distributions of coefficients for the variable $Treat_s$ after running 1,500 placebo regressions based on estimating Equation 6 with a random assignment of the treatment period.

E More on the German institutional setting

E.1 The German G8 reforms

Until 2007 most German high school students had to complete a total of nine years of high school until they could take the final examinations to receive a university entrance qualification (G9 schooling regime). From 2007 onwards, some German states undertook the so-called *G8 reform* that reduced secondary schooling by one year. By design of the reform, there was one year in each affected state in which high schools had two types of graduates: The final cohort of students graduating under the G9 regime and the first cohort graduating under the G8 regime, as illustrated in figure 16. This double cohort of high school graduates was a shock to the number of potential university entrants.

Figure 16: Illustration of double cohort



Note: The figure illustrates the reduction of secondary schooling (grades 5 to 13) by one year and the resulting double cohort.

In Germany, competence in education lies at the level of the federal states (*Länder*), so that each federal state can autonomously undertake educational reforms. The G8 reforms were implemented in different states gradually so that double cohorts of high school graduates were released between 2007 and 2016. The staggered implementation of the reforms is illustrated in figure 17.

Two further reforms took place in the same period: First, between 2006 and 2013 some states allowed their institutions to charge tuition fees up to 500 Euros per semester. Several states introduced tuition fees in 2007. One exception is North Rhine-Westfalia that started charging their students already from 2006 onwards. As illustrated in Figure 24 in Appendix G, each state had at least one neighbouring state that did not charge any tuition fees. Hübner (2012) identifies a small negative effect of the implementation of tuition fees on the extensive margin of studying.

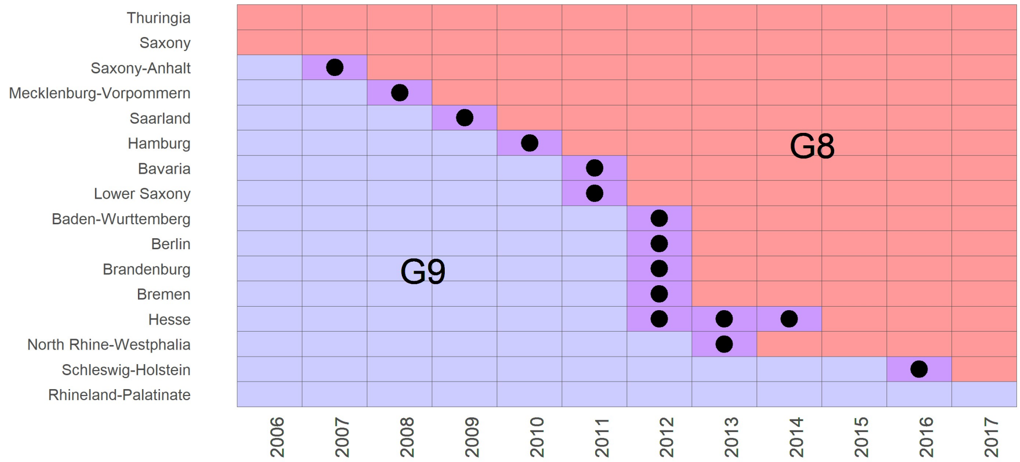
Second, in 2011, the German government decided to put conscription into abeyance. The abolition of military service was unexpected by the population and resulted in a sharp increase of male university entrants in 2011.

Effects on high school students

In order to avoid negative effects on the human capital acquired during secondary schooling, the G8 reforms aimed at reducing the duration of high school without changing the content of the curriculum, i.e. avoiding the usual trade-off between an earlier labor market entry and acquisition of additional human capital. To achieve that aim, the schooling curriculum of nine years was compressed to eight years which resulted in a higher daily workload for high school students.

Previous research on the effects of the G8 reforms on high school students suggests that students of the

Figure 17: Timing of major reforms potentially affecting the number of first-year students

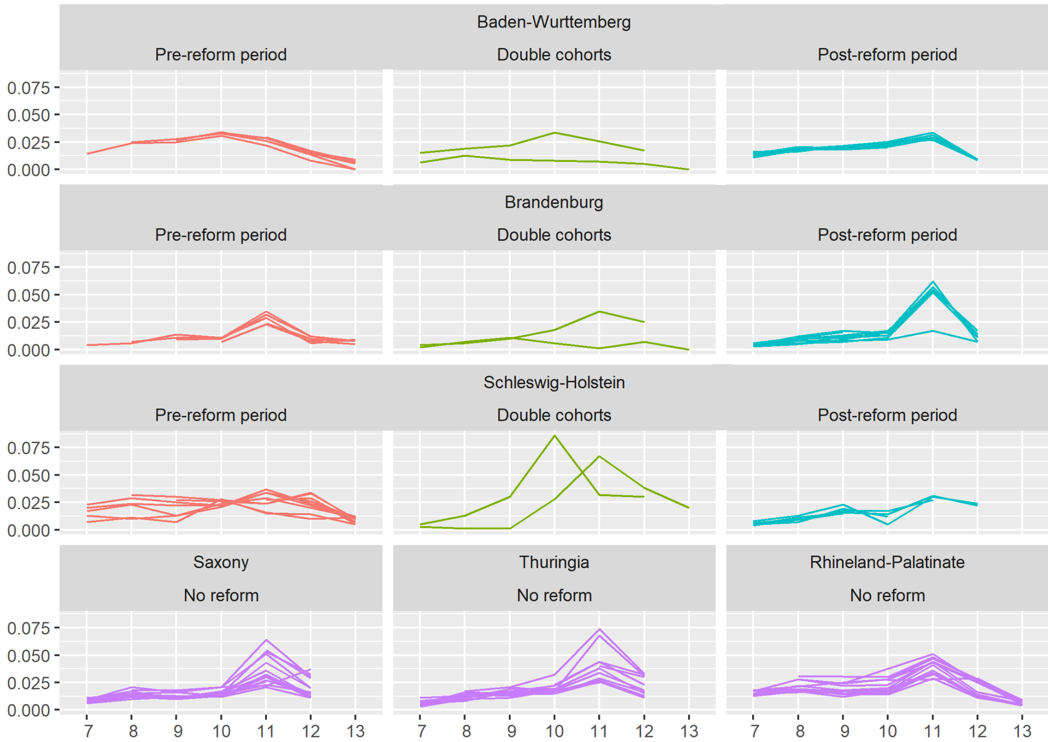


Note: Points indicate the year in which the federal state released two cohorts of high school graduates induced by the change in schooling regime. Hesse released slightly increased cohorts over three years. Rhineland-Palatinate undertook only a policy trial with a few selected schools. Thuringia and Saxony had no reform.

two regimes do not differ in language, but slightly in mathematical skills (Büttner & Thomsen, 2015) . Furthermore, the reform slightly increased grade repetition rates and lowered the final GPAs (Huebener & Marcus, 2017). Compressing instruction time into less years of schooling affected high school students also beyond the performance dimension. A few studies find slight differences in mental health (Quis, 2018), personal traits (Dahmann & Anger, 2014) and drug abuse (Westermaier, 2016) among students of the two schooling regimes. However, these effects are small and mostly based on either survey data or just a single state. Most research to date suggests that the two schooling regimes led to graduates that mainly differ in age but not in competencies or behaviors.

