Adverse effects of increased education efficiency? The impact of shortening high school tenure on graduation age, grade repetitions and graduation rates

This version: May 15, 2014

Mathias Huebener*, Jan Marcus*

DIW Berlin

Abstract

In designing education systems, policy-makers face a trade-off between the provision of higher levels of schooling and earlier labour market entries. A fundamental education reform in Germany tackles this trade-off by increasing education efficiency: The time in high school is reduced by one year while the total number of instruction hours is left unchanged.

Employing administrative data on all pupils in Germany, we exploit both temporal and regional variation in the implementation of the reform and study first indicators of the overall effectiveness of this reform. We find that the shortening of the high school track length by one year reduces the mean high school graduation age by 10 months. We show that grade repetition rates double for pupils in the final years before graduation and that this effect is not quickly fading out over time. However, the number of students that graduate with university entrance qualifications not affected. The results indicate the reform's success in reducing graduation age, though it stays behind its full potential benefits for labour markets, pension schemes and fertility because of higher grade repetition rates.

Keywords: Grade Repetition, Graduation Rates, Learning Intensity, Difference-in-Differences, Education Efficiency, Graduation Age
JEL: I28, D04

*Correspondence:
Email addresses: mhuebener@diw.de (Mathias Huebener), jmarcus@diw.de (Jan Marcus)

1. Introduction

In designing education systems, policy-makers face a trade-off in the optimal allocation of the length of schooling. While increasing years of education improves the economy's human capital stock, it also delays labour force participation. This trade-off is particularly crucial in light of the demographic changes that many industrialised countries are facing.

A broad literature documents the beneficial effects of additional years of schooling both for the individual and for society. Among others, education is found to increase individuals' earnings (Card, 1999), civic engagement (Dee, 2004), and health behaviour (Grossman, 2006), to reduce crime (Lochner and Moretti, 2004) and mortality (Lleras-Muney, 2005) and also to strongly contribute to economic growth (Barro, 2001). On the other hand, earlier entry into the labour market has the potential to increase working experience and the individual's life-time income. Many social security systems are confronting the problem of a smaller group of young workers and an increasing number of older, non-working, individuals. This threatens, for instance, the sustainability of public pay-as-you-go pension schemes. Furthermore, ageing societies require policy-makers to respond to skilled worker shortages. An earlier labour market entrance could mitigate these problems. Additionally, a lower age at school leaving might increase fertility as most women get their children after they finish education and enter the labour market (Skirbekk et al., 2004).

A fundamental reform of the German education system aims at resolving this trade-off by increasing education efficiency. Between 2001 and 2007, several German states passed laws reducing the high school track by one year, but leaving the overall instruction time unchanged. This policy change reduces the time to obtain the university entrance diploma from 13 to 12 years by redistributing the same number of

lessons over the remaining school years. It is referred to as G12 (denoting graduation after 12 years).[1]

In this study, we examine first indicators of the overall effectiveness of the G12 reform by looking at three different outcomes: grade repetitions, mean graduation age and graduation rates. As the reform's principal objective was to reduce the age at which students graduate from high school, our first outcome directly investigates whether the reform has been effective in this regard. Contrary, by the nature of grade repetitions, our second outcome, every pupil that is affected by it cannot benefit from an earlier labour market entry. The two outcomes are strongly related as increased grade repetition rates will translate into higher graduation ages and subsequently in a reduction in the reform's overall effectiveness in reducing graduation age. With the third outcome, graduation rates, we investigate whether the reform has an impact on the society's human capital stock. Graduation rates might be affected by the policy change through increased disutility from schooling due to the higher learning intensity, increased graduation rates and/or lower opportunity costs in terms of years in school.

Employing administrative data on all pupils in Germany, this is one of the first studies exploiting both temporal and regional variation in the implementation of the reform. Difference-in-differences estimates reveal that G12 reduces the graduation age by about 10 months. One possible explanation why it stays behind its potential of one full year lies in increased grade repetition. We find a doubling of grade repetitions in the final two years prior to graduation. Repetition rates in lower

[1] In the German context the reform is often referred to as G8, for 8 years of *Gymnasium* (high school). However, we deem the term G12 more appropriate as in some states *Gymnasium* starts after six years of primary school and, hence, only takes 6-7 years. Conversely, we use the term G13 to refer to the old regime of graduation after 13 years.

grades are seemingly unaffected by the reform. We find some evidence for a decrease of the reform effect on grade repetition over time, although the fading out is slow and far from being complete. Though grade repetition rates increase, we do not find evidence for changes in the overall number of students graduating from high school. Increases in grade repetitions do not translate into more school drop-outs. Generally, we find that boys are more affected by the reform: Due to the reform they repeat a grade more often than females, which translates also into a smaller reduction in the graduation age.

The remainder of this study is structured as follows. Section 2 discusses related studies and section 3 provides additional information regarding the German education system and the G12 reform. Then, we introduce the data (section 4) and the empirical strategy (section 5). In section 6, we report the average reform effects and present a broad range of robustness tests in section 7. Section 8 investigates heterogeneities of the reform effect. Section 9 discusses the findings and concludes.

2. Related literature

Among economists it is beyond question that crucial determinants of labour market outcomes are shaped during childhood and the time at school. Card (1999) summarised convincing evidence for a causal impact of education on labour earnings. An extensive body of literature has aimed at exploring the efficient use of scarce public resources in the formation of human capital through education.

Germany's G12 education reform addresses a central concern of policy makers in the design of education systems: G12 compresses the number of years spent in school, while keeping constant overall instruction time. Related policies are those that increase minimum compulsory schooling, which have widely been used as instrumental variables for years of education in numerous studies (see e.g. Card, 1999, Oreopoulos

et al., 2006, Pischke and von Wachter, 2008, Carneiro et al., 2013, Brunello et al., 2013). Other related studies investigate reductions in the years of education. For instance, Webbink (2007) shows that a reduction of Dutch university duration from five to four years decreased wages by 7-9 %. Similarly, Morin (2013) and Krashinsky (2014) find that a reduction of the high school track length in the Canadian province of Ontario from five to four years resulted in a significantly worse educational performance of the affected cohorts. Yet, the G12 reform differs fundamentally from both, the increases in compulsory schooling and the shortening of further education, as G12 does not change total instruction time.[2]

More similar to the G12 reform is a reform in Germany that aimed at harmonising the nation wide school year in 1966-67. Institutionally, this was realised through the introduction of two short school years. The short school years also increased learning intensity as the material for two full school years was taught in a shorter period of time and the overall curriculum was left unchanged with only minor reductions in requirements. Pischke (2007) examined short-run and long-run effects of this reform and finds that the increase in learning intensity increases grade repetition rates and decreases the number of pupils enrolled in higher secondary school tracks. He does not find evidence for long-run effects of later labour market outcomes. Several other studies examine the effect of term length on achievement and labour market outcomes and typically find insignificant effects (Grogger, 1996, Eide and Showalter, 1998, Card and Krueger, 1992).

In the evaluation of the G12 reform we are looking at, first efforts have been made regarding school achievements and post-schooling decisions. Economists have

[2]That is also the reason why we deem the G12 reform inappropriate as an instrument for education. It remains unclear what the instrument picks up, fewer years of education or increased learning intensity.

gained first insights from a survey providing information of 14 schools in two cities in the federal state of Saxony-Anhalt, which first introduced the G12 reform. Büttner and Thomsen (2010) find that the G12 reform reduces final examination scores in Mathematics for both genders and in English for females. In their analysis of the reform effect on the development of non-cognitive skills, Büttner et al. (2011) find no impact. Using the same set of data, Meyer and Thomsen (2012) identifies a decrease in university enrolment among females.

Despite their valuable contributions, these studies have several limitations due to data restrictions. First, potential reform effects cannot be distinguished from general time trends as only a single state is considered. Second, some of the findings could simply emerge from age effects. Third, if the policy change actually increases grade repetition, these studies cannot account for changes in the sample composition. Fourth, these studies compare the double graduation cohorts, i.e. the last cohort under the old regime to the first cohort under the new regime. Especially the first G12 cohort in Saxony-Anhalt was exposed to a policy surprise, as they were informed in grade 9 that they will graduate one year earlier. For them, there was less time left to distribute the curriculum over the remaining years. Furthermore, incentives and mental pressure might have been different for pupils in this cohort as they directly competed with the older cohort for limited resources (e.g. university places). Therefore, it is questionable whether findings for the first G12 cohort can be generalized to later treatment cohorts.

