

Huebener, Mathias; Marcus, Jan

Article — Accepted Manuscript (Postprint)

Compressing instruction time into fewer years of schooling and the impact on student performance

Economics of Education Review

Provided in Cooperation with:

German Institute for Economic Research (DIW Berlin)

Suggested Citation: Huebener, Mathias; Marcus, Jan (2017) : Compressing instruction time into fewer years of schooling and the impact on student performance, Economics of Education Review, ISSN 0272-7757, Elsevier, Amsterdam, Vol. 58, pp. 1-14, <https://doi.org/10.1016/j.econedurev.2017.03.003>

This Version is available at:

<https://hdl.handle.net/10419/202466>

Standard-Nutzungsbedingungen:

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

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

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

Terms of use:

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

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

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



<http://creativecommons.org/licenses/by-nc-nd/4.0>

Compressing instruction time into fewer years of schooling and the impact on student performance

Mathias Huebener,^{a,b*} Jan Marcus^{a,c}

^a DIW Berlin, ^b Freie Universität Berlin, ^c University of Hamburg

March 2017

Abstract

Is it possible to compress instruction time into fewer school years without lowering education levels? A fundamental reform in Germany reduced the length of academic track schooling by one year, while increasing instruction hours in the remaining school years to provide students with a very similar core curriculum and the same overall instruction time. Using aggregated administrative data on the full population of students, we find that the reform increases grade repetition rates and lowers final grade point averages, without affecting graduation rates. The results suggest adverse reform effects on student performance, but the economic significance of the effects appears moderate.

Keywords: Instruction time; student performance; G8 education reform; difference-in-differences; wild cluster bootstrap

JEL: I21, I24, I28, J18, D04

* Corresponding author: Mathias Huebener, Department of Education and Family, DIW Berlin, Mohrenstraße 58, 10117 Berlin, Germany. Email: mhuebener@diw.de.



I Introduction

High levels of education and early labour market entries are both important for the international competitiveness of economies, but there seems to be a trade-off between the two. This trade-off and the debate on the optimal number of school years is becoming increasingly important as general levels of education increase across OECD countries (OECD, 2015). A recent school reform in Germany provides novel insights into the possibility of educating children up to the same level of skills over a shorter period of time. German students were typically aged 19 or 20 and among the oldest students across all OECD countries to acquire the qualifications needed to study at university. Peers in the UK, the US and Japan typically earn the equivalent qualifications at age 17 or 18 (OECD, 2008, p. 498). In response to this situation, several German federal states reduced the number of years needed to earn the relevant qualification for university entrance (*Abitur*) at academic track schools from 13 to 12 years. Unlike straight reductions in years of schooling, the German reform sought to simultaneously maintain high levels of education by increasing the number of instruction hours in the remaining academic track school years. Consequently, students affected by the reform must cover a very similar curriculum and the same overall instruction time across their school career as before the reform to earn the school leaving certificate – despite having one year less of schooling. Affected students have longer school days in which they cover curriculum content that has been gradually brought forward from previously higher years.

In this study, we investigate the impact of this controversial and major education reform on student performance in school. Using aggregated administrative data covering the entire population of students in academic track schools across several cohorts, we analyse the impact on three important measures of overall student performance: grade repetition rates, graduation rates and final grade point averages (GPA). For the identification of effects, we use the fact that the reform was implemented at different points in time across the German federal states. Our difference-in-differences setting enables us to account for major concerns in the analysis of grade repetition rates, graduation rates and GPAs, such as differences between education systems across federal states, teacher grading biases and general

time trends.

We find that the probability of repeating a grade throughout academic track schooling increases by 3.6 percentage points (26 percent), indicating students' problems with covering more content per school year during longer school days in the remaining school years. The effects are stronger for boys and concentrate in the final years prior to school completion. The reform also reduced the average GPA by about 4.4 percent of a standard deviation. Yet, we do not find evidence of reform effects on the proportion of students completing academic track schooling with the university entrance qualification. The results are robust to a variety of sensitivity checks. We neither find evidence of differential trends in the outcome variables between treatment and control states before the introduction of the reform, nor of changes in the student body due to the reform. Both results strengthen a causal interpretation of our findings.

This study links two strands of the literature on the role of time as an input factor in the educational production process. Previous studies either look at the impact of (i) changes in the amount of instruction time provided in a given school year, or (ii) changes in the number of years of schooling. The first strand finds that expansions in the number of days spent in school (e.g. Marcotte, 2007; Sims, 2008; Fitzpatrick et al., 2011; Herrmann & Rockoff, 2012) or increases in the number of instruction hours provided in a given time period (e.g. Taylor, 2014; Cortes et al., 2015; Lavy, 2015; Rivkin & Schiman, 2015; Huebener et al., 2016) tend to improve student performance.

Regarding the second strand, there is a large economic literature on the benefits of additional years of schooling (e.g. Card, 1999; Lochner, 2011). The effects are often identified through changes related to minimum schooling requirements. The findings relate to students on the verge of early school dropout, who experience high monetary and non-monetary returns to additional years of schooling. However, it is not clear what these studies imply for students who are not on the verge of dropping out of school, i.e. students aiming at earning higher school leaving certificates. A policy experiment in Canada provides initial insights into the benefits of additional years of schooling for students bound for higher education. The Canadian

reform shortened university preparatory high school by one year.¹ Morin (2013) and Krashinsky (2014) examine the impact of the reform on students' academic performance at university. Both studies find a strong negative effect on student performance ranging between 17 to 120 percent of a standard deviation, with larger reductions in student performance for lower-ability students. This suggests that the missing school year had indeed added substantial value to student performance.

While both more instruction time per school year and more years of schooling seem beneficial for student learning, it is not clear whether the effects of reductions in the number of years of schooling can be offset by compensating increases in instruction time over the remaining school years. The German academic track school reform is the first policy experiment providing us with insights into this question, which is particularly important in understanding the trade-off between high levels of education and early labour market entries. We find some evidence of adverse effects, though these effects appear moderate when compared to the findings of studies analysing the elimination of the final year of schooling without compensating instruction hour increases, such as in the Canadian case. Our findings complement the first set of important results by Büttner & Thomsen (2015), the only other economic study assessing the reform effect on student performance at the end of secondary school.² Büttner & Thomsen estimate the reform effect in one out of 13 treatment states for the first affected cohort, using survey data from students in twelve schools and a different econometric approach. They find that the reform resulted in lower final examination results in mathematics, but not in German.

We extend the existing literature in several dimensions. First, we examine three different measures of student performance. Second, we extend the analyses beyond the first affected cohort, and apply a difference-in-differences approach to estimate effects for subsequent treatment cohorts. This is an important refinement as the first

¹The reform in Ontario also reorganised the curriculum in the remaining school years.

²Other studies evaluate the reform effect on non-cognitive skills (Thiel et al., 2014, find no effects; Dahmann & Anger, 2014, provide evidence for higher extraversion and lower emotional stability), on measures of crystalline and fluid intelligence (Dahmann, 2015, finds no general effect, but some evidence for improvements among males) and on student performance in ninth grade when the reform serves as a quasi-experiment to estimate the effect of increased instruction time (Huebener et al., 2016, find small positive average effects and stronger improvements among high-performing students). For overviews on recent findings on effects of this reform, see Huebener & Marcus (2015) and Thomsen (2015).

treated cohort graduated under unusual conditions: It graduated from school in the same year as the last pre-treatment cohort. Therefore, students in the first treated cohort might have different performance incentives as they compete for limited resources at universities (e.g. available places, student-teacher-ratio) and for places in vocational training courses with the preceding cohort (Morin, 2015a,b). Third, studying several treated cohorts allows us to analyse whether the reform effects are only of transitory nature or whether the effects also persist in the medium-run. Fourth, we investigate the reform effect across most treatment states, which allows us to control for cohort effects and other education policies. Fifth, we employ administrative data covering the entire underlying population of students.

Further, this study carries important implications for endogenous sample selection issues in the growing literature on the impact of this education reform on post-secondary education outcomes and on labour market outcomes. The reform effect on grade repetition rates implies that in each affected cohort, the share of students repeating a year is higher. Depending on the sample selection of the researcher, this affects the ability distribution in treatment and control cohorts, thus potentially posing challenges for an unbiased estimation of reform effects. This is less relevant for studies that sample students before grade 10, as we only find reform effects on grade repetitions in later grades. Further, the zero-effect of the reform on the graduation rate suggests that a compositional change in the group of students with the qualification for university entrance is not an important source for sample selection biases.

The remainder of this study is structured as follows. Section II provides information on the German education system and the reform analysed in this study. Section III describes the data, and Section IV outlines the empirical strategy. Section V reports the main reform effects, followed by an analysis of effect heterogeneities in Section VI. Section VII provides a broad range of sensitivity checks and Section VIII concludes.

II Institutional background and the G8-reform

We derive the effects of compressing instruction time on student performance from a reform of German academic track schools. The German education system

tracks students into different school types according to their ability after four years of primary school (when students are typically ten years old). The high-ability, academic track school *Gymnasium* prepares students for university education. This track is the primary route for students to earn the general university entrance qualification in Germany (*Abitur*). About one third of each birth cohort attends an academic track school.³

Historically, academic track students in Germany earned their university entrance qualification around the age of 19 or 20. In response to this relatively high age compared to other OECD countries (OECD, 2008), several federal states implemented a reform which reduced the number of years spent in academic track schools by one year (commonly referred to as the G8-reform). As a result of the reform, students could earn the university entrance qualification in a total of 12 rather than 13 years of schooling, which corresponds to eight years of academic track schooling (G8) rather than nine years (G9). At the same time, the reform sought to maintain high levels of education. It increased the number of instruction hours through grades 5 to 12, such that students had to cover (i) a very similar core curriculum and (ii) the same number of total instruction hours to earn the general university entrance qualification (KMK, 2013).⁴ This is a unique feature of the German G8-reform compared to policies in other countries changing the number of years of schooling.⁵ Following the reform, the number of weekly instruction hours was higher, and curriculum content was taught earlier in a student's school career. The education ministries of the federal states determined the exact changes of the timetables after consulting

³Some federal states offer alternative school tracks that allow students to earn their *Abitur*. These alternative school tracks were generally not affected by the G8-reform. There is no evidence that the reform has an impact on the school track choice. Section VII provides a more detailed discussion on this issue.