Figure 18 shows the shares of students who repeat a class for three states that had a G8 reform (Baden-Wurttemberg, Brandenburg, Schleswig-Holstein) and three comparison states that either always used the G8 regime (Saxony, Thuringia) or always kept the G9 regime (Rhineland-Palatinate). Detailed data about grade retention is not provided by all states. The graphs show that grade repetition mostly takes place in the years prior to graduation. In general, neither the size nor the pattern change much during the double cohort. The main exception is the state of Schleswig-Holstein in which the maximum share of repeaters increased by roughly 3 percentage points from pre-reform periods to the year of the double cohort and went back to the pre-reform level afterwards. Yet, the state of Schleswig-Holstein is excluded from the baseline sample because of its late adoption of the schooling reform. Figure 18 also shows that the share of grade retention is in general small in all states.

Figure 18: Graduates from secondary school in states with G8 reform



Note: The graphs illustrate the share of grade retention in each cohort of secondary school entrants per secondary school year. Each line depicts one cohort. Thex-axis depicts the grade at secondary school. Graduating takes place at the end of grade 12 under the G8 regime and at the end of grade 13 under the G9 regime. Source: Own illustration based on aggregate data about grade retention in different German states provided by Destatis (2019a).

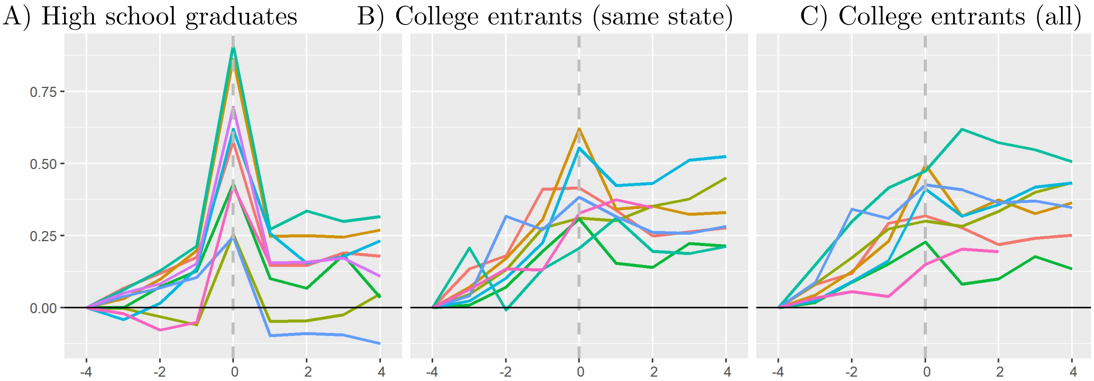
Effects on higher education choices

Four studies analyse the effects of the G8 reforms on high school graduates' higher education choices. Marcus and Zambre (2019) analyse the effect of the compressed schooling curriculum on the probability of going into higher education. They compare the students under the G8 and G9 regimes and conclude that G9 students are more likely not to enroll or to delay their university enrollment. Furthermore, the rate of dropping out of university as well as the change of major are higher among G9 students. These results are in line with findings by Meyer et al. (2019) who focus only on enrollment rates and the probability of spending a gap year between graduating from high school and enrolling to university. Both studies use data from all German states and concentrate on differences between the schooling regimes.

Two earlier studies by Meyer and Thomsen (2016) and Thomsen and Anger (2018) also look at enrollment rates but focus on a single federal state and on the double graduation cohort. They conclude that females were more likely to delay university enrollment and to start vocational training. They could not identify any significant differences between G8 and G9 students in terms of dropout rates or self-reported motivation or abilities.

Based on administrative data, Figure 19 illustrates the relative development in the number of graduates from high school and higher education entrants around the year of the reform. The increased number of high school graduates did not translate into a one-to-one increase in the number of freshmen at colleges.

Figure 19: High school graduates in states with G8 reform



Note: All changes in number of high school graduates are relative to four years prior to the release of the increased cohort. Year 0 indicates the year of the increased cohort. The illustration is based on aggregate data about high school graduates and college entrants provided by Destatis (2019b, 2019c).

Most higher education institutions in Germany do not have a highly selective admission process. Students often enroll to their most preferred institution: A representative survey of more than 15,000 undergraduate students who started their studies in Germany in 2010 has shown that 80 % of the students started studying their most preferred field and 85 % of the students started studying at their most preferred institution in 2010 (Leibniz Institute for Educational Trajectories (LIfBi), 2018). Yet, institutions could have easily reacted to the larger cohorts by restricting the number of admitted students. To the best of my knowledge, there is no evidence how institutions reacted to the increased number of freshmen.

Student mobility and delayed entry to college

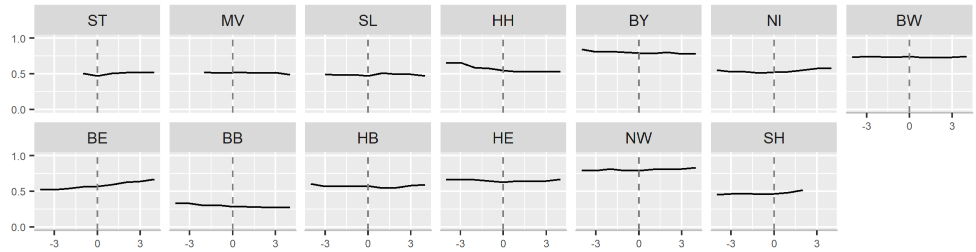
Students have to make the decisions of when, where and what to study. These decisions are commonly referred to as the university choice problem. A student may not enroll to the nearest university if the expected cost for tuition and living are lower than the expected wage premia of attending a particular (selective) institution and to the individual chances of success. Similary, by delaying entry to university, a student may increase his expected return to education by building up additional skills.

Migration to another state is a straightforward way of escaping an expected larger entry cohort in one's home state. All affected states (except Berlin and Bremen) had at least one neighboring state that did not release two cohorts of high school graduates at the same time. Still, aggregate relative migration patterns, as illustrates in figure 20, suggest that students did not behave differently in reform- and non-reform years.