A recent study by Dahmann and Anger (2014) accounts for this problem and is similar to our study in terms of the applied identification strategy. Using data on about 200 G12 pupils and more than 1000 non-treatment students from the Socio-Economic Panel Study (SOEP), they employ a similar difference-in-differences method and find that the G12 reform had some effects on specific personality traits.

3. Institutional background

This section provides some background on the institutional settings of the German education system and introduces the G12 reform in more detail.

3.1. The German school system

Generally, education policy in Germany is a matter for the federal states. Still, the education system exhibits many similarities across states. All states have in common that schooling starts with primary school, in which pupils of all ability types are taught jointly. Pupils enter primary school at age six and stay for either four or six years, depending on the state.[3] After primary school, pupils are tracked into different school forms. The school tracking system intends to best support pupils of different ability types.

Across all states, the high school track *Gymnasium* intends to prepare pupils for university education. This is the only track that is directly affected by the G12 reform we are examining.[4] Currently, about 40% of a cohort enter *Gymnasium* after primary school. The decision about the school track chosen after primary school depends on state regulations. While in some states, the admission to *Gymnasiums* requires the primary teacher's recommendation, other states leave this decision to the parents and the admission by the respective *Gymnasium*. The decision about which school of a certain track is visited depends on the catchment area of the place

[3]State specific regulations can also allow children to enter primary school between age five and seven.

[4]While the *Gymnasium* exists in all states, the structure of alternative tracks, which address pupils with lower abilities, exhibit great variations across federal states. Some states further segregate pupils by there ability types in different school forms, while others teach them together. Students can earn different degrees though the *Gymnasium* is the main track to earn the general university entrance qualification *Abitur*. Every state provides an option next to the *Gymnasium* in which pupils can obtain entrance qualification to technical colleges. This degree does not directly qualify for university education.

of residence. Generally, only in case of well justified requests, parents can apply for schools in other catchment areas.

Once pupils enter *Gymnasium*, they will be prepared for the school leaving certificate *Abitur*, constituting the general university entrance qualification. Students that hold the *Abitur* are allowed to study at any higher education institution in Germany. However, the grade point average (GPA) of the *Abitur* is important for most universities in the admission process to certain subjects that are equipped with study quotas.[5] The final two years at *Gymnasium* constitute the *qualification phase* and only these last two years count towards the final grade point average. These last two years differ from the others in that pupils can drop certain courses and specialise in others.[6]

3.2. The G12 reform shortens high school track tenure

In international comparisons, German pupils were found to enter the labour market at comparably high ages (OECD, 2013) A reduction of the high school track tenure has been suggested such that pupils can accomplish the *Abitur* after 12 years of schooling instead of 13. The policy objective has primarily been to reduce the age at which students can graduate from high school in order to (i) accelerate students' labour market access, (ii) improve their labour market competitiveness, (iii) increase the labour force in reply to shortages of qualified workers and (iv) remedy shortages in the public pension scheme.

Since 2001, this reform has been gradually implemented in German states. Ta-

[5]By August 2013, 68 per cent of regular Bachelor studies at the 20 largest German universities have been restricted to certain GPA averages (Osel and Weiss, 2013)

[6]The *introductory phase*, the last year before the qualification phase is a particularity as it combines elements of the qualification phase (elective courses) and of the first years (grades do not count for final GPA).

ble 1 provides an overview of the timing of policy implementation across different states. The first G12 cohort graduated in 2007 in Saxony-Anhalt. The states of Mecklenburg-Vorpommern, Saarland and Hamburg were next to follow. By 2013, pupils can graduate from high school after 12 years of schooling in 14 out of 16 federal states. The graduation year of the first G12 cohort is at the same time the graduation year of the last pre-treatment cohort. We exclude these double graduation cohorts in most of our analyses as it is difficult to distinguish between them in the data.

While the reform reduces the overall tenure of school, it leaves the minimum required instruction time as well as the amount of holiday unchanged. From grade level 5 onwards, at least 265 hours per week have to be distributed over the remaining 8 rather than 9 years. These additional instruction hours have been differently distributed over the remaining years in different states. The average change in instruction hours per grade level across the states is plotted in figure 1.[7] This information is available for grade 7 onwards only. Where high school already starts in grade 5, not much of the workload has been shifted to these lower grade levels in order to prevent these young pupils from too much workload. Figure 1 reveals that grades 8-10 experienced the highest increase in additional instruction hours, while grades 11-12 experienced more modest increases.

The the final two years before graduation constitute the qualification phase at German high schools. These two years are distinct from the others, which has been preserved under the G12 reform. Now, however, the qualification phase Q1 corresponds to grade 11, rather than 12 under the old regime, and the qualification phase

[7] A uniform distribution of hours across the years would increase the weekly number of instruction hours from $\frac{265}{9} \approx 29.4$ to $\frac{265}{8} \approx 33.1$.

Q2 corresponds to grade 12 rather than 13. This change is depicted in figure 2. In the analysis of the G12 effect on grade repetitions, we distinguish the effect separately for each grade level, accounting for the shift in the qualification phase.[8]

Across all states, pupils that fail to fulfil performance requirements of a grade level have to repeat the same grade level or have to change into a lower school track, alternatively. Next to these performance based criteria, pupils can also voluntarily repeat a grade level once. In the data and in reality, one can hardly distinguish between these two motives. Students who do want to fail a grade might simply refuse to perform sufficiently well. Whatever the reason is for repeating a grade, any grade level repetition induced by the reform hampers the overall gain in time of the reform.

4. Data

General data. Throughout our analyses, we employ administrative data from the Federal Statistical Office.[9] The data cover the universe of all pupils in Germany and contain the relevant information aggregated by year, gender, school type and federal state. The data have not been used much by economists, Pischke (2007) being among the exceptions. We do not consider the federal state of Hesse in our main analysis, as this state introduced the shortening of secondary schooling gradually over a period of three years, which can not be distinguished in the data. Further, to make results comparable and consistent in terms of the sample selection, we drop the federal state of Lower Saxony as this state does not provide information on grade repetitions for the three final years at high school. Our main analysis sample covers the years 2002

[8]This requires to drop grade 11 of the old regime from our main analyses. In the robustness checks, we abstract from the institutional setting and drop grade 13 of the old regime instead.

[9]More specifically, we use data from the *Fachserie 11, Reihe 1 - Allgemein bildende Schulen.*

through 2012 for the three outcomes. Where available, we will use additional years for sensitivity analyses.

The data set has three main advantages. First, it is a full population survey. Second, information about graduation and grade repetitions are not self-reported by the individuals. Hence, individual non-response and social desirability bias are not an issue here. Third, the quality of the data can be regarded as high as the schools are by law required to provide the respective information.[10]

However, the data also have some shortcomings. Generally, there are no socio-economic background variables available in the data. This does not hamper the estimation of the average reform effect. However, it prevents from investigating effect heterogeneities across children's socio-economic background and from investigating changes in group compositions. Second, it is not possible to link individual pupils over time.

We will now comment on the specific data that is used for each of the three outcomes.

Graduation age. For each cohort of *Abitur* graduates, the Federal Statistical Office provides information about the distribution of the graduates' birth years. This data is not immediately available to the public and has been delivered electronically on request. From this information, we calculate the mean graduation age for each state and each graduation year. The data set consists of 142 observations from 14 states across 11 years, where we have to exclude 3 missing observations for the year 2002

[10]The schools provide the information to the statistical offices of the federal states. Then, the Federal Statistical Office harmonizes these state level information and makes them publicly available.

for Saxony Anhalt, Mecklenburg-Vorpommern and North Rhine-Westphalia.[11] We further drop the double graduation cohorts as we cannot distinguished between last pre-treatment observations and first treatment observations in our data (9 observations). The remaining 131 graduation year-state observations contain information from more than 1.7 million students.