⁴Before and after the reform, students must cover at least 265 weekly instruction hours (i.e. the sum of the total number of instruction hours per week across all grades) in the course of academic track schooling to earn the general university entrance qualification. G9-students needed 265 instruction hours per week distributed across nine grades (i.e. $265/9 \approx 29.4$ instruction hours per week on average). G8-students needed 265 instruction hours per week distributed across eight grades (i.e. $265/8 \approx 33.1$ instruction hours per week on average).

⁵The reduction in the length of high school in Ontario, Canada, aimed at reducing the cost of the education system (Morin, 2013; Krashinsky, 2014). The cost-saving argument is less pronounced in the German context as treated students still receive the same amount of total instruction hours in the course of their schooling. Therefore, the personnel expenses (accounting for 80 percent of total public school expenditures; German Federal Statistical Office, 2013) could not be reduced significantly.

education researchers and practitioners to best cover the previous curriculum under the new conditions. Generally, the reform affected cohorts newly entering academic track schools after primary school.⁶ There were no other simultaneous and substantive changes with respect to the number of school holidays, the criteria for moving to the next school year, and the performance requirements for earning the university entrance qualification.

The G8-reform was implemented by 13 out of 16 federal states between 2001 and 2007 (see Table 1). The first treatment cohort left school in 2007 in Saxony-Anhalt. Each year, other federal states followed in the implementation of the reform. Two federal states had always applied the shorter length of academic track schooling.⁷

Note that the shortening of secondary track schooling by one year led to the simultaneous school graduation of the first treatment cohort and the last pre-treatment cohort. The two cohorts are referred to as the double graduation cohort.

III Data

We use national administrative data and examine grade repetition rates, academic track school graduation rates, and final grade point averages (GPAs) as indicators of student performance. Germany does not conduct standardised national final school assessments. Consequently, we deem the indicators employed for the current analysis to be the best nationwide measures of student performance upon graduation. Information is provided by the German Federal Statistical Office (2015a) and by the *Kultusministerkonferenz* (KMK, 2015), the assembly of the 16 federal state ministers of education and cultural affairs.

The administrative data have three main advantages. First, they cover the full student population at academic track schools. Second, the data are reported an-

⁶Exceptions are the federal states of Mecklenburg-Vorpommern and Saxony-Anhalt. The first affected cohorts were already in grade 8 (Mecklenburg-Vorpommern) and grade 9 (Saxony-Anhalt) when they were subject to the G8-reform. We discuss the role of these exceptional, surprised cohorts for our estimated reform effects in the robustness checks in Section VII.

⁷The G8-reform has fuelled a fervid public debate on adverse effects of the G8-reform on students. In response, some federal states have (partially) reversed the reform and gone back to the previous system, but the reversals do not affect the cohorts in our analyses. The following states announced (partial) returns to G9: North Rhine-Westphalia in 2 percent of academic track schools with first graduates in 2020, Hesse and Lower Saxony providing both options for students and with first graduates in 2020 and 2021, and Baden-Württemberg in 10 percent of academic track schools with first graduates in 2022.

nually, thus making it possible to use the regional and temporal variations of the G8-reform implementation to identify treatment effects. Third, information about grade repetitions, graduations and final GPAs are reported by the schools to the statistical offices and education ministries. Non-response and social desirability bias are not an issue. The data also have some shortcomings. Generally, the data is only available in aggregated form at the school year-federal state level, so we cannot link individuals across different outcomes.⁸ There are also no socio-economic background variables available in the data. This limits the options for exploring respective effect heterogeneities. Also, we cannot distinguish between students of the first affected cohort and the last pre-treatment cohort in the aggregated data, as these students leave school in the same year. The federal state of Hesse introduced the reform over a period of three years, and we also cannot distinguish between treated and untreated students of these cohorts in the data. Hesse is therefore excluded from our analyses. Further, the number of observations differs between the three outcome variables (see below) because some states did not provide the relevant information in specific years and because of changes in reporting regulations for one outcome variable. In order to increase the power of our analysis for each outcome we include as many observations as possible. Yet, the results are also robust to using different sample restrictions.

In the following, we describe the three analysed outcome variables and the resulting samples. Summary statistics of the main outcome variables are reported in Table 2.

A. Grade repetition rate

As a first measure of student performance, we follow Lee & Barro (2001) and Pischke (2007), and employ grade repetition rates as an indicator of students' ability to cope with the increased number of weekly instruction hours. Students in Germany need to fulfil certain learning objectives before moving on to the next grade. If they do not meet these objectives, they are required to repeat the grade, or to move to

⁸Aggregating individual-level data at the cohort-state level yields the same results as working with the appropriately weighted individual-level data when the treatment only varies at the federal state level (Angrist & Pischke, 2009, p. 235). As the G8-treatment status also only varies at the federal state level, ecological fallacies are of no concern here.

a lower school track.⁹ We observe grade repetition rates at academic track schools for each federal state and each grade from the school years 1994/1995 to 2012/2013. For our main analysis, we sum the number of students $R_{sc,g}$ repeating grade g from grade 7 until graduation (G) for each cohort c and federal state s , and divide this sum by the cohort size in grade 7 at academic track schools ($N_{sc,7}$):¹⁰

$$y_{sc}^{rep} = \frac{\sum_{g=7}^G R_{sc,g}}{N_{sc,7}} \cdot 100$$

This yields a measure of a cohort’s overall grade repetition rate throughout academic track schooling. Note that for G8, grade repetitions are summed up over six academic years (i.e. $G = 12$), while for G9, they are summed up over seven years (i.e. $G = 13$) as the time to school completion is different. We base our calculations on the number of students in grade 7 rather than grade 5 as this is the first school year in which tracking has taken place in all federal states. We multiply the ratio by 100 for an immediate percentage point interpretation of the coefficient estimates in the results tables. Grade-specific reform effects are examined in Section VI.

The federal state of Lower Saxony lacks information on grade repetitions for the final three years of academic track schooling. This federal state is therefore excluded from the analyses for this outcome. The main sample for the analysis of grade repetition rates consists of 168 state-cohort observations, containing information from the 2002 to 2013 graduation cohorts across 14 federal states.¹¹

⁹In Section VII, we discuss that movements away from academic track schools are unlikely to have a major effect on our findings. Students may also repeat a grade level strategically. We discuss this in more detail in Section VI.C.

¹⁰For example, for the cohort that graduated in the school year 2012/2013 under G8, we count the number of grade repetitions from grade 7 in the school year 2007/2008, grade 8 in 2008/2009, ..., and grade 12 in 2012/13. The numbers are reported at the beginning of the next school year. For an interpretation of reform effects on the individual probability to repeat a grade, one needs to assume that individuals repeat only once. This assumption appears reasonable. In PISA 2012, only 4.1 percent of academic track students who repeated a grade level after primary school repeat again.

¹¹When we refer to the grade repetition rate of a certain “graduation cohort”, we assume the regular time to graduation for students in grade 7. While this is true for the majority of students, some students at academic track schools in grade 7 may graduate later or never. For coherence with the other outcome variables, we stick to the term graduation cohort.

B. Graduation rate

As a second measure of student performance, we analyse the graduation rate as a frequently-used measure for the effectiveness of schooling (see, e.g. Evans & Schwab, 1995; Lochner & Moretti, 2004). This measure is informative about whether the reform had an impact on the current number of students in the economy who left secondary school with the general university entrance qualification. Fewer graduates ultimately reduce the number of students allowed to enrol in university.

We observe the total number of students who earned their university entrance qualification from academic track schools, aggregated by school year and federal state. As the absolute number of academic track school graduates depends on the size of the respective birth cohorts, we divide this number by the respective average cohort size of 18 to 20 year olds living in the federal state (provided by the German Federal Statistical Office, 2015b). Again, we multiply the share by 100. In Section VII, we also calculate the graduation rate with several alternative normalisations, i.e. the average cohort size of 18 to 19 year olds, 18 to 21 year olds, the number of academic track school students in grade 7, and the number of students at regular school entry.

The main sample for the analysis of graduation rates consists of 180 state-cohort observations, containing information from the 2002 to 2013 graduation cohorts across 15 federal states.

C. Grade point average

As a third measure of student performance, we use the mean final grade point average (GPA) of students earning the general university entrance qualification *Abitur*. In Germany, the GPA serves as the main criterion in the university admission process, and it is also a reliable predictor of university success (Trapmann et al., 2007) and labour market outcomes (Piopiunik et al., 2014).

The GPAs are determined by the grades received in school examinations during the final two years and by the results of final school leaving examinations, which all federal states but one have conducted centrally since 2008 (see Table 1). The grades are converted into a final grade point average that ranges from 1.0 (top) to 4.0 (barely a pass). The distribution of GPAs is plotted in Figure A.1 in the appendix.

The effective norms of the calculation of the final grade point average are decreed by the federal education ministries. Students successfully complete academic track schooling by passing the school leaving examinations.

We use information on the frequency distribution of GPAs obtained from KMK (2015), and compute mean GPAs by graduation cohort and federal state. The data contain information on all students being awarded the *Abitur*, i.e. on students from academic track schools affected by the reform, as well as on students from alternative school tracks that are not affected by the reform, but in which the general university entrance qualification can also be earned. While the majority of students covered by this data graduate from academic track schools, 20 percent of students earn their *Abitur* from alternative school tracks, namely from comprehensive schools (*Integrierte Gesamtschulen*) or from schools with a vocational training focus (*Berufsschulen*). The inclusion of students unaffected by the reform in our data introduces a potential attenuation bias to estimated reform effects. In robustness checks reported in Section VII.B, we employ two different strategies to develop an idea of the extent of this attenuation and yield the same conclusions.