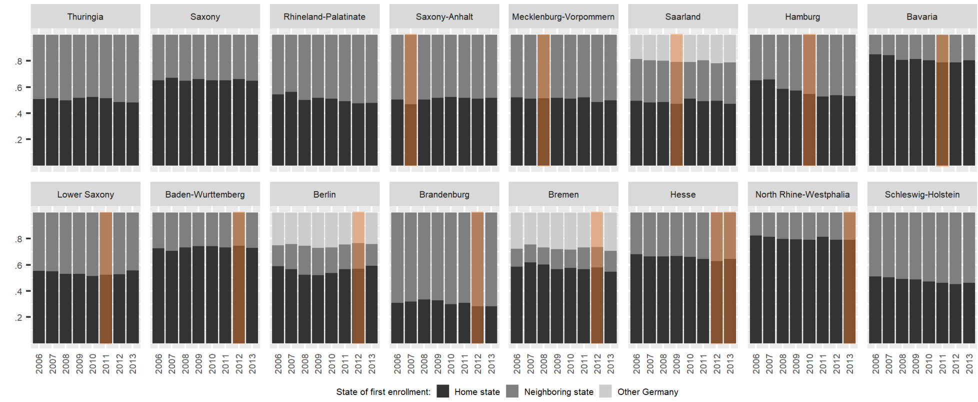
I already described in section 4 and more detailed in Appendix C that migration out of a student's home state does not appear to be a behavioral response to the schooling reform unless a student graduated from secondary school in a district that had a common border with a neighboring state. Overall, it appears that student mobility is low in Germany and has not (or only slightly) been affected by the reform.

Figure 20: Student migration around reform years

A) Share of students staying in home state for higher education



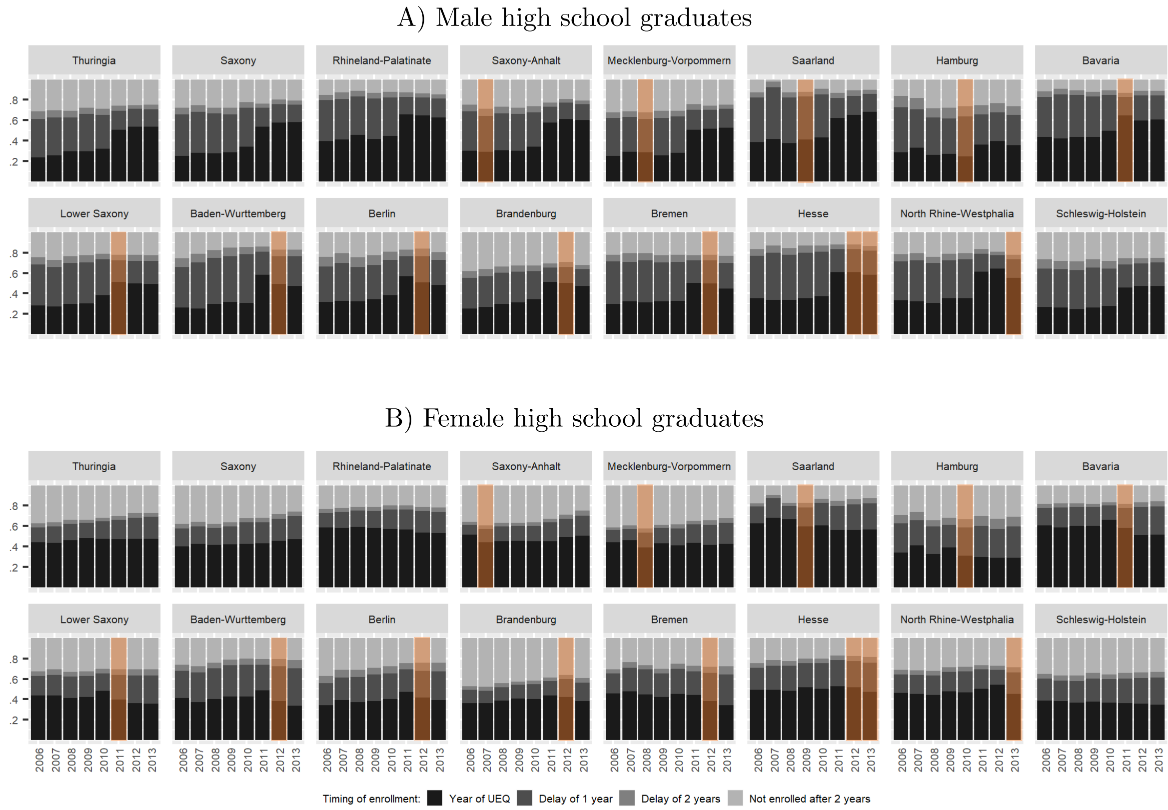
B) Distribution of destinations of migrating students



Note: Panel A illustrates the aggregate share of students who stayed in their home state. The x-axis depicts the years around the reform. Year 0 is the year in which a double cohort of students graduated from secondary school. Panel A includes only states that undertook a G8 reform. Panel B illustrates aggregate relative student migration. The bars depict the share of secondary school students who started university in their home state, in a neighboring state or somewhere else in Germany. The highlighted state-year combinations are those that had a double cohort of graduates from secondary school. Hesse released more high school graduates over three years. The figure is based on aggregate state-level data provided by Destatis (2016, 2015, 2014b, 2014a, 2012b, 2012a, 2011).

The second major behavioral response to the reform is delaying university entry. Until 2011, male students had to delay their college entry by a year in order to complete their military service. Figure 21 provides the state-specific shares of students who enroll to college in the year of high school graduation and who delay their college entry. The majority of high school graduates enrolls to college within one year upon receiving the university entrance qualification.

Figure 21: Direct enrollment and delay of college entry between 2006 and 2013



Note: Panels A and B illustrate the timing of enrollment to higher education for male and female high school graduates. The years on the x-axis refer to the years of high school graduation. The y-axis depicts the share of the cohort of high school graduates that enrolled to college in the year of graduation, a year after, two years after or did not enroll within two years upon high school graduation. The highlighted state-year combinations are those that had a double cohort of graduates from secondary school. Hesse had a different G8 reform that released a slightly increased number of high school graduates over three years. The figure is based on aggregate state-level data provided by Destatis (2016, 2015, 2014b, 2014a, 2012b, 2012a, 2011).

While student mobility is low in Germany, delaying college entry appears to be popular. If systematic changes in the probability to delay college entry leads to changes in the ability distribution of cohorts of college entrants, the identification is challenged because research by Carrell, Fullerton, and West (2009) and Adam S. Booij and Oosterbeek (2017) shows that the ability distribution in college can affect study performance. Unfortunately, the Exam Registry does not provide any information on a student's ability prior to enrolling. To understand, whether students with and without gap years after high school differ in their ability, I use survey data from the German National Education Panel Study.