Grade repetitions. The data on grade repetition rates is also provided by the Federal Statistical Office and provides the additional advantage of disaggregation at the grade level, resulting in more variation in the data. In each year, in each state the number of students who repeated a specific grade at high school is provided. However, the nature of its recording introduces a source of potential measurement error. At the beginning of the new school year (usually in September), it is recorded how many pupils repeated the respective grade level. The vast majority of pupils who did not pass a grade repeats the grade at the same high school. Pupils who do not repeat the grade at the same high school can repeat it at a different high school in the same state. This does not introduce measurement error in our sample (as our data is aggregated at the state level). However, pupils that repeat a grade level at a lower school type in the same state, at a high school in another state or at a lower school type in another state are potential sources of measurement error. Also, measurement errors are introduced if pupils leave the German school system completely instead of repeating a grade.

If this kind of measurement error in the dependent variable varies randomly with the introduction of the policy reform, our estimates will not be biased; only

[11]Prior to 2002, Saxony-Anhalt and Mecklenburg-Vorpommern experienced the G12 regime already and re-introduced high school graduation after 13 years. Consequently, the two states do not have a 2002 graduation cohort. 2002 information on North Rhine-Wesphalia is missing.

the standard errors will increase. If this measurement error correlates with the introduction of shortening secondary schooling, this will bias the results. We argue that this will most likely result in an underestimation of the reform's effect on grade repetitions: If the policy change increases grade repetition rates in general and also increases the ratio of repeaters who do not repeat the grade in the same state and the same school track, this will result in a downward bias of the overall reform effect on grade repetition rates.

We drop double graduation cohorts in each treatment state as we cannot distinguish grade repeaters in the treatment group from grade repeaters in the control group for these cohorts. Hence, overall our sample for the period 2002-2012 consists of 826 state-year-grade observations.[12] These 826 state-year-grade observations contain information from almost 13 million student-year observations.

Abitur graduates. The Federal Statistical Office also reports the total number of students that obtain their general university entrance qualification *Abitur* from high school by year and state. As the number of high school graduates depends heavily on the size of the respective birth cohorts, we standardise it.[13] For this purpose, we use information on the state's number of *Abitur*-aged individuals and divide the number of graduates by the average cohort size of 18-20 year old living in a specific state in a specific year. In robustness analyses, we experiment with different ways of standardisations which does not affect our conclusion (e.g. number of individuals of the same cohort when they were in grade 7). For this outcome, the number of

[12]From the original sample, we drop the information about the double graduation cohorts of nine federal states in six grades, and the double graduation cohort information of one state in one grade (we do not observe the double graduation cohorts for North Rhine-Westphalia in the two last years before graduation (2013)). Hence, our sample consists of 826 observations.

[13]This standardisation is not only relevant due large difference between the size of the states, but also because the number of births dropped sharply in East Germany after reunification.

observations reduces from potentially 154 (14 states and 11 years) to 139, as (a) 9 double graduation cohorts are dropped for which we cannot distinguish the treatment status among the total number of graduates, (b) there are no graduates in Saxony-Anhalt and Mecklenburg Vorpommern in 2002, and (c) there are no observations for Baden-Württemberg for 2002-2005. The 139 state-year observations comprise information on almost 2 million graduates.

5. Empirical strategy

Our empirical strategy makes use of the institutional peculiarity of statehood in educational affairs. This state sovereignty leads to a quasi-experimental setting in which time and regional variations allow for estimation of difference-in-differences type regressions. We estimate the effect of shortening secondary schooling on mean graduation age and *Abitur* graduation rates as dependent variables y for state s at time t with the following model:

$$y_{st} = \beta \cdot G12_{st} + \delta_s + \kappa_t + X'_{st} \cdot \lambda + \varepsilon_{st}, \tag{1}$$

$G12_{st}$ is a binary variable that identifies the treatment status of state s at time t. β is the coefficient of core interest and identifies the reform effect. δ_s, a set of state dummy variables, controls for general outcome differences between different states. κ_t refers to a set of indicator variables for each year, thereby taking into account general changes in the outcome variables over time. X_{st} denotes a number of time varying control variables, which take into account changes in the state's economic and demographic situation. It includes the state's GDP growth, the state's general unemployment rate and the state's youth unemployment rate as measures of the economic situation. Furthermore, in order to monitor changes in the population

composition, X_{st} includes the state's population density and the state's share of teenagers in the state's total population. We allow for correlations of the error term ε_{st} within states and provide standard errors that are clustered at the state level.

For mean graduation age and overall graduation rates, we only have one observation per state and year. For grade repetition rates, the data is additionally disaggregated at the grade level. In order to benefit from this variation, we adjust the model in the following way:

$$y_{gst} = (G12_{gst} \cdot grade_g)' \cdot \beta + \gamma_g + \delta_s + \kappa_t + \mu_{gs} + \nu_{gt} + X'_{st}\lambda + \varepsilon_{gst}, \qquad (2)$$

where y_{gst} is the fraction of pupils repeating the grade level g in state s at time t. As the reform effect might differ between grade levels and particularly between the qualification phase (last two years before graduation) and the earlier years, we estimate the treatment separately for each grade level. We interact the $G12$ indicator with dummies $grade_g$ for each grade level from 7-10 and the qualification phase Q1 and Q2 such that β is now a vector comprising of the reform effect estimate for each grade level. δ_s, κ_t and X_{st} are defined as before. We further include a set of grade-fixed effects, captured by γ_g. This takes into account general differences in repetition rates between grades. μ_{gs} indicates a set of binary variables for each grade-state combination. This set of control variables captures grade-specific outcome differences that differ between the states. For instance, passing a specific grade might be more difficult in one state than in others, even when general differences in the passing probability between the states are taken into account. Similarly, ν_{gt} controls for each grade-time interaction using binary variables.

6. Results

This section reports our estimation results of the G12 reform effect on (i) the mean age at which students graduate with *Abitur*, (ii) on high school grade repetition rates and (iii) on *Abitur* graduation rates.

6.1. Abitur graduation age

The reform's principal objective was to reduce the age at which students graduate from school. This section first investigates whether the reform has been effective in this regard. A first descriptive inspection of figure 3 reveals that graduation age has dropped on average by 0.9 years from about 19.7 to about 18.8 years in the treatment states.

The regression results obtained from estimating equation (1) support this finding (see table 2). The baseline DiD specification suggests that the reform reduces the mean graduation age by 0.82 years, about 10 months. In model (2), we additionally control for factors that characterise the economic environment in which students obtain their education. We include the GDP growth of the federal state, the general and youth unemployment rate, the population density, and the share of *Gymnasium*-aged children in the total population. This specification yields a point estimate of -0.86. This is a noticeable decrease due to the reform, but there is strong statistical evidence that the reform effect reduces the mean graduation age by less than one year. Thereby, the reform stays behind its potential of one year.

Why does the reform not unfold its full potential? One possible explanation is that grade repetition rates increase because of the G12 reform. This potential mechanism will be inspected next.

6.2. Grade repetition rates

There are several reasons to believe that shortening high school tenure has an additional impact on both strategic and involuntary grade repetitions. First, students experience a higher learning intensity because of the G12 reform. This could deteriorate pupils' capability of mastering the material. If this motive was the only reason, the effect is expected to be strongest in the grade levels that receive the highest increase in learning intensity (grade 8-10, see figure 1). However, students may also want to repeat a grade level for strategic reasons. Students might want to trade the gained one year for a better performance in school which then potentially improves their grade point average for university applications. Alternatively, they may want to use this additional year to decide about their post-schooling decisions. If this is the main driving force, the effect is expected to be strongest in the final two years where grades count towards the final grade point average of the *Abitur*. Both motives are captured by our data and would lead to a reduction in the overall effectiveness of the reform in reducing graduation age.

Because of these different motives, we analyse grade repetition rates separately for each grade level as there are various reasons to assume that the reform effect might differ across grade levels.

Before turning to the estimation results, in figure 4 we present a descriptive illustration of mean grade repetition rates in each grade across all treatment states two years before and two years after the introduction of the G12 reform. It immediately appears that grade repetition rates have changed in the qualification phase (grades 12 and 13 under the old regime, grades 11 and 12 under the new regime) after the implementation of the G12 reform. The change in grades 7-9 is less apparent.

Furthermore, figure 5 plots grade repetition rates for qualification phase Q1 over time. For illustration, we consider the states of Saxony-Anhalt (ST), Mecklenburg-

Vorpommern (MV), Saarland (SL), Hamburg (HH) and Bavaria (BY) as treatment group, whereas all pre-treatment observations across the remaining states constitute the control group. In the period before the gradual introduction of the G12 reform, the time trend of the treatment and the control group appears to be parallel. When the reform is gradually introduced in the treatment states, the difference in grade repetition rates between both groups increases noticeably.