For ease of interpretation, we standardise the mean GPA using the mean and the standard deviation of the GPA distribution in our main sample and reverse the scale so that higher values indicate better grades.¹² We multiply this measure by 100, so that our point estimates can be interpreted as percent changes of a standard deviation.

Our main sample for the analysis of GPAs consists of 120 state-cohort observations containing information from the 2006 to 2013 graduation cohorts across 15 federal states. We have additional information on the 2002 to 2005 graduation cohorts, but changes in reporting regulations created a structural break in the data that makes us less convinced by the legitimacy of this information for our analy-

¹²Neither linear transformation affects our conclusions, and the results are robust to using raw GPAs.

ses.¹³ The estimated reform effect on the GPA captures the effect of compressing instruction time into fewer years of schooling and of the resulting younger age-at-test of students. This joint effect is of significant policy relevance as the final GPA is a measure of students' educational performance and a good indicator of their preparation for post-secondary education paths given that they are now younger.

IV Empirical strategy

To identify G8-reform effects, we make use of the fact that the reform was implemented at different points in time across the federal states. We exploit this variation with a difference-in-differences (DiD) approach. We estimate the G8-reform effect on grade repetition rates, graduation rates and GPAs as dependent variables y using the following model:

$$y_{sc} = \beta \cdot G8_{sc} + \mu_s + \kappa_c + \delta \cdot DGC_{sc} + X'_{sc} \cdot \lambda + P'_{sc} \cdot \phi + \varepsilon_{sc}. \quad (1)$$

$G8_{sc}$ is a binary variable that indicates whether graduation cohort c in federal state s is affected by the G8-reform. β is the coefficient of core interest and identifies the G8-reform effect. A set of state fixed effects (μ_s) captures general, cohort-invariant differences in the outcome variables between federal states, such as state-specific requirements for grade repetitions, state differences in grading policies, or state differences in the level of difficulty of final exams. The state fixed effects also account for a large share of the time-invariant component in a potential teacher bias in the outcome variables, as the labour market for teachers is very persistent over time. Cohort fixed effects (κ_c) capture general changes in the outcome variables across cohorts, as well as shocks common to all federal states. The nationwide suspension of compulsory military service in 2011, birth cohort effects, or general changes in grading over time (such as grade inflation) are thereby accounted for. We further include an indicator for the double graduation cohort (DGC), consisting of students

¹³Between 2002 and 2005, the GPA information provided by the KMK (2015) contains data on students graduating with the general university entrance qualification from academic track schools and from comprehensive schools. From 2006 onward, the data also contains GPAs of students graduating from school with a vocational training focus. The different school tracks cannot be distinguished in the data. KMK (2015) explicitly recommends avoiding comparisons across the structural break. We perform a robustness check in Section VII.B.

from the first G8-treatment cohort and students from the final pre-treatment cohort. The data do not allow us to distinguish between these cohorts, as they leave school simultaneously. Furthermore, performance incentives for students of the double graduation cohort may differ compared to other cohorts.¹⁴

When we present our results in Section V, we gradually extend our baseline model by further control variables. This serves two main purposes. First, it relaxes the common trend assumption, a key assumption of our identification strategy. Second, it allows us to check for the orthogonality of the G8-reform indicator to other factors. In the first extension of the model, we include a vector of economic control variables (X_{sc}). Several studies document the impact of business cycles on educational decisions, such as grade repetitions, university enrolment and university graduation (see, e.g. Edwards, 1976; Gaini et al., 2013). We therefore include GDP growth and the general unemployment rate (German Federal Statistical Office, 2015b) measured in the year prior to graduation to account for changes in federal states' economic environments that might impact the dynamic path of our outcome variables across federal states. Furthermore, children born in Eastern Germany shortly after the fall of the Wall are also included in our sample. These birth cohorts experienced an unprecedented decline in overall fertility rates (Goldstein & Kreyenfeld, 2011), with evidence of adverse parental selection in these birth cohorts (Chevalier & Marie, 2016). Hence, we indicate observations of Eastern German cohorts born up to two years after the fall of the Wall with a separate dummy variable, with birth calculations based on regular school duration until graduation.

The model is flexible enough to check for the potentially confounding role of other education reforms that are not captured by state or cohort fixed effects. We include a vector of binary variables for other significant education reforms (P_{sc}) passed by some federal states during the sampling period (see Table 1). None of the policy changes considered are collinear to the G8-reform. The first relevant change is the introduction of centralised school leaving examinations, which may impact student

¹⁴The double graduation cohort is roughly double the size of a regular cohort, with students competing for resources at universities (e.g. available places, student-teacher-ratio), and for places on vocational training courses and volunteer programmes. This might alter performance incentives for students compared to other cohorts (Morin, 2015a,b). The robustness checks in Section VII show that the results are insensitive to completely excluding the double graduation cohort from the sample.

performance (Jürges et al., 2005). Furthermore, some federal states combined the middle- and low-ability track in the German three-tier system, which may impact the school track choice and the student composition within tracks (Bauer & Riphahn, 2006; Malamud & Pop-Eleches, 2011). Finally, the subject choices in the final years of academic track schooling were reduced in some federal states, which may also affect our outcomes (Görlitz & Gravert, 2016).

The remaining variation in the outcomes is captured by the error term ε_{sc} . In difference-in-differences settings, a general problem is that ε_{sc} might be correlated within federal states (Bertrand et al., 2004). A common way of accounting for the correlation structure of the error term is to cluster standard errors at the federal state level. For statistical inference with a small number of clusters – as encountered in our analyses – Cameron et al. (2008) suggest wild cluster bootstrapping procedures, which we employ throughout our analyses.¹⁵ For the main results, we also report p -values for conventional clustering at the federal state level. Both procedures yield the same conclusions.

V Results

We start the presentation of our results with a graphical inspection. Figure 1 displays the development of the three outcome variables in the treatment states around the introduction of the G8-reform, adjusted for state and cohort fixed effects. The figure shows that grade repetition rates increased discontinuously after the introduction of the G8-reform. Also the GPA declines discontinuously after the introduction of the G8-reform. The graduation rate does not show such a discontinuous change after G8 is introduced.

We now present the estimation results of the empirical model (equation 1) in Table 3. Panel A provides the estimates on the grade repetition rate. The baseline difference-in-differences specification in column 1 only includes the G8-treatment indicator, federal state and cohort fixed effects, and a double cohort dummy. The estimation result suggests an increase in the grade repetition rate of 4.3 percentage

¹⁵We rely on a specification with Mammen weights, 999 replications, and testing under H_0 with 14 federal state clusters in the analyses of grade repetition rates, and 15 federal state clusters in the analyses of graduation rates and GPAs.

points. The effect is not precisely estimated, a common phenomenon in difference-in-differences regressions (Brewer et al., 2013). Gradually adding further control variables (X_{sc} and P_{sc}) increases the precision of our estimates. In our main specification in column 3, the G8-reform increases the grade repetition rate at academic track schools by 3.6 percentage points, or 26 percent.¹⁶ The effect is statistically significant at the 5-percent level, irrespective of whether the conventional clustering or the wild cluster bootstrapping procedure is used to calculate p -values. This result provides empirical evidence that an increased number of students fail to satisfy the required learning goals. The reform increased weekly instruction hours to cover additional curriculum content. Weaker students on the verge of repeating a school year may encounter difficulties understanding the increased volume of curriculum content in each school year. Huebener et al. (2016) analyse the G8-reform effect on the performance of students in grade 9. They find that low-performing students benefit least from the increased classroom time, which is in line with the increase in grade repetition rates typically occurring among the weakest students.

Panel B of Table 3 reports the G8-reform effect on graduation rates. The estimated effect is small in magnitude and statistically insignificant across the three model specifications. This suggests that despite increased grade repetition rates, the share of individuals graduating from an academic track school with the university entrance qualification is not affected. The findings somewhat contrast Eide & Showalter (2001) and Jacob & Lefgren (2009), who provide evidence that grade repetitions also reduce high school completion rates. Compared to these studies, our case focuses on students in the high-ability track. These grade repeaters have a strong performance incentive to leave school with the university entrance qualification, a prerequisite for university education. This result is also consistent with the literature on sheepskin effects in the returns to education, suggesting that the completion of school degrees, rather than the number of school years generate significant returns in the labour market (see, e.g. Hungerford & Solon, 1987; Jaeger & Page, 1996; Dickson & Smith, 2011).

¹⁶Calculations of percentage changes are based on the counterfactual means for the treatment group in the treatment period, i.e. $E(y|G8 = 1, post = 1) - \hat{\beta}$. The post-treatment means are reported in the notes to Table 3.

Panel C of Table 3 reports the G8-reform effect on the GPA. The baseline difference-in-differences estimate for the mean GPA is -6.9 percent of a standard deviation. When the full set of control variables is included, the coefficient estimate is -4.4 percent of a standard deviation. The estimated reform effects are attenuated as our data also includes approximately 20 percent of students from alternative school tracks that were not affected by the G8-reform.¹⁷ Overall, our findings are in line with Büttner & Thomsen (2015), who analyse the G8-reform effect on final exam grades in mathematics and language arts. Such final exam grades are one component in the final GPA that we analyse.

The shortened school duration was accompanied by significant increases in instruction hours, such that the total instruction time and the school curriculum provided to students remained very similar. Nevertheless, the estimated reform effects imply that students' performance is poorer: Grade repetition rates increase and GPAs decline. There are at least two explanations that can rationalise the findings. First, fatigue and the declining concentration of students over the extended school days may diminish the benefits of additional instruction time on a given day (Rivkin & Schiman, 2015). Second, students may lack maturity for certain classroom material or fail to carry out sufficient preparation to enable them to process the curriculum content that now appears relatively earlier in their academic careers. Findings from a high school programme in the US support this argument. As a consequence of teaching more difficult algebra courses in middle school rather than in high school, Allensworth et al. (2009) and Clotfelter et al. (2015) find negative effects on student performance in later mathematics courses.