The German National Educational Panel Study (NEPS) is a sequential multi-cohort panel study that tracks six starting cohorts over their life course. Please see Hans-Peter Blossfeld and von Maurice (2011) for a detailed description. I use data from the starting cohorts SC4 and SC5 that started tracking $N_{SC4} = 16,425$ high school students and $N_{SC5} = 17,910$ college entrants from 2010 through their years of education and their entrance into working life. All caculations are based on version 10.0.0 of NEPS SC4 and on version 12.0.0 of NEPS SC5.[13] I apply the same sample selection procedure as with the Exam

[13]When using the NEPS data, the following statement is required: This paper uses data from the National Educational Panel Study (NEPS): Starting Cohort 4 – Grade 9, `10.5157/NEPS:SC4:12.0.0` and Starting Cohort 5 – First-Year Students, `doi:10.5157/NEPS:SC5:8.0.0`. From 2008 to 2013, NEPS data were collected as part of the Framework Programme for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research (BMBF). As of 2014, the NEPS survey is carried out by the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg in cooperation with a nationwide network.

Registry (see section 3.1). First, I select only first-year undergraduate students at public universities studying towards a Bachelor's degree in a non-selective field. This reduces the sample size in SC4 heavily because only a small fraction of pupils finally enroll to college. Moreover, I exclude all mature students and early entrants to college (*Frühstudium*).

The remaining datasets comprises of $N = 2,931$ students for NEPS SC4 and $N = 5,066$ students for NEPS SC5. Table 13 compares some sample means of the NEPS SC5 and the Exam Registry. Table 13 suggests that students in the NEPS are comparable to those in the Exam Registry.

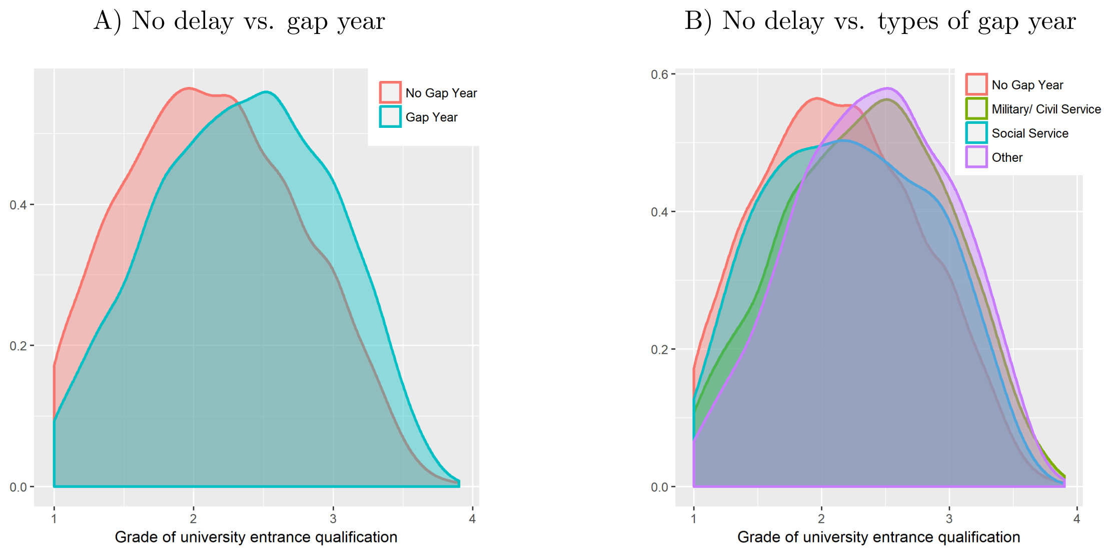
Table 13: Sample means of NEPS and Exam Registry

	Entry cohort of 2010 as of	
	NEPS	**Exam Registry**
Mean age at enrollment	20.48	20.40
Share of:		
Female students	0.46	0.45
German citizenship	0.97	0.93
Gap year	0.46	0.46
Number of observations	5,066	24,521

Note: The sample means refer to the entry cohorts of 2010 in NEPS SC5 and sample of the Exam Registry as described in section 3.1.

Figure 22 illustrates the distribution of high school GPAs among students with and without a gap year in panel A. It shows that students who did not delay college entry had a higher GPA when graduating from high school. However, as of 2010, this was driven mostly by students who started military service (see Panel B in figure 22). Voluntary gap years such as completing a year of voluntary social service are not associated with much lower GPAs. I conclude that the slightly larger share of students with a gap year is unlikely to significantly affect the distribution of ability in a cohort of first-year entrants. If at all, the cohort would be positively selected, so that more talented students start their college degree. In such a case, one would expect an increase (and not a decrease) in study performace.

Figure 22: Distribution of high school GPA across types of gap years



Note: The figure depicts group-specific distributions of first-year students' high school GPA. Panel A compares the distribution of high school GPA of those students who enrolled in the year of receiving their UEQ (*No Gap Year*) and of those who delayed college entry by at least one year (*Gap Year*). Panel B shows the same distributions but splits the group of students who delay college entry up into different groups according to the type of gap year. Own illustration based on NEPS SC5.

F Additional tables

Table 14: Average cohort sizes and college expenditure

Year	Entry cohort		College expenditure	
	Subject	Program	Personnel	Buildings
2007	33.30	52.92	85,658,662	14,107,566
2008	34.30	55.94	87,209,861	14,939,510
2009	35.19	56.61	93,590,306	16,701,196
2010	34.87	56.12	100,163,957	17,464,540
2011	40.11	63.89	104,567,739	18,272,106
2012	34.42	52.50	110,266,060	19,079,938
2013	27.21	38.25	11,5410,780	20,584,911

Own calculations based on the sample generated with RDC (2019a, 2019b, 2019c). Cohort sizes are mean values for each subject or program, respectively. Expenses are mean values for all colleges in the sample.

Table 15: Average number of staff per college

Year	Academic staff			Administration	
	Professors	Graduate RAs	Student RAs	Administrative Staff	Librarians
2007	165.49	820.57	729.29	333.43	75.49
2008	159.24	821.76	746.05	330.29	69.74
2009	165.32	908.68	838.39	340.16	69.97
2010	169.74	940.18	864.61	351.58	73.39
2011	172.18	988.84	915.08	363.50	73.63
2012	172.71	1025.66	912.13	383.21	71.61
2013	178.00	1060.68	943.00	391.43	72.89

Own calculations based on the sample generated with RDC (2019a, 2019b, 2019c). Cohort sizes are mean values for each subject or program, respectively. Expenses are mean values for all colleges in the sample.

Table 16: On-time completion and forced de-registration - larger sample (10 semesters)

	Dependent variable:	
	On-time completion	*Forced de-registration*
	(1)	(2)
Treat x 2007	0.051	-0.032^{**}
	(0.058)	(0.015)
Treat x 2008	0.020	-0.012^{***}
	(0.027)	(0.003)
Treat x 2009	-0.020	-0.013^{*}
	(0.021)	(0.006)
Treat x 2011	-0.056^{***}	-0.002
	(0.008)	(0.008)
Treat x 2012	-0.050^{**}	0.026^{**}
	(0.024)	(0.012)
Female	0.099^{***}	-0.035^{***}
	(0.016)	(0.010)
Constant	0.307^{***}	0.039^{**}
	(0.043)	(0.016)
Observations	137,321	137,321
Adjusted R^2	0.087	0.042

Note: All regressions include fixed effects for years, states, and fields. Standard errors are clustered on the level of states and years. Significance levels: *p<0.1; **p<0.05; ***p<0.01.