Table 3 reports the results of the difference-in-differences model from equation (2), separately for each grade level. Our first model in column (1) of table 3 accounts for state, year and grade level variations as well as state specific and time specific grade level variations. There are no significant effects on grade repetition rates in grade 7-9, whereas there are significant effects in the qualification phase that amount to more than a doubling of grade repetition rates. The specification in column (2) includes further the aforementioned time-varying control variables capturing changes in the states' economic and demographic characteristics. This is our preferred specification. Accordingly, grade repetition rates are unaffected in grade 7 to 9, then increase by 1.32 percentage points, or about 70%, in grade 10. The strongest effect in absolute terms can be observed in grade Q1 (two years before graduation) where the repetition rates increase by 3.39 percentage points. Also in the final year, the effect of 1.28 percentage points is highly significant and amounts to an increase by more than 100 %.

As outlined, the reform effect on grade repetitions could arise through at least two channels: involuntary and strategic grade repetitions. Notice that the effect on grade repetition rates does not correspond one-to-one to the shift in work load. In the discussion, we collect some arguments that favour the one or the other motive.

The higher rate of grade repetitions in the last high school years is a plausible explanation why the reform's potential of reducing high school graduation age has

not been fully exploited.

6.3. Abitur graduates

This section investigates whether the reform has an impact on society's human capital stock, measured by the share of *Abitur* graduates. There are several channels through which the G12 reform could impact *Abitur* graduation rates. On the one hand, an increase in the weekly school instruction time could cause problems for some students in coping with the material. This would increase the disutility from schooling and favour drop outs or the attendance of a lower school track, which would also reduce *Abitur* rates. On the other hand, a shortening of the number of years required to get awarded the same educational degree, reduces the opportunity costs of schooling, which could justify an increase in *Abitur* rates.

Furthermore, we find that due to the G12 reform grade repetition rates increase in the final years at high school and several other studies relate grade repetitions and graduation rates. However, there is no general agreement in the literature as to whether grade repetitions increase or decrease drop-out rates. While Manacorda (2012) shows that grade repetitions increase drop-out rates, Eide and Showalter (2001) do not obtain statistically significant effects of grade repetitions on school drop-outs. Jacob and Lefgren (2009) suggest that the effects of grade repetitions might differ between the grades: They show that grade repetition among older pupils increases drop-out rates, while grade repetitions among younger pupils do not change high school completion rates. The expected direction of the G12 effect on *Abitur* graduates remains unclear.

In table 4, we analyse the impact of the G12 reform on the share of *Abitur* graduates in the state's population aged 18 to 20. The baseline DiD specification in column (1) yields negative estimates that are not only statistically insignificant

but also close to zero. The estimated effect is also small and insignificant when we control for economic and demographic changes (column 2). Hence, we conclude that there is no significant effect of shortening high school tenure on *Abitur* graduation rates.

7. Robustness analysis

In this section, we perform various sensitivity analyses of the results in the previous section. We apply the same set of robustness checks for all outcome variables and perform further robustness tests that are outcome-specific. Table 5 summarises the results for different model specifications, sample restrictions, estimation issues and placebo policy reforms.

The identifying assumption in these quasi-experimental analyses - that of a common time trend of the outcome variable between treatment and control states in the absence of treatment - is untestable. The first set of robustness checks investigates potential threats to the validity of this identification assumption. In column (2) of table 5, we add linear state-specific time trends to check whether the estimated reform effect is driven by generally differing time trends between the states.

Another threat to our identification strategy are co-treatments, i.e. other policies that have been implemented around the same time and that might also affect graduation age, grade repetition rates or the share of graduates. Mainly following the weak performance of German pupils in the PISA 2001 comparison, several other educational reforms have been implemented during our analysis period:

- Centralised school leaving examinations: Some states allowed for the design of school leaving examinations at the school level, which were replaced by

centralised school leaving examinations leading to the same final examination questions for all pupils in that state.

- Reduction in subject choice options: Several states implemented policies that reduce pupils' subject choice options and predetermined in which subjects final examinations have to be passed.

Table A.1 in the appendix provides an overview over these reforms and indicates the first affected cohort in each state. Similarly to G12, these policies were implemented at different times in different states. As none of the other policy changes is perfectly collinear to the G12 reform, there is enough variation in the data to distinguish the G12 effect from other policy changes.

Another crucial assumption of our difference-in-differences estimation strategy is the exclusion of cross-border migration induced by the reform, and therewith a change in the treatment and control group. Only individuals living close to a border can change the state without moving homes - when they are willing to go through the bureaucratic process of changing to another school district. Moving to the school system in a different state is linked to further obstacles due to state-varying formal admission requirements. Compared to the entirety of German students captured by our data, we consider this source of selective migration minor. Entire moves from treatment to non-treatment states are related to considerable costs, such as movement costs, the change of job of the parents and abandoning the social environment. The concerned share seems rather negligible. In their micro-level based G12 study, Dahmann and Anger (2014) find no such signs of selective migration.[14]

[14]If there is selective migration from treatment to control states, which is independent of students' ability, the effect on mean graduation age and grade repetition rates would not be affected. The reform effect on graduation rates would be downward biased (negative effects are exaggerated,

The reform effects could further be distorted due to anticipatory behaviour of students. For example, a student may not want to repeat a grade in the penultimate cohort of the old regime, as she would then belong to the double graduation cohort which induces several drawbacks for post-schooling decisions. These cohorts were of double size and they were prevalently competing for places in vocational trainings or volunteering activities or for resources at universities (e.g. available places, pupil-teacher-ratio). In the specification in column (4), we include two dummy variables indicating whether a certain cohort is the last or penultimate cohort of the old regime through which we account for anticipatory effects.

In column (5), we extend the sample period to 1995-2002 for grade repetition rates in order to check for the sensitivity of our results with respect to a different sampling period. Column (6) excludes the first three treatment cohorts in Saxony-Anhalt and Mecklenburg-Vorpommern from the sample as they were already in grade 7-9 when they were informed that they can finish high school one year earlier. For these surprised, exceptional cohorts, the additional workload has been distributed over fewer remaining years.

While in our main specification each observation has the same weight, in column (7) and (8) we account for the fact that our data comprises of averages generated from the entirety of German high school students. Column (7) weights the observations by the respective number of pupils per year and state (and grade level, where possible) that generated this average. This assigns a higher weight to states with a larger number of pupils. Column (8) uses frequency weights and accounts for the fact that the data is an aggregate of all students. This model performs the analysis

positive effects attenuated). When low ability students migrate to non-treatment states (in order to avoid the higher learning intensity of the G12 regime), the effect on mean graduation age and grade repetition rates is downward biased.

pretending we have data at the individual level which naturally increases the number of observations while leaving the number of clusters (states) unchanged.

We also run a series of placebo reform tests to further check for potential violations of the crucial common time trend assumption between treatment and control states. We pretend that the policy change took place 2 to 4 years before the actual reform and analyse the effects of these placebo treatments.[15] For these analyses, we only consider G13 observations and generate a new treatment variable with the respective timing. In order to have enough pre-treatment observations the effect on grade repetitions in the lower grades, for grade repetitions, we use the extended sample 1995-2012.

Graduation age. The effect of the G12 reform on graduation rates is strikingly robust: Across all specifications on the real policy reform effect, the estimates prove to be significantly smaller than zero and significantly larger than -1, reinforcing the picture that the reform does not fully meet its potential. The estimate is closest to -1 when surprised cohorts are excluded. Simultaneously, this reduces the point estimates of the effect on grade repetition rates, strengthening the argument that grade repetitions are one channel through which the reform's effect on the reduction in graduation age is attenuated.

Grade repetition rates. The picture that we gained in our main analysis proves to be robust across the various model specifications. The largest decline in the reform effect on grade repetitions rates in the qualification phase can be observed - as we would expect - when surprised cohorts are excluded from the sample. Still the effect equals more than a doubling of grade repetition rates due to the G12 reform. This

[15]We do not present results for a placebo policy change one year earlier as this does not only leave few post-treatment observations, but also features the risk to capture the described effects of anticipatory behaviour.

pattern is consistent with the explanations of higher grade repetition rates because of increased workload. The placebo regressions (columns 9-11) provide support for the identification assumption that treated states show a similar time trend in grade repetition rates as control states under the old regime.