Although we find evidence of negative effects on student performance, the economic significance of the effects seems moderate for three reasons. First, it should be acknowledged that the majority of students successfully graduate in the shorter period of time. A rather small proportion of students cope with the increased learning intensity by repeating a school year, but ultimately also graduate with the university entrance qualification. As a consequence of the reform, academic track school graduates are, on average, 0.86 years or 10.4 months younger (see Table A.1 in the

¹⁷In Section VII, we perform different robustness checks to adjust the coefficient estimates for the potential attenuation bias.

appendix). The second reason we think that treatment effects are moderate is that our estimated G8-effects on the GPA are much smaller than the effects of the shortened secondary school duration in Ontario. Depending on the group of students, performance at university declined by between 17 and 120 percent of a standard deviation in the Canadian case, with low-performing students suffering the most (Morin, 2013; Krashinsky, 2014).¹⁸ These effects might even be lower-bound estimates as they are estimated in the group of students moving on to university and not for the entire group of affected students at the end of high school. Third, note that the estimated effects on grade repetitions and GPAs not only capture the effect of one year less of schooling (with the same total instruction time), but they also include effects of treated students facing certain classroom material and sitting final exams at a younger age. This younger age may itself lead to lower student performance. Unfortunately, our data does not allow to separate the treatment effect of receiving the same number of instruction hours in one year less of schooling from the effect of being one year younger at degree completion. However, for the reform in Ontario, Canada, age-effects can only account for a small fraction of the large, overall reform effect (Krashinsky, 2014). We therefore believe that age-at-test effects do not play a significant role for our findings. The presence of such effects would strengthen the impression that the negative reform impact on student performance is moderate.

If we now use our results to project G8-reform effects on students' university success and labour market outcomes, we would expect the effects on these outcomes to also be moderately negative. First studies examine the G8-reform effects on post-secondary education decisions and university performance and indeed only find small effects. Meyer et al. (2015) and Meyer & Thomsen (2016) find that some students delay the start of university by one year. By the second year after graduation, the effects on the transition to university have almost vanished. They also find no effects on university subject choices. Insights regarding the reform's effect on student performance at university are still missing due to the recency of the reform.

¹⁸Parts of the substantial negative effects of the Ontario reform could also be explained by the reorganisation of the curriculum. In particular, certain topics in mathematics appeared several years earlier, which may itself have a negative impact on student performance if students are not sufficiently prepared (e.g. Allensworth et al., 2009; Clotfelter et al., 2015).

We conclude that redistributing instruction hours and the school curriculum over fewer school years cannot entirely compensate for the deterioration in student performance caused by reductions in the years of schooling, but the adverse effects are moderate.

VI Effect heterogeneities

This section analyses whether average treatment effects mask heterogeneities in the effects by gender, over time, and across grades.

A. By gender

A broad, interdisciplinary literature documents gender differences in brain structures, brain functioning and hormonal levels (see, e.g. Cahill, 2006), in socialisation (see, e.g. Bertrand & Pan, 2013) and in personality and non-cognitive skills (see, e.g. Spinath, 2014). These differences may cause a differential impact of the G8-reform on girls and boys, because they may differ in their capacity to concentrate on longer school days or to organise their activities within the tighter time schedule, for example.

Table 4 reports the estimated G8-reform effects on grade repetition rates and graduation rates separately by gender. The GPA information is not available separately by gender. Boys experience higher absolute reform-induced increases in grade repetition rates than girls. This is in line with findings by Büttner & Thomsen (2015) who find larger negative effects on boys' performance in final mathematics exams. However, considering the original level of grade repetition rates for boys and girls, we find that the relative impact of the reform on grade repetition rates is almost equal. The G8-reform effect on graduation rates is close to zero and insignificant for both boys and girls, indicating no important gender differences.

B. Over time

To shed light on the transitory nature of the effects, we examine the evolution of the treatment effects over time. We interact the G8-dummy with indicators for each year after the implementation of the reform to estimate the reform effect separately for cohorts following the implementation (see Table 5). Each reform-interaction effect is identified by fewer cohorts than the main effect which reduces

the statistical power. Yet, the results across cohorts are very similar to our main findings and there is no evidence that the reform effects on grade repetition rates and GPAs diminish for subsequent cohorts. Again, we do not find any significant effects on graduation rates. We conclude that the effects are not only transitory implementation effects.

C. By grade level

The data on grade repetitions allow us to differentiate the reform effect by grade levels. As the increase in instruction hours and learning intensity occurred across all grades in academic track school, this analysis sheds some light on whether learning deficiencies accumulate over time. We adjust equation 1 by interacting the G8-dummy with indicators for each grade. We additionally include grade fixed effects, grade-state fixed effects and grade-cohort fixed effects. The previous sample size of $N = 168$ increases to $N \cdot 6 = 1008$ observations.¹⁹

As between grade 7 and school completion there are seven school years for G9-students but only six years for G8-students, we have to omit one grade under the G9-regime for a grade-to-grade comparison. For grades 7 to 9, we always compare the respective grades. Furthermore, we compare the respective final two years at academic track schools as they constitute the years in which students earn grades that count towards the final GPA, i.e. we compare the penultimate grade to the penultimate grade (*qualification phase 1*, denoted by Q1), and the final grade to the final grade (*qualification phase 2*, denoted by Q2). Less clear is whether grade 10 in the G8-regime should be compared to the previous grade 10, or to the previous grade 11, as its new role comprises institutional characteristics of both. We report both comparisons in Table 6.

We find significant and sizeable effects especially in the final years at academic track school. The effect size and the significance in grade 10 depends on the comparison group. The relative effect sizes on grade repetition rates increase with grades. As the majority of grade repeaters repeat only once in the course of schooling, an increased probability to repeat one grade lowers the probability to repeat a subsequent

¹⁹We examine the effects on grade repetitions starting in grade 7 as this is the earliest grade in which students are tracked into academic track schools across all federal states.

grade. Some students who would have repeated a higher grade in the absence of the reform may now already do so in an earlier grade. While overall grade repetition rates are not affected by such shifts between grades, estimated grade level effects may be distorted towards lower grades. Yet, the small and insignificant effects in grades 7 to 9 indicate that these distortions are not very strong. The concentration of grade repetition effects in the final years of schooling support the idea of accumulated learning deficits of students. Weaker students on the verge of repeating a school year may have accumulated learning deficits across the different grades. While in grades 7 to 9, to a certain extent, it is still at the discretion of teachers as to whether or not students are pushed on to the next grade, this scope for the teacher to affect students' progress into the next year declines as learning deficits increase and as centralised exams begin to feature.²⁰

An alternative explanation for the pattern across grade levels could be voluntary, strategic grade repetitions. The incentive to postpone school graduation is particularly strong in the double graduation cohort when more students compete for the same post-secondary schooling resources (Morin, 2015b). However, the double cohort does not contribute to our effect estimates. Furthermore, the costs associated with grade repetitions are high, and they may increase as shorter academic track schooling becomes the standard.²¹ However, the grade repetition effects can also be observed six years after the double cohort (see Table 5).²² If some of the grade repetitions still occur for strategic reasons, our interpretation of moderate reform effects on student performance would be strengthened.²³

²⁰Moving to other school tracks is unlikely to explain the pattern in the null-effects in earlier grades, which we discuss in more detail in Section VII.A.

²¹For example, students need to resit all courses and exams, the social environment is interrupted, and grade repetition attaches social stigma to students. Moreover, grade repeaters could be identified more easily in the labour market as time passes since the reform implementation. See, e.g. Eide & Goldhaber (2005) for the economic costs of grade repetition. As the German education system is grade-based rather than credit point-based, accelerated graduation by completing courses from higher grades in earlier grades is generally not possible.

²²Another argument refers to gender differences in competitive situations. Morin (2015b) shows that boys behave more strategically than girls in the competitive environment of the double cohort in Canada. As we show in Section VI.A, the relative G8-reform effects on girls and boys are very similar.

²³We also examined whether the treatment effects differ by state characteristics. As states differ in many more dimensions than there are states, it is not clear whether differences in treatment effects are due to the respective state characteristic or due to other correlated dimensions. While the effects on grade repetitions rates are higher in states where a higher share of students attends the academic school track, treatment effects do not differ much between poorer and richer states

VII Sensitivity checks

A. Threats to the identification strategy

This section deals with the two main assumptions of our identification strategy required for the causal interpretation of the estimated reform effects: the common trend assumption and the absence of reform-induced compositional changes.

With respect to the common trend assumption, our data contain a series of pre-treatment observations that allow us to simulate different placebo reforms within our empirical model. We pretend that the reform would have taken place one to four years before the actual reform, and drop all treatment observations from our sample. We substitute the G8-treatment indicator in equation 1 by placebo reform indicators one at a time. The estimation results for these placebo treatments are reported in columns 2 to 5 of Table 7. All coefficient estimates are small in magnitude and insignificant, suggesting that treatment and control states did not follow different time trends prior to the treatment. As an alternative check for the common trend assumption, we substitute the set of control variables (X_{sc} and P_{sc}) in equation 1 with linear state-specific time trends (see column 6 of Table 7). This specification also confirms the main pattern in our findings. Co-treatments with other education reforms may also violate the common trend assumption. Note, however, that our findings are very similar irrespective of whether we include or exclude controls for other significant education reforms (compare columns 2 and 3 of Table 3). In sum, the specifications presented here support the assumption of a common time trend in student performance of treatment and control states in the absence of treatment.