Table 17: Regression results for subsample of non-moving and non-delaying students

	Dependent variable:	
	On-time completion	*Forced de-registration*
	(1)	(2)
Treat x 2007	0.049	−0.025
	(0.045)	(0.018)
Treat x 2008	0.022	0.001
	(0.018)	(0.010)
Treat x 2009	−0.012	0.008
	(0.033)	(0.010)
Treat x 2011	−0.044***	−0.011
	(0.014)	(0.014)
Treat x 2012	−0.033	0.018
	(0.026)	(0.025)
Treat x 2013	−0.080***	0.041
	(0.028)	(0.025)
Female	0.093***	−0.053***
	(0.012)	(0.009)
Constant	0.408***	0.074**
	(0.027)	(0.018)
Observations	63,237	63,237
Adjusted R^2	0.079	0.069

Note: All regressions include fixed effects for years, states, and fields. Standard errors are clustered on the level of states and years. Significance levels: *p<0.1; **p<0.05; ***p<0.01.

Table 18: Regression results when excluding two leads

	Dependent variable:	
	On-time completion	*Forced de-registration*
	(1)	(2)
Treat x 2007	0.080	−0.015
	(0.068)	(0.017)
Treat x 2008	0.048*	0.007
	(0.026)	(0.007)
Treat x 2011	−0.036**	0.008
	(0.015)	(0.011)
Treat x 2012	−0.019	0.038**
	(0.025)	(0.016)
Treat x 2013	−0.050**	0.028**
	(0.020)	(0.015)
Female	0.090***	−0.046***
	(0.016)	(0.013)
Constant	0.443***	0.078***
	(0.037)	(0.018)
Observations	127,723	127,723
Adjusted R^2	0.079	0.061

Note: All regressions include fixed effects for years, states, and fields. Standard errors are clustered on the level of states and years. Significance levels: *p<0.1; **p<0.05; ***p<0.01.

Table 19: State-specific regression results for on-time completion with and without neighboring states in control group

	Dependent variable: On-time completion									
	Bavaria		Lower Saxony		Brandenburg		Baden-Wurttemberg		N.-Rhine Westfalia	
	all	w/o neigh.	all	w/o neigh.	all	w/o neigh.	all	w/o neigh.	all	w/o neigh.
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)	(5a)	(5b)
Treat x 2007	0.214***	0.263**	−0.033	−0.028	−0.044	−0.085	0.027	0.066	0.065	0.102
	(0.062)	(0.116)	(0.065)	(0.056)	(0.061)	(0.083)	(0.062)	(0.082)	(0.065)	(0.089)
Treat x 2008	0.019	0.020	0.020	−0.037***	−0.004	0.010	0.041	0.087*	0.009	0.053
	(0.035)	(0.071)	(0.034)	(0.005)	(0.032)	(0.043)	(0.034)	(0.045)	(0.034)	(0.045)
Treat x 2009	−0.038	−0.009	−0.044*	−0.086*	−0.113***	−0.080***	−0.014	−0.013	−0.025	−0.023
	(0.026)	(0.033)	(0.026)	(0.050)	(0.027)	(0.023)	(0.026)	(0.055)	(0.026)	(0.054)
Treat x 2011	−0.064***	−0.059***	−0.061***	−0.080***	0.002	0.013	−0.059***	−0.052***	−0.015**	−0.008
	(0.007)	(0.012)	(0.007)	(0.013)	(0.011)	(0.015)	(0.007)	(0.011)	(0.007)	(0.008)
Treat x 2012	−0.072***	−0.056***	0.268***	0.268***	−0.032**	−0.017	−0.077***	−0.102***	−0.013	−0.037***
	(0.015)	(0.015)	(0.014)	(0.030)	(0.014)	(0.015)	(0.017)	(0.011)	(0.015)	(0.008)
Treat x 2013	−0.048***	−0.037***	−0.310***	−0.331***	−0.038**	−0.017	−0.081***	−0.087**	0.011	0.001
	(0.017)	(0.003)	(0.018)	(0.033)	(0.018)	(0.015)	(0.018)	(0.034)	(0.017)	(0.033)
Female	0.111***	0.109***	0.126***	0.123***	0.092***	0.091***	0.079***	0.078***	0.093***	0.098***
	(0.010)	(0.011)	(0.024)	(0.033)	(0.017)	(0.021)	(0.014)	(0.017)	(0.012)	(0.015)
Constant	0.426***	0.445***	0.429***	0.218***	0.426***	0.412***	0.407***	0.406***	0.467***	0.460***
	(0.047)	(0.061)	(0.065)	(0.080)	(0.070)	(0.090)	(0.069)	(0.076)	(0.047)	(0.062)
Obs.	63,771	53,682	44,426	33,659	34,508	29,126	67,659	54,870	48,688	35,899
Adj. R^2	0.093	0.089	0.112	0.098	0.080	0.084	0.083	0.088	0.092	0.101

Note: All regressions are based on the formula outlined in Equation 1 but are ran seperately on each state-specific subsample. Each subsample includes one treated state and all untreated states (columns 1a, 2a, 3a, 4a, 5a) or only those untreated states that do not have a common border with the treated state (columns 1b, 2b, 3b, 4b, 5b). All regressions include fixed effects for years, states, and fields. Standard errors are clustered on the level of states and years. Significance levels: *p<0.1; **p<0.05; ***p<0.01.

Table 20: State-specific regression results for forced de-registration with and without neighboring states in control group