Table 6 reports the results of further robustness checks that are specific to grade repetition rates. Column (1) includes state-grade time trends, which leaves the general picture unchanged. In column (2), we ignore institutional characteristics and compare grade 11 and 12 in the old regime with grade 11 and 12 in the new regime. Observations on grade 13 rather than grade 11 are now dropped from the control group sample. The estimates are very similar to the visual illustration in figure 4. Grade repetitions generally appear to be higher one year prior graduation than in the final year. We now compare repetition rates of the pre-treatment cohorts one year prior graduation to the repetition rates in the final year of the treatment cohorts.[16] By the repetition rate profile across grades, this increases the effect size in qualification phase Q1, but leaves the effect insignificant and even negative in qualification phase Q2.

As grade repetitions are calculated at the beginning of the new school year, specification (3) of table 6 calculates grade repetitions based on the number of students in the previous school year from which the repeating students come. This accounts for varying cohort sizes. The pattern of effects is the same, though the point estimates slightly decrease. This indicates declining cohort sizes in treatment states.

In the difference-in-differences framework, the validity of the identification assumption also depends on the measurement scale of the outcome (see e.g. Lechner,

[16]In order to be consistent with this sample restriction, in this model we substitute the grade level dummies (and interaction thereof) by school year dummies.

2010). In order to examine the sensitivity of our results to a transformation of the outcome variable, in column (4) we take the logarithm of grade repetition rates as the dependent variable. Though the point estimates are no longer comparable, also in this specification, the G12 effect is significant for the final three years, whereas it remains insignificant for lower grade levels.

Finally, we investigate the reform effect on basic and medium track schools as a placebo reform outcome. Although the reform addressed only the high school track directly and although students cannot earn the general university entrance qualification in basic and medium track schools (see section 3), the G12 reform might still have an indirect impact on grade repetition rates at these two school types. By the nature of our data, it is possible that some of the individuals that are counted as grade repeaters at basic or medium schools come from treated *Gymnasiums*. As it turns out in column (5), the effects are very imprecisely estimated and there is no statistical evidence for an impact of the G12 reform on grade repetition rates in other school forms.

Abitur graduates. Throughout the different specifications, in table A.1 we do not find evidence for a statistically significant effect of the G12 reform on *Abitur* graduation rates. Even the sign of the estimates varies between the specifications. As for the other outcomes, the placebo regressions do not point to different pre-treatment trends in graduation rates between treated and untreated states.

Table 7 reports results for different definitions of *Abitur* graduation rates. While in our preferred specification we standardise the number of high school graduates with the average cohort size of the 18-20 year old, for standardisation in table 7 we instead use (i) the average cohort size of the 18-19 year old (column 2), (ii) the graduation cohort's size in high school when the cohort was in grade 7 (column 3), and (iii) the

graduation cohort's size in all schools when the cohort was in grade 7 (column 4). Additionally, column (5) reports the results for the natural logarithm of the total number of graduates without any standardisation. We do not find any statistical evidence for a significant effect of the G12 policy change on overall graduation rates.

8. Effect heterogeneities

This section separates the average G12 effect in our sample by gender and over time after the implementation of the reform to inspect whether average treatment effects mask effect heterogeneities.

8.1. By gender

Boys and girls differ in various aspects like genetic endowments, socialisation and stress resistance. It is therefore conceivable that there are gender differences in the responsiveness to the reform in terms of managing the increased hourly workload or the change in opportunity costs. The previously estimated average treatment effects could mask heterogeneous effects by gender.

Table 8 displays the estimated G12 effects separately for males and females in columns (2) and (3).[17] There is a slight difference in the point estimates of the reform impact on mean graduation age. While this effect is closer to -1 for females, the difference between females and males is statistically not significant.

For grade repetition rates we see that males exhibit higher absolute increases in grade repetition rates than females in the final three years at high school. Both are seemingly unaffected in grade 7-9. The relative reform effect is also stronger for males

[17]These specifications differ from the main specifications only with respect to the outcome variable. The new outcomes now only refer to females and males, respectively.

in the final two years at high school. This makes it plausible that the reduction in mean graduation age is more pronounced for females than for males.

Finally, table 8 shows that the absence of a statistically significant reform effect on overall graduation rates does not arise through heterogeneous treatment effects by gender. Neither the effect for males nor the effect for females is significantly different from zero.

8.2. Evolution of the treatment effect over time

In this section, we investigate whether the estimated G12 effects remain constant over time or whether they fall back to pre-treatment levels after some time. This question is of high policy relevance as it is crucial for policy-makers to understand differences between short-run and medium-/long-run effects when evaluating the success of this reform. Due to the reform's timeliness, it is not yet possible to investigate the long-term effects of the reform. We estimate effects separately over 5 years.

We will focus on the evolution of the treatment effect on grade repetitions because for grade repetitions we can make use of a longer post-treatment period: While the other two outcomes can be observed only after a cohort's graduation, grade repetitions occur before graduation. Furthermore, we find strong evidence that the mean graduation age does not decrease by a full year due to increased rates of grade repetitions. Hence, any trend in the reform's effect on grade repetitions is likely to translate into changes in the effect on mean graduation age.

We expect some levelling out of the effect on grade repetition rates for three reasons. First, teachers and education policy-makers might improve the curriculum under the new G12 regime, so that it will be easier for pupils to follow the course material. Second, pupils and teachers might improve learning and teaching strategies, respectively. Third, pupils under the G12 regime might be more prone to repeat a

grade voluntarily, when the last G13 cohorts are not far away as repeating a grade feels less like the loss of a year.

For the sake of clarity, we will focus on the effect in the qualification phase Q1. We suspect that differences in the treatment effect over time will be most apparent in this grade as for this grade we observe the largest effect. Additionally, for qualification phase Q1 we have more post-treament observations than for qualification phase Q2. In the appendix we also report the results for qualification phase Q2.

Table 9 displays the reform effect over time on grade repetition rates for qualification phase Q1.[18] Column (1) reports the results for all available observations. This specification does not provide evidence that the reform effect is fading out over time as the effect for five years after the reform is of similar magnitude as the effect for one and two years after the reform. However, since the policy change was implemented at different points in time in different states, each year effect is identified by a different number of states. For instance, while the effect five years after the reform is only identified by the two states that were the first to switch to G12, the effect two years after the reform is identified by five states. Hence, differences in the reform effect between states could drive the results in column (1).

Therefore, we balance the sample in the post-treatment period in columns (2) to (5).[19] The results of table 9 point to some decline of the G12 effect over time. Yet,

[18]For the specifications in table 9, we replace the overall G12 identifier in equation (1) by indicator variables for each year after the reform was implemented, i.e. the number of cohorts after the double graduation cohort.

[19]All pre-treatment observations as well as all observations from Saxony and Thuringia, which always had G12, remain in the sample. Yet, from the remaining observations under the G12 regime, column (2) only includes the first five years after the G12 introduction and only for states with at least five post-treatment observations for qualification phase Q1. G12 observations in states with fewer than five post-treatment observations are dropped from this sample. For instance, the treatment effects in column (2) are only identified by Saxony-Anhalt and Mecklenburg-Vorpommern. For the identification of the effect across four years in column (3), Saarland can be added to the

even 5 years after the implementation we do not find evidence for a complete washing out of the effect. For instance, for Saxony-Anhalt and Mecklenburg-Vorpommern (column 2), after 5 years, the G12 reform effect of 5.01 percentage points is still highly significant, down from an estimated effect of 7.37 percentage points in the first year. The other columns provide similar evidence for a decline of the effect over time but not for a quick fading out.[20] Table A.2 in the appendix confirms this notion for qualification phase Q2.

9. Discussion and conclusion

Using data from the German Federal Statistical Office covering the entirety of all pupils in Germany, we analyse the short and medium run impacts of the fundamental G12 education reform in Germany. This reform tackles the trade-off between the provision of higher levels of schooling and earlier labour market entries by increasing education efficiency: The time in high school is reduced by one year while the total number of instruction hours is left unchanged and distributed over the remaining school years.