The second key assumption is that the G8-reform had no impact on the composition of the group of students in treatment and control states. Federal states that implemented the reform passed new regulations regarding the length of academic track schooling, and all academic track schools were obliged to comply with the regulation. Students trying to avoid the reform had to (i) opt for another school track

(in terms of public debts) and between states performing on average below and above the German PISA 2000 mean. We also looked at the G8-reform effect at various quantiles of the GPA distribution. These unreported regressions did not provide evidence of differential effects across the GPA distribution. Note, however, that ceiling effects on the GPA measure may hide differential effects across the GPA distribution.

or (ii) move to another state that had not (yet) implemented the reform. Choosing another school track can have lasting consequences for individuals. Academic track schools are the most common way of earning the qualification required for university education. Commuting or moving to other federal states entails high costs for the child and its family, and becomes even less feasible as more federal states implement the reform. Any behaviour of avoiding academic track schools should become evident from enrolment rates in academic track schools. We explicitly check for G8-reform effects on the share of grade 7 students enrolled in academic track schools by using the ratio between the number of grade 7 students enrolled in academic track schools and the total number of grade 7 students across all school tracks as an outcome variable in equation 1. There is no evidence for the G8-reform having an impact on the share of a cohort enrolled in academic track schools (see Table A.1 in the appendix). Further, it may be that students still enrol in academic track schools, but switch to other school tracks later. First, note that moving between school tracks is very rare. In the school year 2012/2013, 2.4 percent of all students switched between different school types (Autorengruppe Bildungsberichterstattung, 2014). Second, Huebener et al. (2016) test for group compositional changes of students in grade 9 at academic track schools in representative PISA data and do not find any evidence of G8-reform-related changes in student composition by gender, parental education, migration background and age. Third, any drop in the share of students attending academic track schools would translate into lower *Abitur* graduation rates. Our findings, however, suggest a zero-reform effect on the graduation rate. In sum, the evidence suggests that selective migration to other school tracks or federal states is unlikely to be an important source of bias for our findings.

B. Model specifications

This section explores the sensitivity of the findings to different model specifications. First, we examine the role of exceptional circumstances in the introduction of the reform in the federal states of Saxony-Anhalt and Mecklenburg-Vorpommern. The first three treatment cohorts in Saxony-Anhalt were in grades 7 to 9, and the first two treatment cohorts in Mecklenburg-Vorpommern were in grades 7 to 8 when they were assigned to G8. Consequently, the additional workload has been redistributed over fewer remaining years for these “surprised” cohorts, which could affect

the results. In column 7 of Table 7, we incorporate a dummy indicating these cohorts and find that controlling for these circumstances does not alter our conclusions.

Further, we examine the sensitivity of the results to different ways of accounting for the double graduation cohort. First, we exclude them from the sample (see column 8 of Table 7). Second, we interact the federal state dummies with the double graduation cohort dummies (column 9 of Table 7). The method of dealing with the double cohort also has no impact on our findings.

We also check whether the way we calculate the graduation rate impacts on our findings. While in our main specification we normalise the number of graduates with the average cohort size of the 18 to 20 year old individuals, our results are robust to normalising the number of graduates with the average cohort size of 18 to 19 year olds, 18 to 21 year olds, the number of academic track school students in grade 7, and the number of students at regular school entry (see Table A.2 in the appendix). The different normalisations confirm the conclusion that the G8-reform had no impact on the graduation rate.

The next set of sensitivity checks is dedicated to the GPA. We first address the aforementioned attenuation bias in the estimates (see columns 2 and 3 of Table A.3 in the appendix). Approximately 20 percent of the students in the GPA data earned their university entrance qualification from alternative school tracks which were not subject to the G8-reform. Therefore, we interact the G8-indicator with the share of graduates from academic track schools among all students leaving school with the university entrance qualification. This approach scales the treatment effect by the actual share of treated students in the cell. In the second test, we exclude observations from the three treatment states with the highest shares of *Abitur* graduates from alternative school tracks: Baden-Württemberg (31 percent), Berlin (30 percent) and Hamburg (28 percent). As we would expect, in both specifications the point estimates increase, but the magnitude of the attenuation bias does not alter our conclusion on moderate overall reform effects on the final GPA. Finally, our findings on the GPA may be attenuated because the GPA is only defined for individuals who pass the final examinations. If more students fail the final examinations because of the G8-reform, our estimates would be biased towards zero. We estimate the G8-effect on the share of students failing the final examinations, and

find a small positive, but insignificant effect of 0.30 percentage points (compared to a mean of 4.41, see Table A.1 in the appendix).²⁴ Incorporating failing students in the GPA measures slightly increases the magnitude of the estimated reform effect, but our conclusions are the same (see column 4 of Table A.3).²⁵

The last GPA-related sensitivity check addresses the structural break in the reporting of aggregated GPA information described in Section III. Between 2002 and 2005, the aggregated data did not include graduates from schools with a vocational training focus. They account for about 14 percent of all graduates. In column 5 of Table A.3, we add information for the 2002-2005 graduation cohorts before this structural break to our sample, which only increases the number of pre-treatment observations. While the point estimate is similar, it is estimated with more noise.

C. Construct validity and interrelations between the outcome variables

The validity of the employed outcomes as indicators of student performance would be compromised if the outcomes were prone to policy-makers' manipulation. There are several reasons why this manipulation is unlikely to play a major role in the present case. First, education policy-making is decentralised and in the responsibility of the 16 federal states. Isolated attempts of policy-makers to affect the outcome variables are unlikely to have an impact on our findings as we identify treatment effects across several treatment states. Concerted actions across federal states would have been noticed by the general public. Second, we are looking at three different measures of student performance, none of which can be manipulated by policy-makers alone. Decisions on grade repetitions are highly decentralised and made at the teacher and school level. Further, we did not find any evidence of official instructions to change grade repetition rates as part of the reform. Also, graduation rates are unlikely to be systematically manipulated by policy-makers. Graduating from an academic track school requires a student to meet graduation requirements and these were not changed by the G8-reform. Furthermore, the GPA is calculated based on two components. The first component consists of grades awarded at the

²⁴Students with insufficient performance in the final two years are not admitted to sit the final examinations. These students can repeat the final grade. Students who fail the final examinations can also repeat the final grade.

²⁵For this specification, we assigned individuals who failed the final exams the grade "5.0" (before the standardisation), the usual grade associated with failing in Germany.

class level during the final two years. These grades are very difficult to manipulate for policy-makers. The second component in the GPA is the grades earned in final examinations. Attempts by individual states to make them easier are unlikely to have a major impact on our findings as we identify reform effects across several federal states. If policy-makers were still able to impact our measures of student performance, our estimated reform effects would likely be attenuated.

After having discussed the construct validity of our outcome variables, we now turn to a discussion of how the outcome variables and the findings of reform effects on them relate to each other. We first discuss how an increased number of grade repetitions affect our findings on GPAs and graduation rates. One could think of grade repeaters as being misassigned to 12 years of schooling, while they received 13 years of schooling in reality. If we assume that these students catch up in the extra school year, such that their GPA is not affected by the reform, our estimated effects on GPAs would be (slightly) biased towards zero. Back-of-the-envelope calculations suggest very small changes in effect sizes, as the estimated treatment effect on the GPA (-4.41 percent of a standard deviation) would then be a weighted average of the zero-effect for about 3.6 percent of students additionally repeating a grade and the actual treatment effect for students who did not repeat a grade due to the G8-reform (about 96.4 percent). Under this assumption, the actual treatment effect for the non-repeaters would be -4.57 rather than -4.41 percent of a standard deviation.²⁶

How do grade repetitions relate to the graduation rate? In any G8-cohort, additional students repeat a grade due to the G8-reform and move to the younger cohort. This younger cohort receiving the grade repeaters also loses students because of higher grade repetition rates.²⁷ If these outgoing and incoming students balance in total, there is no mechanic relationship between increased grade repetition rates and graduation rates. Only if the treatment effect on grade repetitions substantially changes over time, incoming and outgoing students would not balance. However, we neither find evidence for this (see Table 5), nor do we find evidence for

²⁶If we assume that the additional grade repeaters eventually performed even better due to the reform, the actual treatment effect for non-repeaters would be more negative, but the magnitude of effect differences is very small.

²⁷This does not hold for the first G8-cohort, which does not contribute to the identification of effects in our analysis as it is part of the double cohort.

changes in treatment effects on graduation rates over time.

Next, we discuss the implications of the zero-effect on graduation rates for the other two outcomes. Students who do not pass a grade after compulsory schooling could decide to drop out of school instead of repeating the grade.²⁸ Our estimates for the grade repetition outcome would be biased towards zero, if the reform induced more students to drop out of school. However, increased dropout rates would be reflected in negative point estimates of the effect on graduation rates. This also implies that the effects on the GPA are unlikely to be distorted by fewer or more graduates.

D. External validity

The regional and temporal variations in the implementation of the reform arguably provide good internal validity of our findings. Can our findings also be transferred to other contexts, however? First, it is important to note that we use data capturing the universe of academic track school students in the German states. We therefore estimate *the* average G8-reform effect for the affected cohorts in our sample. One question is whether the treatment effects will differ for future treatment cohorts. Policy experiments carry the risk of observed effects being only transitory implementation effects, e.g. because of learning effects of school administrations, teachers and students. In our analysis, the first affected cohort does not contribute to identifying the reform effects as it is part of the double graduation cohort. Furthermore, the analyses on treatment effects over time (reported in Section VI.A) show that the effects are still present six years after the double graduation cohort, suggesting that they are not only transitory and will translate to subsequent cohorts.

Another question to consider is what would happen if other countries implemented a similar reform reducing the length of schooling and increasing the instruction hours in the remaining school years to compensate for the shortened school duration. Though one can only speculate about this type of external validity without further insights from other contexts, we believe that such a reform would also have similar consequences elsewhere. One likely reason for our finding of weaker

²⁸Compulsory schooling laws require German students to stay at least 9 or 10 years at school, depending on the federal state (Vossenkuhl, 2010).

school performance is that students may benefit at a diminishing rate from additional instruction hours on longer school days. Rivkin & Schiman (2015) provide evidence of such diminishing benefits in international PISA data, which suggests that this is not specific to the German reform. Furthermore, covering additional material on longer school days may overstretch at least some students. Students have less time outside school hours to process and understand the material and revise for tests and exams and they may not be prepared well enough to cope with the new curriculum content. Overtaxing students is also a reason that is likely to apply across institutional contexts. Finally, we cannot entirely rule out that G8-students covered less material than G9-students, or not to the same depth, as weaker students may request repeated explanations of classroom material more often. Such behaviour does also not seem to be specific to the German context.

Based on these considerations, we believe that, also in other countries, a proportionate increase of instruction hours in earlier school years could not entirely compensate for reductions in the years of schooling.