	Dependent variable: *Forced de-registration*									
	Bavaria		Lower Saxony		Brandenburg		Baden-Wurttemberg		N.-Rhine Westfalia	
	all	w/o neigh.	all	w/o neigh.	all	w/o neigh.	all	w/o neigh.	all	w/o neigh.
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)	(5a)	(5b)
Treat x 2007	-0.038^{*}	0.003	-0.010	-0.026	-0.007	0.020^{**}	-0.050^{**}	-0.069^{***}	-0.013	-0.028
	(0.020)	(0.016)	(0.022)	(0.044)	(0.018)	(0.009)	(0.021)	(0.023)	(0.019)	(0.024)
Treat x 2008	-0.003	-0.002	0.003	0.011^{*}	0.009	0.005	-0.017^{***}	-0.023^{***}	-0.008^{*}	-0.015^{**}
	(0.006)	(0.009)	(0.007)	(0.006)	(0.007)	(0.009)	(0.005)	(0.008)	(0.005)	(0.007)
Treat x 2009	-0.028^{***}	-0.026^{***}	-0.007^{***}	-0.010^{***}	0.0003	0.002^{**}	-0.009^{***}	-0.008^{*}	-0.015^{***}	-0.014^{***}
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.001)	(0.002)	(0.004)	(0.002)	(0.004)
Treat x 2011	0.011^{**}	0.006^{**}	-0.005	-0.005	-0.004	-0.007	-0.016^{***}	-0.009	-0.014^{**}	-0.006
	(0.006)	(0.003)	(0.006)	(0.010)	(0.005)	(0.007)	(0.006)	(0.008)	(0.006)	(0.009)
Treat x 2012	0.069^{***}	0.070^{***}	0.026^{***}	0.022^{**}	-0.012^{***}	-0.009^{**}	0.010^{***}	0.013^{***}	-0.018^{***}	-0.016^{***}
	(0.003)	(0.002)	(0.005)	(0.009)	(0.004)	(0.005)	(0.003)	(0.004)	(0.003)	(0.004)
Treat x 2013	0.048^{***}	0.054^{***}	-0.052^{***}	-0.061^{***}	-0.043^{***}	-0.034^{***}	0.036^{***}	0.036^{*}	-0.035^{***}	-0.037^{**}
	(0.010)	(0.010)	(0.009)	(0.016)	(0.010)	(0.004)	(0.009)	(0.019)	(0.009)	(0.019)
Female	-0.065^{***}	-0.066^{***}	-0.060^{**}	-0.064	-0.073^{***}	-0.078^{**}	-0.060^{***}	-0.042^{***}	-0.064^{***}	-0.038^{***}
	(0.019)	(0.023)	(0.029)	(0.039)	(0.028)	(0.032)	(0.020)	(0.008)	(0.025)	(0.011)
Constant	0.075^{**}	0.082^{*}	0.081^{*}	0.170^{***}	0.107^{**}	0.126^{**}	0.096^{***}	0.065^{***}	0.084^{**}	0.039^{***}
	(0.037)	(0.046)	(0.045)	(0.058)	(0.046)	(0.050)	(0.033)	(0.019)	(0.041)	(0.012)
Obs.	63,771	53,682	44,426	33,659	34,508	29,126	67,659	54,870	48,688	35,899
Adj. R^2	0.071	0.082	0.077	0.080	0.070	0.082	0.045	0.043	0.056	0.042

Note: All regressions are based on the formula outlined in Equation 1 but are ran seperately on each state-specific subsample. Each subsample includes one treated state and all untreated states (columns 1a, 2a, 3a, 4a, 5a) or only those untreated states that do not have a common border with the treated state (columns 1b, 2b, 3b, 4b, 5b). All regressions include fixed effects for years, states, and fields. Standard errors are clustered on the level of states and years. Significance levels: *p<0.1; **p<0.05; ***p<0.01.

Table 21: Regression results for placebo treatment in 2010

	Dependent variable:	
	On-time completion	*Forced de-registration*
	(1)	(2)
Treat x 2007	0.062	−0.018
	(0.079)	(0.020)
Treat x 2008	0.032	−0.0005
	(0.024)	(0.004)
Treat x 2010	−0.007	0.006
	(0.017)	(0.005)
Female	0.091^{***}	-0.032^{***}
	(0.015)	(0.008)
Constant	0.366^{***}	0.042^{*}
	(0.038)	(0.017)
Observations	84,382	84,382
Adjusted R^2	0.075	0.046

Note: Instead of the real treatment in 2011, I assume a placebo treatment period for 2010. All regressions include fixed effects for years, states, and fields. Standard errors are clustered on the level of states and years. Significance levels: *p<0.1; **p<0.05; ***p<0.01.

Table 22: Regression results on subsamples according to fresher-professor-ratio

	Dependent variable: *On-time completion*				
	Increase of fresher-professor-ratio by:				
	any increase	a minimum of 10 %	a minimum of 25 %	a minimum of 50 %	no increase
	(1a)	(1b)	(1c)	(1d)	(2)
Treat x 2007	0.046	0.058	0.073	0.048	0.043
	(0.057)	(0.058)	(0.064)	(0.057)	(0.067)
Treat x 2008	0.045	0.048	0.027	0.054	0.004
	(0.030)	(0.032)	(0.038)	(0.034)	(0.027)
Treat x 2009	−0.006	−0.010	0.006	0.055^{*}	−0.037
	(0.023)	(0.024)	(0.023)	(0.025)	(0.021)
Treat x 2011	-0.054^{***}	-0.062^{***}	-0.074^{***}	-0.100^{***}	-0.076^{***}
	(0.007)	(0.009)	(0.011)	(0.006)	(0.007)
Treat x 2012	-0.057^{***}	-0.066^{***}	-0.051^{***}	-0.087^{***}	−0.001
	(0.014)	(0.013)	(0.011)	(0.015)	(0.039)
Treat x 2013	-0.049^{*}	-0.067^{***}	−0.038	-0.116^{***}	-0.134^{***}
	(0.021)	(0.018)	(0.021)	(0.023)	(0.035)
Female	0.095^{***}	0.090^{***}	0.107^{***}	0.107^{***}	0.081^{**}
	(0.013)	(0.010)	(0.011)	(0.015)	(0.026)
Observations	82,838	75,202	47,538	32,262	80,459
Adjusted R^2	0.081	0.080	0.092	0.106	0.092

Note: All regressions are based on subsamples according to the changes in the ratio of freshers to professors. The regressions also include fixed effects for years, states, and fields as well as a constant (see Equation 1). Standard errors are clustered on the level of states and years. Significance levels: *p<0.1; **p<0.05; ***p<0.01.

Table 23: Regression results on subsamples according to fresher-assistant-ratio

	Dependent variable: On-time completion				
	Increase of fresher-assistant-ratio by:				
	any increase	a minimum of 10 %	a minimum of 25 %	a minimum of 50 %	no increase
	(1a)	(1b)	(1c)	(1d)	(2)
Treat x 2007	0.026	0.008	0.045	0.034	0.061
	(0.055)	(0.056)	(0.057)	(0.058)	(0.069)
Treat x 2008	0.048	0.026	0.047	0.023	0.008
	(0.030)	(0.034)	(0.032)	(0.045)	(0.026)
Treat x 2009	−0.003	−0.017	0.016	0.042	−0.031
	(0.023)	(0.027)	(0.021)	(0.031)	(0.020)
Treat x 2011	−0.071***	−0.073***	−0.100***	−0.086***	−0.056***
	(0.008)	(0.010)	(0.006)	(0.011)	(0.008)
Treat x 2012	−0.078***	−0.077***	−0.055**	−0.071**	0.012
	(0.014)	(0.013)	(0.015)	(0.026)	(0.036)
Treat x 2013	−0.072***	−0.062**	−0.053**	−0.110***	−0.104**
	(0.018)	(0.021)	(0.018)	(0.023)	(0.039)
Female	0.090***	0.098***	0.102***	0.104***	0.089***
	(0.010)	(0.007)	(0.013)	(0.014)	(0.023)
Observations	77,920	67,662	38,180	32,727	85,710
Adjusted R^2	0.077	0.087	0.098	0.104	0.097

Note: All regressions are based on subsamples according to the changes in the ratio of freshers to research assistants. The regressions also include fixed effects for years, states, and fields as well as a constant (see Equation 1). Standard errors are clustered on the level of states and years. Significance levels: *p<0.1; **p<0.05; ***p<0.01.