Exploiting the sequential implementation of the reform in the German states, we study first indicators of the overall effectiveness of this reform by looking at three outcomes: (i) the mean age at which students graduate from high school, (ii) high school grade repetition rates and (iii) overall high school graduation rates. Difference-in-differences estimates reveal that the policy change reduces the graduation age by about 10 months. One possible explanation why it stays behind its potential of one

effect identifying states, as it provides four post-treatment observations. Columns (4) and (5) follow analogously.

[20] Comparing the treatment effects across columns in table 9 points to substantial differences in the reform effect across states. Especially, Saxony-Anhalt and Mecklenburg-Vorpommern exhibit high increases in grade repetition rates due to G12.

full year lies in increased grade repetition rates. We find a doubling of grade repetition rates in the final two years prior to graduation. Repetition rates in lower grades are seemingly unaffected by the reform. Though grade repetition rates increase and graduation age decreases, we do not find evidence for changes in overall graduation rates.

The estimates prove to be stable in a broad range of robustness checks and do not change when we control for other education policies that came into effect during our period of analysis. Furthermore, placebo regressions do not indicate differing pre-treatment trends in the three outcome variables. Inspections of treatment effect heterogeneity reveal that male pupils are more affected by the reform: Due to the reform they repeat a grade more often than females, which translates also into a slightly smaller reduction in the graduation age. We find some evidence for a decrease of the strong reform effect on grade repetition rates over time, although the fading out is slow and far from being complete.

Our results point out that shortening school tenure can decrease school leaving age without affecting the human capital endowment in terms of high school graduates. While the results do not identify an adverse effect of the reform on the state's human capital endowment in terms of high school graduates, we cannot comment on the reform effect on the quality of the human capital stock (e.g. cognitive and non-cognitive skills).

However, it is possible that the costs of the reform are borne at the individual level. Given the same amount of resources, grade repetitions increase the student-teacher ratio and thereby potentially harm other students' benefit from education (among others, see Bandiera et al., 2010, for recent contributions). Furthermore, it is unclear whether there are more costs to be borne by the individual due to involuntary grade repetitions. We cannot distinguish whether the identified effects

are induced by strategic or involuntary grade repetitions. It can be noticed, that the increase in grade repetition rates does not one-to-one match the shift in the work load across grade levels. Also, the increase is highest in the final two years where the incentive for strategic grade repetition is highest. On the other hand, we do not observe a fading out over time. The incentive for strategic grade repetitions decreases over time as potential employees get more used to younger high school graduates. The negative stigma of grade repetitions would prevail. This favours the view that students are struggling with the requirements. We believe that the increase in grade repetitions is a mixture of both motives.

The findings are also interesting for other analyses of the G12 reform, particularly for studies that rely on comparisons of the double graduation cohorts. Depending on the motive of grade repetitions, it is conceivable that low-ability students are selected out of the first treatment cohort. Considering test performance comparisons of the last pre-treatment and the first treatment cohort, even in the absence of treatment effects, such selection processes would naturally lead to an increase in average test scores for the G12 cohort. Upcoming evaluations of the G12 reform should be aware of the fact that G12 increases grade repetitions and that the student composition between the last pre-treatment and first treatment cohort might ultimately differ.

Our findings are also interesting for education policy makers and inform other industrialised countries that face similar challenges as Germany, such as skilled worker shortages and decreasing fertility rates.

References

Bandiera, O., Larcinese, V., and Rasul, I. (2010). Heterogeneous Class Size Effects: New Evidence from a Panel of University Students. *The Economic Journal*, 120(549):1365–1398.

Barro, R. (2001). Human Capital and Growth. *American Economic Review*, 91(2):12–17.

Brunello, G., Fabbri, D., and Fort, M. (2013). The Causal Effect of Education on Body Mass: Evidence from Europe. *Journal of Labor Economics*, 31(1):195–223.

Büttner, B., Thiel, H., and Thomsen, S. L. (2011). Variation of learning intensity in late adolescence and the impact on noncognitive skills. *ZEW Discussion Papers*, 11-007.

Büttner, B. and Thomsen, S. L. (2010). Are we spending too many years in school? Causal evidence of the impact of shortening secondary school duration. *ZEW Discussion Papers*, 10-011.

Card, D. (1999). The causal effect of education on earnings. *Handbook of labor economics*, 3:1801–1863.

Card, D. and Krueger, A. B. (1992). Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States. *Journal of Political Economy*, 100(1):1–40.

Carneiro, P., Meghir, C., and Parey, M. (2013). Maternal Education, Home Environments, and the Development of Children and Adolescents. *Journal of the European Economic Association*, 11:123–160.

Dahmann, S. and Anger, S. (2014). The Impact of Education on Personality: Evidence from a German High School Reform. *IZA Discussion Paper*, (8139).

Dee, T. S. (2004). Are there civic returns to education? *Journal of Public Economics*, 88(9-10):1697–1720.

Eide, E. and Showalter, M. H. (1998). The effect of school quality on student performance: A quantile regression approach. *Economics Letters*, 58(3):345–350.

Eide, E. R. and Showalter, M. H. (2001). The effect of grade retention on educational and labor market outcomes. *Economics of Education Review*, 20(6):563–576.

Grogger, J. (1996). Does School Quality Explain the Recent Black/White Wage Trend? *Journal of Labor Economics*, 14(2):231–53.

Grossman, M. (2006). Education and nonmarket outcomes. In Hanushek, E. A. and Welch, F., editors, *Handbook of the Economics of Education*, volume 1, pages 577–633. Elsevier, Amsterdam.

Jacob, B. A. and Lefgren, L. (2009). The Effect of Grade Retention on High School Completion. *American Economic Journal: Applied Economics*, 1(3):33–58.

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/Revue canadienne d'économique*, 47(1):70–97.

Lechner, M. (2010). The estimation of causal effects by difference-in-difference methods. *University of St. Gallen Department of Economics Working Paper Series*, 28.

Lleras-Muney, A. (2005). The relationship between education and adult mortality in the United States. *Review of Economic Studies*, 72(1):189–221.

Lochner, L. and Moretti, E. (2004). The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports. *American Economic Review*, 94(1):155–189.

Manacorda, M. (2012). The Cost of Grade Retention. *Review of Economics and Statistics*, 94(2):596–606.

Meyer, T. and Thomsen, S. L. (2012). How Important is Secondary School Duration for Post-school Education Decisions? Evidence from a Natural Experiment. *Diskussionspapiere der Wirtschaftswissenschaftlichen Fakultät, Universität Hannover*, 509.

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/Revue canadienne ...*, (613):1–36.

OECD (2013). *Education at a Glance 2013: OECD Indicators*. Education at a Glance. OECD Publishing.

Oreopoulos, P., Page, M. E., and Stevens, A. H. (2006). The intergenerational effects of compulsory schooling. *Journal of Labor Economics*, 24(4):729–760.

Osel, J. and Weiss, M. (2013). Numerus clausus wird zur Regel. *Süddeutsche Zeitung*.

Pischke, J.-S. (2007). The Impact of Length of the School Year on Student Performance and Earnings: Evidence From the German Short School Years. *Economic Journal*, 117(523):1216–1242.

Pischke, J.-S. and von Wachter, T. (2008). Zero returns to compulsory schooling in Germany: Evidence and interpretation. *Review of Economics and Statistics*, 90(3):592–598.

Skirbekk, V., Kohler, H.-P., and Prskawetz, A. (2004). Birth Month, School Graduation, and the Timing of Births and Marriages. *Demography*, 41(3):547–568.

Webbink, D. (2007). Returns to university education: Evidence from a Dutch institutional reform. *Economica*, 74(293):113–134.