VIII Conclusion

This paper examines an education reform in Germany that shortened academic track schooling by one year, and increased instruction hours in the remaining years at academic track schools to provide the same total instruction time and school curriculum. We focus on the reform effects on student performance observed in administrative data. Our results suggest that compressing instruction time into fewer years of schooling has adverse effects on educational performance. Some students cope with the increased learning intensity by repeating a school year, especially students in the final years of academic track schooling. However, we find no evidence for reform effects on graduation rates. The effect estimates on the GPA are consistently negative, and we argue that several potential sources of attenuation bias strengthen our conclusions on the lower educational performance of students at the end of their school careers. This suggests that affected students have a poorer command of the school material than peers who covered the same material over a longer period of time. These effects are not only of a transitory nature, but are also apparent six years after the introduction of the reform.

Although we find adverse effects on student performance, we argue that the effects are moderate in economic terms. The majority of affected students graduate one year earlier. A small proportion of students deals with the increased learning intensity by repeating a school year, but ultimately also completes school successfully. The effect on the GPA is also small if compared to the high school reform in Ontario, Canada, in which the final year was removed without compensating students with more instruction time in the remaining school years.

The potential costs of the reform due to adverse effects on student performance must be weighed against the economic gains from earlier labour market entries that can mitigate skilled worker shortages, increase social security contributions and raise individuals' lifetime earnings. Compressing instruction time into fewer years of schooling may therefore constitute a previously undiscussed policy option for economies where upper-secondary schooling is comparably long.

Acknowledgements We would like to thank Maximilian Bach, Jörg Breitung, Tarjei Havnes, Mandy Huebener, Nicolai Kristensen, Adam Lederer, Jan van Ours, Carla Welch, and two anonymous referees. Geraldine Frantz and Jakob Simonsen provided excellent research assistance. Mathias Huebener gratefully acknowledges support from the German National Academic Foundation. Jan Marcus greatly appreciates the support from the College for Interdisciplinary Education Research.

References

- Allensworth, E., Nomi, T., Montgomery, N., & Lee, V. E. (2009). College preparatory curriculum for all: Academic consequences of requiring Algebra and English I for ninth graders in Chicago. *Educational Evaluation and Policy Analysis*, 31(4), 367–391.
- Angrist, J. D. & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.
- Autorengruppe Bildungsberichterstattung (2014). *Bildung in Deutschland 2014* (1st ed.). Bielefeld: Bertelsmann.
- Bauer, P. & Riphahn, R. T. (2006). Timing of school tracking as a determinant of intergenerational transmission of education. *Economics Letters*, 91(1), 90–97.
- Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, 119(1), 249–275.
- Bertrand, M. & Pan, J. (2013). The trouble with boys: Social influences and the gender gap in disruptive behavior. *American Economic Journal: Applied Economics*, 5(1), 32–64.
- Brewer, M., Crossley, T. F., & Joyce, R. (2013). Inference with difference-in-differences revisited. *IZA Discussion Paper*, 7742.
- 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.
- Cahill, L. (2006). Why sex matters for neuroscience. *Nature Reviews Neuroscience*, 7(6), 477–484.
- Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, 90(3), 414–427.
- Card, D. (1999). The causal effect of education on earnings. In O. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics*, volume 3A of *Handbook of Labor Economics* chapter 30, (pp. 1801–1863). Amsterdam: North Holland.
- Chevalier, A. & Marie, O. (2016). Economic uncertainty, parental selection, and children's educational outcomes. *Journal of Political Economy*, forthcoming.
- Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2015). The aftermath of accelerating Algebra: Evidence from district policy initiatives. *Journal of Human Resources*, 50(1), 159–188.
- Cortes, K. E., Goodman, J. S., & Nomi, T. (2015). Intensive math instruction and educational attainment: Long-run impacts of double-dose algebra. *Journal of Human Resources*, 50(1), 108–158.

- Dahmann, S. (2015). How does education improve cognitive skills? Instructional time versus timing of instruction. *SOEPpapers on Multidisciplinary Panel Data Research*, 769.
- Dahmann, S. & Anger, S. (2014). The impact of education on personality: Evidence from a German high school reform. *IZA Discussion Paper*, 8139.
- Dickson, M. & Smith, S. (2011). What determines the return to education: An extra year or a hurdle cleared? *Economics of Education Review*, 30(6), 1167–1176.
- Edwards, L. N. (1976). School retention of teenagers over the business cycle. *Journal of Human Resources*, 11(2), 200–208.
- Eide, E. R. & Goldhaber, D. D. (2005). Grade retention: What are the costs and benefits? *Journal of Education Finance*, 31(2), 195–214.
- Eide, E. R. & Showalter, M. H. (2001). The effect of grade retention on educational and labor market outcomes. *Economics of Education Review*, 20(6), 563–576.
- Evans, W. N. & Schwab, R. M. (1995). Finishing high school and starting college: Do Catholic schools make a difference? *The Quarterly Journal of Economics*, 110(4), 941–974.
- Fitzpatrick, M. D., Grissmer, D., & Hastedt, S. (2011). What a difference a day makes: Estimating daily learning gains during kindergarten and first grade using a natural experiment. *Economics of Education Review*, 30(2), 269–279.
- Gaini, M., Leduc, A., & Vicard, A. (2013). School as a shelter? School leaving-age and the business cycle in France. *Annales d’Economie et de Statistique*, 111-112, 251–270.
- German Federal Statistical Office (2013). Bildungsausgaben: Ausgaben je Schüler/-in 2010. *Statistisches Bundesamt*, Wiesbaden.
- German Federal Statistical Office (2015a). Allgemeinbildende Schulen: Fachserie 11, Reihe 1. *Several years*. <https://www.destatis.de/DE/Publikationen/Thematisch/BildungForschungKultur/Schulen/BroschuereSchulenBlick.html>.
- German Federal Statistical Office (2015b). Genesis-Online Datenbank. <https://www-genesis.destatis.de>.
- Goldstein, J. R. & Kreyenfeld, M. (2011). Has East Germany overtaken West Germany? Recent trends in order-specific fertility. *Population and Development Review*, 37(3), 453–472.
- Görlitz, K. & Gravert, C. (2016). The effects of the high school curriculum on school dropout. *Applied Economics*, 48(54), 5314–5328.
- Herrmann, M. A. & Rockoff, J. E. (2012). Worker absence and productivity: Evidence from teaching. *Journal of Labor Economics*, 30(4), 749–782.

- Huebener, M., Kuger, S., & Marcus, J. (2016). Increased instruction hours and the widening gap in student performance. *DIW Discussion Paper*, 1561.
- Huebener, M. & Marcus, J. (2015). Empirische Befunde zu Auswirkungen der G8-Schulzeitverkürzung. *DIW Roundup Politik im Fokus*, 57.
- Hungerford, T. & Solon, G. (1987). Sheepskin effects in the returns to education. *The Review of Economics and Statistics*, 69(1), 175–177.
- Jacob, B. A. & Lefgren, L. (2009). The effect of grade retention on high school completion. *American Economic Journal: Applied Economics*, 1(3), 33–58.
- Jaeger, D. A. & Page, M. (1996). Degrees matter: New evidence on sheepskin effects in the returns to education. *The Review of Economics and Statistics*, 78(4), 733–740.
- Jürges, H., Schneider, K., & Büchel, F. (2005). The effect of central exit examinations on student achievement: Quasi-experimental evidence from TIMSS Germany. *Journal of the European Economic Association*, 3(5), 1134–1155.
- KMK (2013). Vereinbarung zur Gestaltung der gymnasialen Oberstufe in der Sekundarstufe II. Beschluss der Kultusministerkonferenz vom 07.07.1972 i.d.F. vom 06.06.2013. *Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland*, Bonn/Berlin.
- KMK (2015). Uebersicht der Abiturnoten im Ländervergleich. *Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland*, <https://www.kmk.org/dokumentation-und-statistik/statistik/schulstatistik/abiturnoten.html>.
- Krashinsky, H. (2014). How would one extra year of high school affect academic performance in university? Evidence from an educational policy change. *Canadian Journal of Economics*, 47(1), 70–97.
- Lavy, V. (2015). Do differences in schools' instruction time explain international achievement gaps? Evidence from developed and developing countries. *The Economic Journal*, 125(588), F397–F424.
- Lee, J.-W. & Barro, R. J. (2001). Schooling quality in a cross-section of countries. *Economica*, 68(272), 465–488.
- Lochner, L. (2011). Nonproduction benefits of education: Crime, health, and good citizenship. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the Economics of Education*, volume 4 chapter 2, (pp. 183–282). Amsterdam: Elsevier.
- Lochner, L. & Moretti, E. (2004). The effect of education on crime: Evidence from prison inmates, arrests, and self-reports. *American Economic Review*, 94(1), 155–189.
- Malamud, O. & Pop-Eleches, C. (2011). School tracking and access to higher education among disadvantaged groups. *Journal of Public Economics*, 95(11-12), 1538–1549.