Table 24: Regression results on subsamples according to fresher-temporary staff-ratio

	Dependent variable: On-time completion				
	Increase of fresher-temporary staff-ratio by:				
	any increase	a minimum of 10 %	a minimum of 25 %	a minimum of 50 %	no increase
	(1a)	(1b)	(1c)	(1d)	(2)
Treat x 2007	0.052	0.068	0.060	0.076	0.062
	(0.052)	(0.051)	(0.054)	(0.054)	(0.072)
Treat x 2008	0.038	0.046	0.042	0.058	0.0002
	(0.031)	(0.031)	(0.034)	(0.032)	(0.029)
Treat x 2009	−0.006	0.0003	−0.014	0.003	−0.033
	(0.024)	(0.024)	(0.025)	(0.022)	(0.018)
Treat x 2011	-0.049^{***}	-0.045^{***}	-0.050^{***}	-0.064^{***}	-0.074^{**}
	(0.001)	(0.001)	(0.002)	(0.009)	(0.020)
Treat x 2012	-0.055^{**}	-0.049^{**}	-0.063^{***}	-0.061^{**}	0.013
	(0.016)	(0.017)	(0.016)	(0.017)	(0.041)
Treat x 2013	-0.043^{*}	-0.038^{*}	-0.061^{**}	-0.079^{**}	-0.106^{**}
	(0.018)	(0.019)	(0.018)	(0.021)	(0.030)
Female	0.088^{***}	0.085^{***}	0.079^{***}	0.078^{***}	0.101^{**}
	(0.011)	(0.016)	(0.012)	(0.020)	(0.028)
Observations	87,341	80,240	69,427	46,238	65,490
Adjusted R^2	0.084	0.095	0.095	0.094	0.090

Note: All regressions are based on subsamples according to the changes in the ratio of freshers to temporary academic staff. The regressions also include fixed effects for years, states, and fields as well as a constant (see Equation 1). Standard errors are clustered on the level of states and years. Significance levels: *p<0.1; **p<0.05; ***p<0.01.

Table 25: Regression results on subsamples according to fresher-student assistants-ratio

	Dependent variable: On-time completion				
	Increase of fresher-student assistants-ratio by:				
	any increase	a minimum of 10 %	a minimum of 25 %	a minimum of 50 %	no increase
	(1a)	(1b)	(1c)	(1d)	(2)
Treat x 2007	0.046	0.059	0.100	0.057	0.065
	(0.058)	(0.060)	(0.057)	(0.052)	(0.069)
Treat x 2008	0.073^{**}	0.073^{**}	0.092^{**}	0.068^{***}	-0.005
	(0.026)	(0.026)	(0.029)	(0.017)	(0.028)
Treat x 2009	0.011	0.008	0.033	0.060^{***}	-0.052^{*}
	(0.020)	(0.020)	(0.021)	(0.012)	(0.024)
Treat x 2011	-0.042^{***}	-0.044^{***}	-0.045^{**}	-0.061^{*}	-0.057^{***}
	(0.006)	(0.009)	(0.018)	(0.030)	(0.008)
Treat x 2012	-0.053^{*}	-0.050^{*}	-0.047^{*}	-0.026	-0.007
	(0.024)	(0.025)	(0.023)	(0.036)	(0.037)
Treat x 2013	-0.065	-0.063	-0.035	-0.012	-0.104^{**}
	(0.035)	(0.037)	(0.029)	(0.059)	(0.037)
Female	0.093^{***}	0.094^{***}	0.098^{***}	0.111^{***}	0.092^{***}
	(0.020)	(0.019)	(0.020)	(0.020)	(0.019)
Observations	72,525	69,427	52,794	36,012	83,048
Adjusted R^2	0.086	0.089	0.075	0.097	0.086

Note: All regressions are based on subsamples according to the changes in the ratio of freshers to student assistants. The regressions also include fixed effects for years, states, and fields as well as a constant (see Equation 1). Standard errors are clustered on the level of states and years. Significance levels: *p<0.1; **p<0.05; ***p<0.01.

Table 26: College fixed effects and state fixed effects

	Dependent variable:			
	On-time completion		*Forced de-registration*	
	Incl. fixed effects for:		Incl. fixed effects for:	
	Colleges	States	Colleges	States
	(1a)	(1b)	(2a)	(2b)
Treat x 2007	0.083	0.051	−0.021	−0.029
	(0.052)	(0.061)	(0.019)	(0.018)
Treat x 2008	0.010	0.021	0.009	−0.007
	(0.036)	(0.030)	(0.012)	(0.005)
Treat x 2009	−0.029	−0.027	−0.006	-0.013^{**}
	(0.022)	(0.024)	(0.005)	(0.006)
Treat x 2011	-0.062^{***}	-0.063^{***}	−0.004	−0.005
	(0.006)	(0.008)	(0.007)	(0.009)
Treat x 2012	−0.037	-0.045^{*}	0.024	0.025^{*}
	(0.021)	(0.026)	(0.013)	(0.013)
Treat x 2013	-0.080^{***}	-0.077^{***}	0.015	0.015
	(0.021)	(0.022)	(0.013)	(0.014)
Female	0.072^{***}	0.092^{***}	-0.042^{***}	-0.045^{***}
	(0.011)	(0.015)	(0.010)	(0.011)
Observations	152,244	152,244	152,244	152,244
Adjusted R^2	0.115	0.077	0.087	0.059