Figures

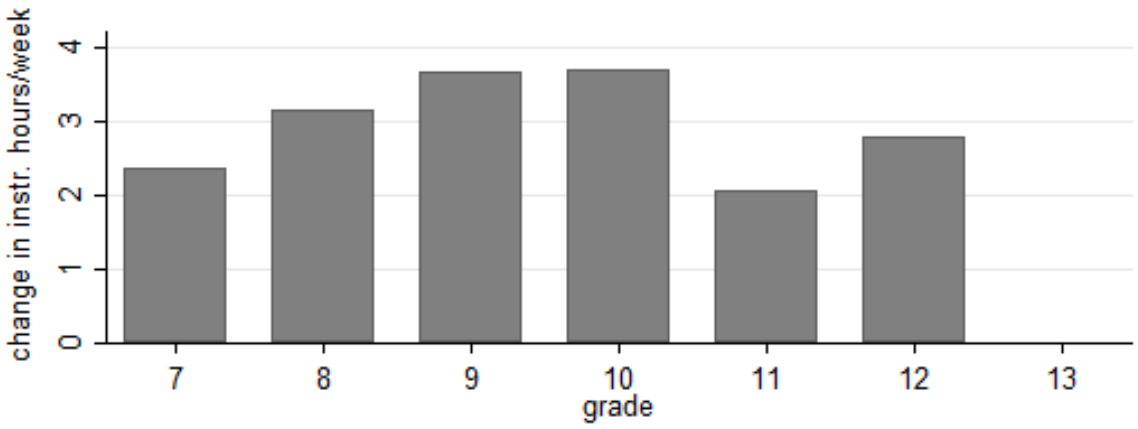

Figure 1: The bar chart plots the average change in the number of instruction hours per week and per grade level before and after the G12 reform. Grade 13 has been removed by the reform. Calculations are based on 561 state-grade-year combinations for 2002-2012, thereof 298 post treatment observations. Data has been provided by the *Kultusministerkonferenz* ('standing conference of the ministers of education and cultural affairs').

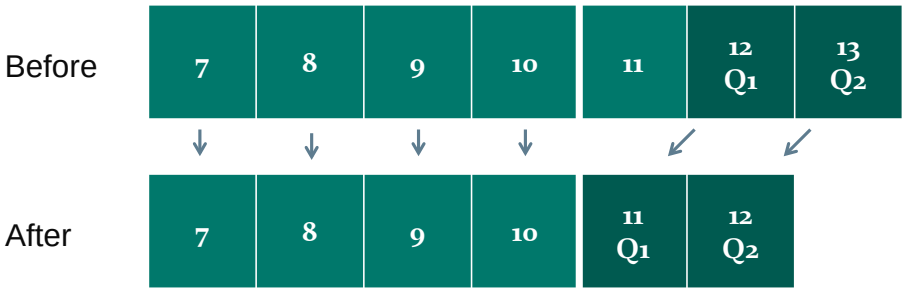

Figure 2: Comparison groups

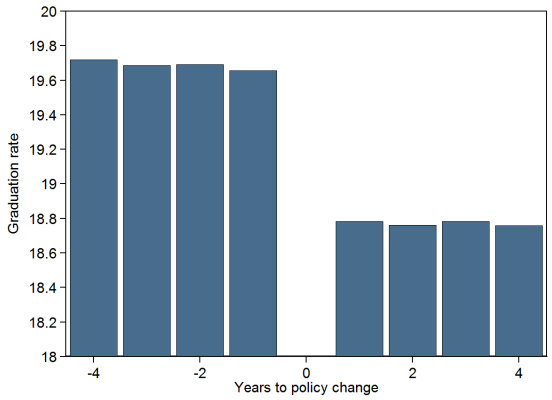

Figure 3: Mean *Abitur* graduation age - before and after the G12 reform. Each bar represents a cohort by its distance in years to the time of the policy implementation in their state.

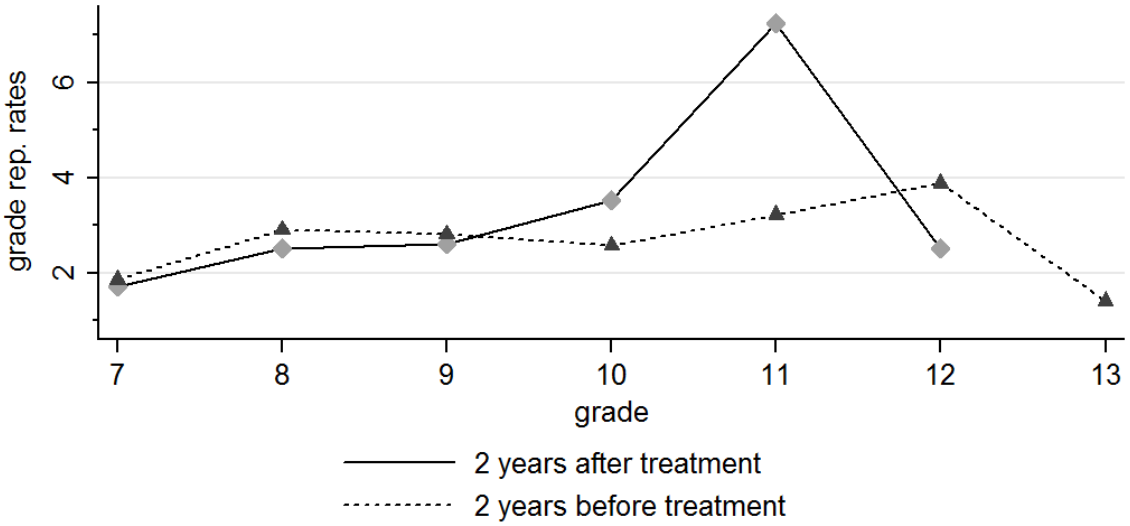

Figure 4: Mean grade repetition rates in treatment states separately by grades two years before and two years after the introduction of the G12 reform.

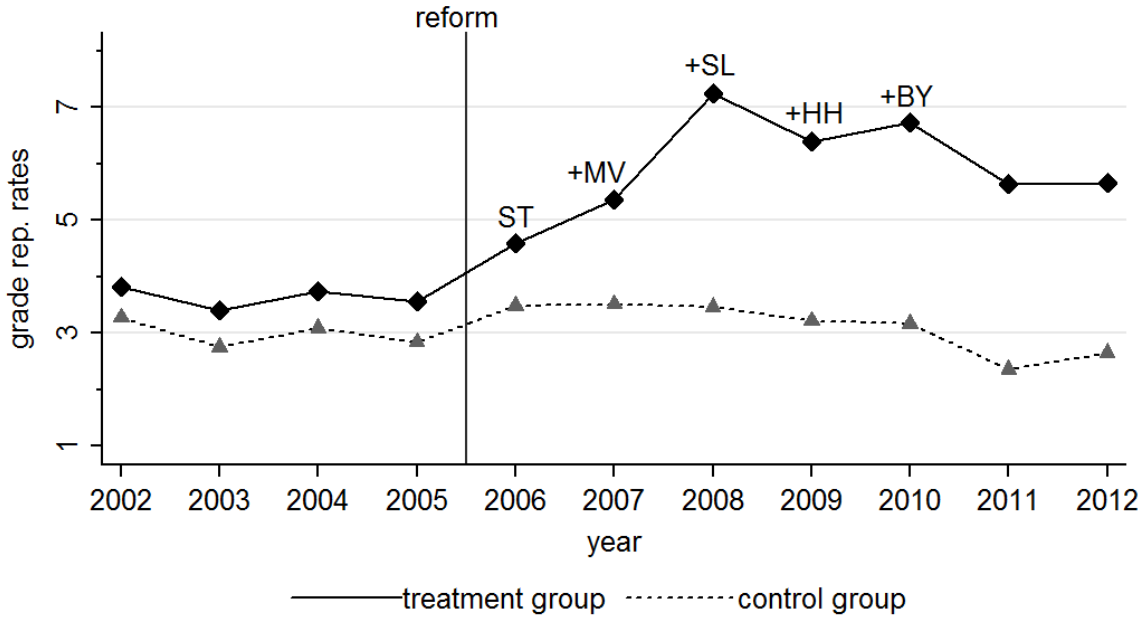

Figure 5: Mean grade repetition rates two years before graduation in treatment and control states over time. Here, considered treatment states are Saxony-Anhalt (ST), Mecklenburg-Vorpommern (MV), Saarland (SL), Hamburg (HH) and Bavaria (BY) that introduced the reform one after the other.

Tables

Table 1: Implementation of G12 in the federal states

State	First G12 graduates	Implementation
Change from G13 to G12		
Saxony-Anhalt	2007	2003
Mecklenburg-Vorpommern	2008	2004
Saarland	2009	2001
Hamburg	2010	2002
Bavaria	2011	2004
Baden-Württemberg	2012	2004
Bremen	2012	2004
Berlin	2012	2006
Brandenburg	2012	2006
North Rhine-Westphalia	2013	2005
Schleswig-Holstein	2016	2007
Always G12		
Saxony		
Thuringia		
Always G13		
Rhineland-Palatinate		
Excluded from sample		
Lower-Saxony	2011	2004
Hesse	2012-14	2004

Notes: The table specifies the timing of the G12 reform for each federal state by their announcement and by the first affected graduation cohort.