- Marcotte, D. E. (2007). Schooling and test scores: A mother-natural experiment. *Economics of Education Review*, 26(5), 629–640.
- 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.
- Meyer, T., Thomsen, S. L., & Schneider, H. (2015). New evidence on the effects of the shortened school duration in Germany: An evaluation of post-school education decisions. *IZA Discussion Paper*, 9507.
- Morin, L.-P. (2013). Estimating the benefit of high school for university-bound students: Evidence of subject-specific human capital accumulation. *Canadian Journal of Economics*, 46(2), 441–468.
- Morin, L.-P. (2015a). Cohort size and youth earnings: Evidence from a quasi-experiment. *Labour Economics*, 32, 99–111.
- Morin, L.-P. (2015b). Do men and women respond differently to competition? Evidence from a major education reform. *Journal of Labor Economics*, 33(2), 443–491.
- OECD (2008). Education at a glance 2008: OECD indicators. *OECD Publishing*.
- OECD (2015). Education at a glance 2015: OECD indicators. *OECD Publishing*.
- Piopiunik, M., Schwerdt, G., & Woessmann, L. (2014). Zentrale Abschlussprüfungen, Signalwirkung von Abiturnoten und Arbeitsmarkterfolg in Deutschland. *Zeitschrift für Erziehungswissenschaft*, 17(1), 35–60.
- Pischke, J.-S. (2007). The impact of length of the school year on student performance and earnings: Evidence from the German short school years. *The Economic Journal*, 117(523), 1216–1242.
- Rivkin, S. G. & Schiman, J. C. (2015). Instruction time, classroom quality, and academic achievement. *The Economic Journal*, 125(588), F425–F448.
- Sims, D. P. (2008). Strategic responses to school accountability measures: It’s all in the timing. *Economics of Education Review*, 27(1), 58–68.
- Spinath, B. (2014). The roles of intelligence, personality and motivation in girls’ outperforming boys at school. *Personality and Individual Differences*, 60(Supplement), S45.
- Taylor, E. (2014). Spending more of the school day in math class: Evidence from a regression discontinuity in middle school. *Journal of Public Economics*, 117, 162–181.
- Thiel, H., Thomsen, S. L., & Büttner, B. (2014). Variation of learning intensity in late adolescence and the effect on personality traits. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 177(4), 861–892.

- Thomsen, S. L. (2015). The impacts of shortening secondary school duration. *IZA World of Labor*, 166(July), 1–10.
- Trapmann, S., Hell, B., Weigand, S., & Schuler, H. (2007). Die Validität von Schulnoten zur Vorhersage des Studienerfolgs - eine Metaanalyse. *Zeitschrift für Pädagogische Psychologie*, 21(1), 11–27.
- Vossenkuhl, A. (2010). (Berufs-) Schulpflicht in Deutschland. *Berufsbildung in Wissenschaft und Praxis*, 39(6), 53–54.

Figures

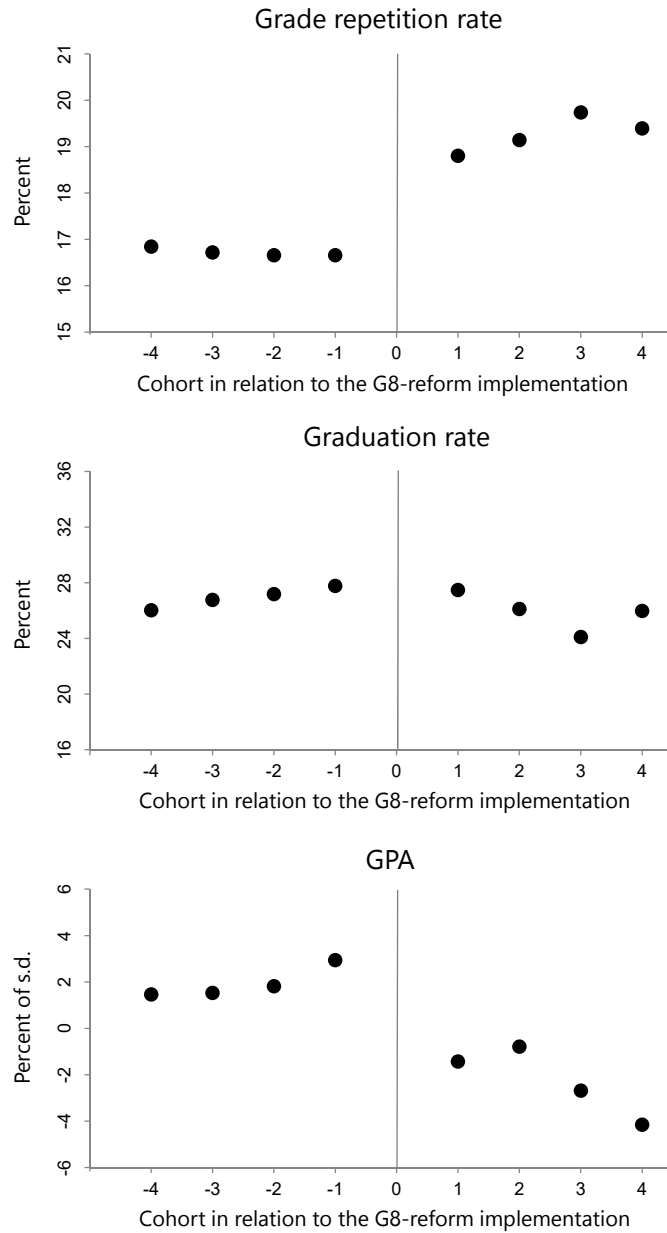


Figure 1: Evolution of student performance around the implementation of the G8-reform.

Notes: The graphs plot the evolution of the three measures of student performance for cohorts preceding and following the implementation of the G8-reform in the treatment states. Cohort “0” refers to the double graduation cohort. The graphs display the residuals of regressions of the outcome variables on state and cohort fixed effects, and a dummy for the double graduation cohort. For ease of interpretation, overall means were added to the grade repetition rate and the graduation rate.

Source: Own illustration based on data from the German Federal Statistical Office (2015a) and KMK (2015).

Tables

Table 1: Implementation of the G8-reform and other education reforms in the federal states

	G8	Centralised exit examination	Two-tier system	Restricted upper-secondary subject choice
Change from G9 to G8				
Saxony-Anhalt	from 2007	all	from 2006	from 2005
Mecklenburg-Vorpommern	from 2008	all	from 2010	from 2008
Saarland	from 2009	all	from 2006	from 2010
Hamburg	from 2010	from 2005	none	from 2011
Bavaria	from 2011	all	none	from 2011
Lower-Saxony	from 2011	from 2006	none	from 2008
Baden-Württemberg	from 2012	all	none	from 2004
Bremen	from 2012	from 2007	from 2012	all
Berlin	from 2012	from 2007	none	all
Brandenburg	from 2012	from 2005	from 2012	none
Hesse	from 2012	from 2007	none	from 2005
North Rhine-Westphalia	from 2013	from 2007	none	all
Always G8				
Saxony	all	all	all	from 2010
Thuringia	all	all	all	from 2011
Always G9 (in sample period)				
Rhineland-Palatinate	none	none	none	from 2011
Schleswig-Holstein	from 2016	from 2008	none	from 2011

Notes: The table reports how the graduation cohorts in our sample period of 2002-2013 are affected by different education reforms and institutional changes. *Centralised school exit examinations* shift the responsibility for the design of exit exams from schools to federal state institutions such that all students in the specific state sit the same exit exam. *Two-tier system* indicates reforms that combine the low- and middle-ability track in the traditional German three-tier school track system. *Restricted upper-secondary subject choice* indicates reforms in which students' set of subject choices for the final two years at academic track school has been restricted.

Source: Numerous sources, available from the authors on request.

Table 2: Summary statistics of the main outcome variables

Outcome variable	Mean			Sample		
	Overall	Girls	Boys	Graduation cohorts	Number of federal states	No. of state-cohort observations
Grade repetition rate	18.14 (8.25)	14.32 (7.28)	22.54 (9.33)	2002-2013	14	168
Graduation rate	26.36 (7.39)	30.29 (8.33)	22.65 (6.72)	2002-2013	15	180
Grade point average (GPA), original scale	2.50 (0.63)	—	—	2006-2013	15	120

Notes: The table reports summary statistics of the main outcome variables. Standard deviations are reported in parentheses. The GPA mean is based on the original scale.

Source: Own calculations based on German Federal Statistical Office (2015a) and KMK (2015).

Table 3: G8-reform effects: Main estimation results

	(1) Baseline DiD	(2) Economic controls	(3) Main specification
Panel A: Grade repetition rate			
G8-reform effect	4.33 [0.110] (0.187)	3.35* [0.092] (0.090)	3.59** [0.016] (0.033)
%-change	32.81	23.59	25.70
No. of state-cohort observations (N)	168	168	168
No. of federal states	14	14	14
Panel B: Graduation rate			
G8-reform effect	-0.07 [0.923] (0.956)	-0.27 [0.793] (0.816)	0.22 [0.863] (0.868)
%-change	-0.25	-0.98	0.79
No. of state-cohort observations (N)	180	180	180
No. of federal states	15	15	15
Panel C: GPA			
G8-reform effect	-6.88*** [0.000] (0.039)	-7.35*** [0.000] (0.028)	-4.41* [0.080] (0.195)
No. of state-cohort observations (N)	120	120	120
No. of federal states	15	15	15
Economic controls	No	Yes	Yes
Education reform controls	No	No	Yes

Notes: The table reports estimates of the G8-reform effects on grade repetition rates, graduation rates, and GPAs. The regressions are based on equation 1. They include federal state fixed effects, cohort fixed effects, and a double graduation cohort dummy. Regressions in columns 2 and 3 gradually include economic control variables (GDP growth, unemployment rate in the year before graduation, dummy for Eastern German cohorts born up to two years after the fall of the Wall) and education reform controls (centralised school leaving examinations, restricted upper-secondary subject choice, two-tier tracking systems, see Table 1 for details). The post-treatment means for the grade repetition rate is 17.54, for the graduation rate it is 27.85. The GPA is reversed and standardised with mean 0 and a standard deviation of 100, i.e. coefficient estimates refer to changes in percent of a standard deviation. P -values obtained from wild cluster bootstrapping are reported in brackets. P -values obtained from conventional clustering of the error terms are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Source: Own calculations based on German Federal Statistical Office (2015a) and KMK (2015).

Table 4: G8-reform effects by gender

	(1) Girls	(2) Boys
Panel A: Grade repetition rate		
G8-reform effect	2.89** [0.032]	4.27*** [0.000]
%-change	26.83	24.28
N	168	168
Panel B: Graduation rate		
G8-reform effect	-0.12 [0.991]	0.54 [0.627]
%-change	-0.40	2.22
N	180	180

Notes: The table reports the estimation results of the G8-reform effects on grade repetition rates and graduation rates separately by gender. The dependent variable is gender specific. All models are based on the main specification, outlined in equation 1. *P*-values obtained from wild cluster bootstrapping are reported in brackets. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Source: Own calculations based on German Federal Statistical Office (2015a).