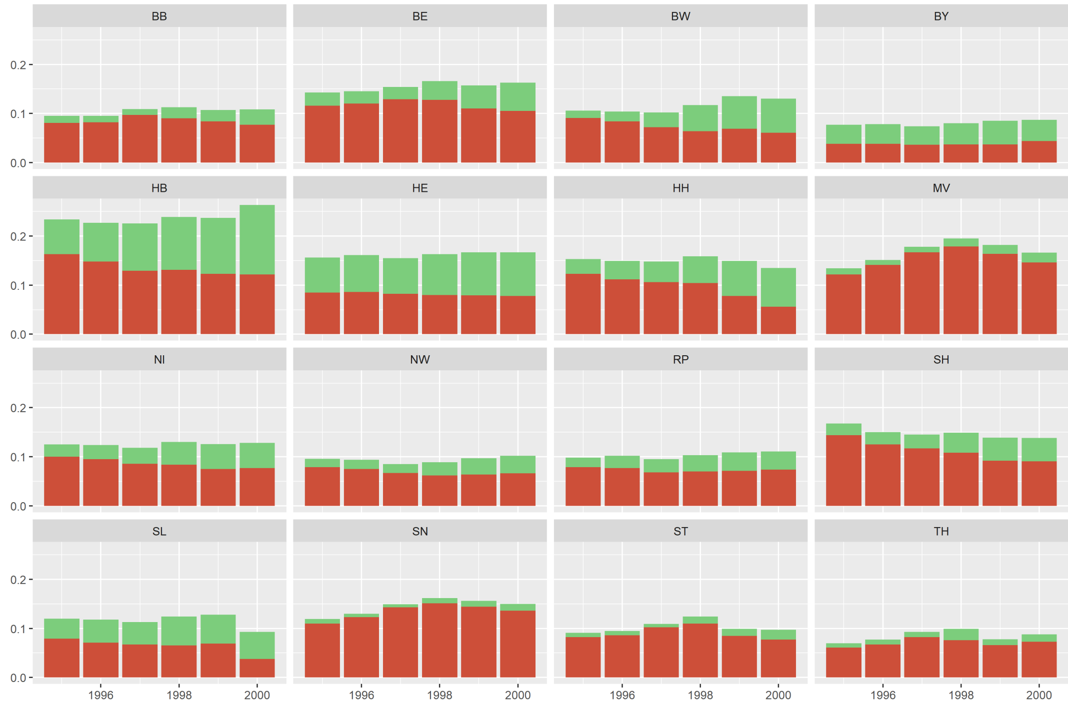
Note: All regressions are based on Equation 1. All regressions also include fixed effects for years, fields, and institutions (columns 1a and 2a) or years, fields, and states (columns 1b and 2b). Significance levels: *p<0.1; **p<0.05; ***p<0.01.

G Additional figures

Early and late entry to primary school in Germany

Students who first enrolled in undergraduate courses between 2006 and 2016 must have started primary school in the mid- to end-1990s. Figure 23 illustrates the shares of pupils who had an early (green) or late (red) entry to primary school in the different German states between 1995 and 2000. It shows that there is a lot of heterogeneity across states. In Bremen (HB), for instance, roughly one quarter of a birth cohort either postpones primary school entry or starts school earlier than required. These shares are similar in some states such as Saarland (SL) and Hesse (HE). In other states such as Mecklenburg-Vorpommern (MV) or Saxony (SN) there are almost no kids who start school earlier than they are required to.

Figure 23: Early and late entry to primary school in German states



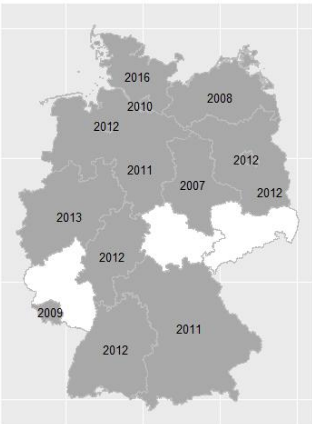
Note: The figure illustrates early and late entry to primary school in all German states between 1995 and 2000. Red bars indicate a late entry, green bars indicate an early entry to primary school. The y-axis indicates the share of non-regular school entrants as part of the primary school entry cohort. Source: Own illustration based on data from Autorengruppe Bildungsberichterstattung (2018).

Location and timing of double cohorts and tuition fees

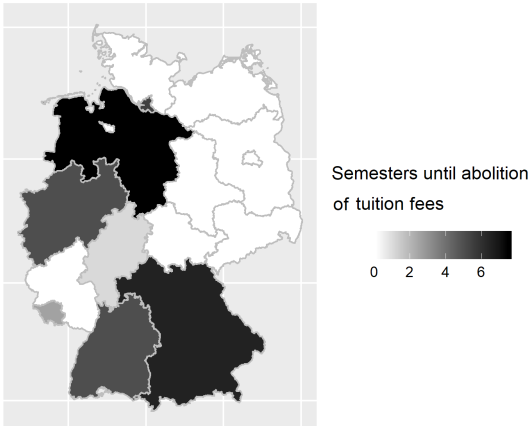
Germany has sixteen different states, each sovereign over its educational system. Panel A in Figure 24 illustrates the timing and spatial variation of the secondary schooling reforms (G8 reforms) that affected the number of potential first-year university students. Panel B illustrates the location of states that charged tuition fees between 2006 and 2014.

Figure 24: Spatial and temporal variation in major reforms

A) Years and states of double cohorts



B) Length and states of tuition fees

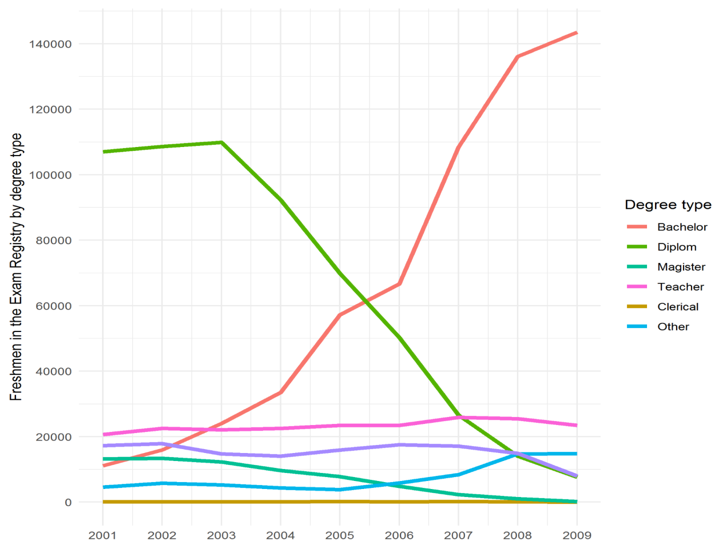


In Panel A all states coloured in gray undertook the schooling reform and released a double cohort from secondary school in the year indicated. Panel B visualizes the spatial distribution of states that charged tuition fees. Coloured states introduced tuition fees in 2007, only North Rhine-Westfalia started charging their students from 2006 onwards. Own illustration.

Bologna Reform: College entrants by degree type

Figure 25 illustrates the number of college entrants per undergraduate degree type. The Bologna Process, a transnational higher education reform that started in 1999 in order to establish a unified European Higher Education Area, aimed at establishing standardized study programs and degrees across Europe. An essential element of this convergence process in Germany was the creation of a two-tier system of professional qualifications (Bachelor's and master's degrees). As a result, traditional German degree programs such as the *Diplom* had to be converted into the new two-tier system. Figure 25 illustrates the number of registered first-year students by type of their expected degree. It was not until 2006 that there were more college entrants in Bachelor's than Diploma programs.

Figure 25: Number of college entrants by undergraduate degree types



Note: The figure illustrates the number of college entrants in different degree types that completed their undergraduate degree until the end of 2016. The figure is based on all German higher education institutions and all fields of study. Source: Own illustration based on data from RDC (2019c).