Table 5: G8-reform effects over time

	(1)	(2)	(3)
	Dependent variable		
	Grade repetition rate	Graduation rate	GPA
G8 · 1 year after	3.36*** [0.008]	0.80 [0.434]	-4.09* [0.066]
G8 · 2 years after	3.83** [0.012]	-0.06 [1.000]	-4.30 [0.260]
G8 · 3 years after	4.13* [0.082]	-1.78 [0.404]	-6.95 [0.116]
G8 · 4 years after	3.48 [0.128]	-0.10 [0.945]	-9.37* [0.088]
G8 · 5 years after	6.65** [0.020]	1.61 [0.669]	-8.61* [0.078]
G8 · 6 years after	5.19** [0.038]	-1.72 [0.657]	-9.09 [0.124]
N	168	180	120

Notes: The table reports the estimation results of the G8-reform effects for different years after the implementation of the G8-reform. The regressions are based on equation 1. P -values obtained from wild cluster bootstrapping are reported in brackets. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Source: Own calculations based on German Federal Statistical Office (2015a) and KMK (2015).

Table 6: G8-reform effects on grade repetition rates by grade level

Dependent variable: Grade specific grade repetition rate		
	(1)	(2)
	Grade 10 under G8 compared to	
	Grade 10 under G9	Grade 11 under G9
G8 · grade 7	0.05 [0.921]	0.07 [0.903]
G8 · grade 8	0.38 [0.400]	0.43 [0.292]
G8 · grade 9	0.20 [0.719]	0.24 [0.659]
G8 · grade 10/11	1.03* [0.056]	0.57 [0.108]
G8 · grade Q1	2.77*** [0.002]	2.81*** [0.000]
G8 · grade Q2	1.13** [0.018]	1.17*** [0.002]
N	1008	1008
%-change grade 7	2.87	4.01
%-change grade 8	15.82	18.44
%-change grade 9	8.11	9.95
%-change grade 10	48.12	21.76
%-change grade Q1	83.66	85.83
%-change grade Q2	103.73	110.67

Notes: The table reports the estimation results of the G8-reform effects on grade repetition rates separately by grade. All models are based on the specification outlined in equation 1 and additionally include grade fixed effects, grade-state fixed effects and grade-cohort fixed effects. *P*-values obtained from wild cluster bootstrapping are reported in brackets. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Source: Own calculations based on German Federal Statistical Office (2015a).

Table 7: Sensitivity checks

	(1)	(2)			(3)			(4)			(5)			(6)			(7)			(8)			(9)		
		Main	$t-1$	$t-2$	$t-3$	$t-4$	State-time trends	Common trend	Placebo treatments	w/o surprised cohorts	w/o double cohort	Exceptional cohorts	Interacted double cohort												
Panel A: Grade repetition rate																									
G8-reform effect	3.59**	0.97	0.80	0.89	0.63	2.65*	3.29**	3.54***	3.54***	3.54***	3.54***	3.54***													
	[0.016]	[0.388]	[0.486]	[0.310]	[0.336]	[0.088]	[0.044]	[0.004]	[0.002]	[0.004]	[0.002]	[0.002]													
N	168	120	120	120	120	168	168	158	168	158	168	168													
Panel B: Graduation rate																									
G8-reform effect	0.22	1.01	0.32	0.24	0.32	-0.95	0.22	-0.02	-0.85	-0.02	-0.85														
	[0.863]	[0.242]	[0.573]	[0.452]	[0.152]	[0.146]	[0.893]	[1.000]	[0.496]	[1.000]	[0.496]														
N	180	130	130	130	130	180	180	169	180	169	180														
Panel C: GPA																									
G8-reform effect	-4.41*	0.71	-0.58	-1.13	0.28	-7.14**	-4.47*	-4.20	-4.58	-4.20	-4.58														
	[0.080]	[0.653]	[0.779]	[0.595]	[0.911]	[0.024]	[0.058]	[0.118]	[0.110]	[0.118]	[0.110]														
N	120	78	78	78	78	120	120	109	120	109	120														

Notes: The table reports sensitivity checks for the estimated G8-reform effects on grade repetition rates, the graduation rate and standardised GPAs. The regressions are based on equation 1. Section VII provides more information on the specific models. *P*-values obtained from wild cluster bootstrapping are reported in brackets. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. *Source:* Own calculations based on German Federal Statistical Office (2015a) and KMK (2015).

Appendix

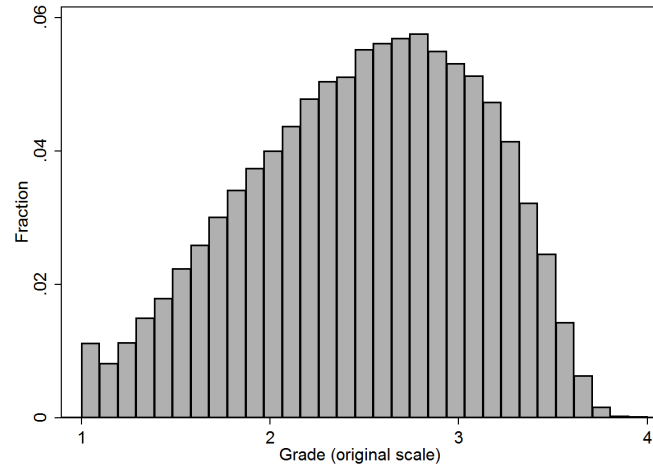


Figure A.1: Distribution of final GPAs (original scale) in the main sample

Notes: The graph plots the proportion of students with the respective GPA according to the original scale ranging from 1.0 (top) to 4.0 (barely a pass) in the analysis sample.

Source: Own illustration based on KMK (2015).

Table A.1: Further outcomes: G8-reform effects on graduation age, the share of students in grade 7 at academic track schools, and the final exam failure rate

	(1)	(2)	(3)
	Dependent variable		
	Graduation age	Share of students in academic track schools in grade 7	Final exam failure rate
G8-reform effect	-0.86*** [0.000]	0.88 [0.212]	0.30 [0.591]
$p(\beta = -1)$	0.00	—	—
No. of state-cohort observations (N)	167	180	117
No. of federal states	15	15	15

Notes: The table reports the estimated G8-reform effects on the graduation age, the share of students enrolled at academic track schools in grade 7, and the share of students who fail the final exit examinations. All models are based on the specification outlined in equation 1. The post-treatment means for the graduation rate is 18.74, for the share of students in grade 7 at academic track schools it is 36.91, and for the final exam failure rate it is 4.41. For graduation age, the potential 180 state-cohort observations (15 federal states over 12 years) reduce to 167 observations as we lack information about the graduation age for the pre-treatment years 2002-2005 in Brandenburg, Hamburg and Baden-Württemberg as well as for Lower Saxony in 2007. For each graduation cohort in each state, the mean graduation age is calculated based on information on the graduates' distribution of birth years. The drop in the graduation age is statistically different from -1, the reduction by one full year as indicated by the p -value of the test $\beta = -1$. The second outcome is computed by dividing the number of students enrolled in academic track schools in grade 7 by the total number of students in grade 7 enrolled in any school track. This ratio is then multiplied by 100. P -values obtained from wild cluster bootstrapping are reported in brackets. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Source: Own calculations based on German Federal Statistical Office (2015a) and KMK (2015).

Table A.2: Outcome-specific sensitivity checks: Graduation rate normalisation

	(1) Post-treatment mean	(2) Baseline DiD	(3) Economic controls	(4) Main specification
Panel A: 18-20				
G8-reform effect	27.85	-0.07 [0.897]	-0.27 [0.807]	0.22 [0.885]
N		180	180	180
Panel B: 18-19				
G8-reform effect	29.19	0.40 [0.727]	-0.05 [0.937]	0.60 [0.685]
N		180	180	180
Panel C: 18-21				
G8-reform effect	26.53	-0.29 [0.775]	-0.29 [0.743]	0.10 [0.939]
N		180	180	180
Panel D: Grade 7				
G8-reform effect	76.19	-0.09 [0.939]	0.23 [0.929]	3.05 [0.156]
N		180	180	180
Panel E: School entry				
G8-reform effect	29.08	0.64 [0.505]	0.40 [0.643]	1.00 [0.320]
N		165	165	165
Economic controls		No	Yes	Yes
Education reform controls		No	No	Yes

Notes: The table reports the estimation results of the G8-reform effect on graduation rates with different normalisations of the outcome variable. Column 1 reports the post-treatment means. The regressions in columns 2-4 are based on equation 1. Panel A reports the original estimate for the share of graduates based on the average population aged 18-20 years. Panels B and C normalise the number of graduates by the average size of the populations aged 18-19 and 18-21 years. Panel D normalises the number of graduates by the size of the graduation cohort in grade 7 (the data does not contain information on school entries for the 2002 cohort and the sample size drops by one observation per state) and Panel E normalises by the size of the regular school entry cohort. *P*-values obtained from wild cluster bootstrapping are reported in brackets. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Source: Own calculations based on German Federal Statistical Office (2015a) and KMK (2015).

Table A.3: GPA-specific sensitivity checks

Dependent variable: GPA					
	(1)	(2)	(3)	(4)	(5)
	Measurement error			Extended period	
	Main	Interacted	w/o states	with failure	2002-2013
G8-reform effect	-4.41*	-5.96*	-4.74	-5.14*	-5.01
	[0.086]	[0.090]	[0.174]	[0.070]	[0.210]
N	120	120	96	117	172

Notes: The table reports sensitivity checks for the G8-reform effect on the GPA. All models are based on the specification outlined in equation 1. The GPA is standardised with mean 0 and standard deviation 100. Column 2 reports the coefficient estimate of the G8-treatment indicator interacted with the share of graduates from academic track schools. Column 3 excludes observations from Baden-Württemberg, Berlin and Hamburg, the three treatment states with the highest shares of *Abitur* graduates from alternative school tracks. Column 4 reports estimates for the whole 2002-2013 period (including the structural break in the data reporting between 2005 and 2006). *P*-values obtained from wild cluster bootstrapping are reported in brackets. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Source: Own calculations based on KMK (2015).