Table 2: Reform effect on high school graduation age

	baseline DiD (1)	main (2)
G12 reform effect	-0.82***	-0.86***
	(0.06)	(0.04)
Adj. R-Squared	0.98	0.99
N	142	142
State dummies	Yes	Yes
Year dummies	Yes	Yes
Time-varying controls	No	Yes

Notes: The table reports the effect of the G12 reform on mean age at graduation from high school for the years 2002-2012. Estimation is based on 14 states (excluding Hesse and Lower Saxony). Clustered standard errors are reported in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 3: Overall effects on grade repetitions

	baseline DiD (1)	main (2)
G12 · grade 7	-0.32	-0.14
	(0.29)	(0.29)
G12 · grade 8	0.12	0.30
	(0.26)	(0.28)
G12 · grade 9	0.01	0.06
	(0.37)	(0.34)
G12 · grade 10	1.40*	1.32*
	(0.67)	(0.70)
G12 · grade Q1	3.57***	3.39***
	(1.09)	(0.99)
G12 · grade Q2	1.51***	1.28**
	(0.38)	(0.44)
Adj. R-Squared	0.78	0.80
N	826	826
%-change grade 7	-16.98	-8.18
%-change grade 8	5.35	14.41
%-change grade 9	0.57	2.51
%-change grade 10	77.70	69.94
%-change grade Q1	125.82	112.40
%-change grade Q2	182.82	120.38
State dummies	Yes	Yes
Year dummies	Yes	Yes
Grade level dummies	Yes	Yes
State * grade level	Yes	Yes
Time * grade level	Yes	Yes
Time-varying controls	No	Yes

Notes: The table reports the effect of the G12 reform on the share of grade repetitions (in %) from OLS regressions. Clustered standard errors are reported in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 4: Overall effects on share of *Abitur* graduates in 18-20 years old pop

	baseline DiD (1)	main (2)
G12 reform effect	-0.69	-0.42
	(1.56)	(1.61)
Adj. R-Squared	0.75	0.77
N	139	139
%-change	-2.64	-1.64
State dummies	Yes	Yes
Year dummies	Yes	Yes
Time-varying controls	No	Yes

Notes: The table reports the effect of the G12 reform on the share of grade repetitions (in %) from OLS regressions. Parantheses contain standard errors clustered by federal states. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 5: Robustness checks

		additional controls			sample restrictions		estimation issues		placebo policy change[a]		
	main	time trend	further policies	anticip. effects	extended period	w/o surpr. cohorts	analy. weights	frequ. weights	$t-2$	$t-3$	$t-4$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel A: Abitur graduation age											
G12 reform effect	-0.86***	-0.84***	-0.89***	-0.89***	—	-0.90***	-0.89***	-0.89***	0.00	-0.02	-0.02
	(0.04)	(0.06)	(0.03)	(0.04)		(0.04)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
N	131	131	131	131	—	127	131	1754207	94	94	94
p(β=-1)	0.004	0.023	0.002	0.014	—	0.027	0.007	0.003	0.000	0.000	0.000
Panel B: Grade repetition rates											
G8 · grade 7	-0.14	-0.00	-0.14	-0.24	0.03	-0.24	-0.38	-0.38	0.25	0.26	0.05
	(0.29)	(0.31)	(0.30)	(0.36)	(0.31)	(0.26)	(0.32)	(0.29)	(0.24)	(0.25)	(0.32)
G8 · grade 8	0.30	0.49*	0.30	0.19	0.54	0.21	0.20	0.20	0.13	0.32	0.17
	(0.28)	(0.27)	(0.28)	(0.35)	(0.35)	(0.25)	(0.22)	(0.20)	(0.24)	(0.28)	(0.32)
G8 · grade 9	0.06	0.24	0.06	-0.05	0.32	-0.01	-0.12	-0.12	0.12	0.07	0.31
	(0.34)	(0.34)	(0.34)	(0.40)	(0.45)	(0.30)	(0.27)	(0.25)	(0.20)	(0.21)	(0.19)
G8 · grade 10	1.32*	1.48**	1.31*	1.20	1.32*	1.15*	0.73	0.73	0.32	0.19	-0.00
	(0.70)	(0.67)	(0.71)	(0.74)	(0.68)	(0.64)	(0.59)	(0.53)	(0.22)	(0.19)	(0.12)
G8 · grade Q1	3.39***	3.50***	3.35***	3.30***	3.24***	2.65***	2.58**	2.58**	-0.10	-0.11	-0.14
	(0.99)	(0.94)	(0.95)	(1.03)	(0.97)	(0.86)	(1.05)	(0.94)	(0.40)	(0.32)	(0.29)
G8 · grade Q2	1.28**	1.45***	1.20**	1.19**	1.28**	1.21**	1.92***	1.92***	0.06	0.06	-0.10
	(0.44)	(0.40)	(0.54)	(0.47)	(0.44)	(0.51)	(0.54)	(0.48)	(0.28)	(0.28)	(0.34)
N	826	826	826	826	1301	804	826	12941574	858	858	858
Panel C: Abitur graduates											
G12 reform effect	-0.42	-0.45	1.21	0.93	—	0.27	-0.51	-0.51	0.51	0.91	0.63
	(1.61)	(1.00)	(1.29)	(1.58)	—	(1.46)	(1.19)	(1.06)	(0.65)	(0.51)	(0.38)
N	139	139	139	139	—	135	139	1998393	102	102	102

Note: The table reports various robustness tests for the effect of the G12 reform on the mean graduation age, high school grade repetition rates, Abitur graduates and the share of student in the high school track and in alternative school tracks. Parentheses show standard errors clustered on the state level. All models include fixed effects for state, year and time. Grade repetition rate analyses further include state-grade and time-grade interactions. Additionally, time-varying control variables on the state level (GDP growth, general and youth unemployment rate, population density, and the share of *Gymnasium*-aged children in the total population) are included. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.
[a] t denotes the year of the first G12 cohort.

Table 6: Further robustness checks for grade repetitions

	state-grade time trend (1)	grade level (2)	# students prev. year (3)	log(*rep. rate*) (4)	Placebo outcome (5)
G12 · grade 7	0.15	-0.14	-0.21	-0.00	-0.70
	(0.39)	(0.31)	(0.28)	(0.17)	(0.63)
G12 · grade 8	0.48	0.30	0.17	0.14	-0.39
	(0.30)	(0.29)	(0.26)	(0.14)	(0.70)
G12 · grade 9	0.36	0.06	-0.07	0.01	0.82
	(0.50)	(0.36)	(0.32)	(0.16)	(0.68)
G12 · grade 10	2.00**	1.31*	1.10	0.50**	0.64
	(0.87)	(0.72)	(0.65)	(0.20)	(0.80)
G12 · grade Q1	3.47***	3.97***	2.69***	0.78***	—
	(0.87)	(0.94)	(0.83)	(0.23)	
G12 · grade Q2	1.59***	-0.73	1.09**	0.76***	—
	(0.43)	(0.84)	(0.44)	(0.21)	
Adj. R-Squared	0.87	0.79	0.81	0.81	0.76
N	826	914	825	826	527

Notes: The table reports further robustness checks for the reform effect of the G12 reform on grade repetition rates. Clustered standard errors are reported in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 7: Further robustness checks for graduation rates

	# graduates as share of				
	population in age		pupils in grade 7 at		log(#
	18-20	18-19	high school	all schools	graduates)
	(1)	(2)	(3)	(4)	(5)
G12 reform effect	-0.42	-0.52	-0.11	1.06	0.02
	(1.61)	(1.82)	(2.64)	(1.04)	(0.06)
Adj. R-Squared	0.77	0.81	0.90	0.92	0.99
N	139	139	139	139	139

Notes: The table reports robustness checks for the reform effect of the G12 reform on *Abitur* graduation rates using various measures of *Abitur* graduation rates as dependent variables. Clustered standard errors are reported in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 8: Reform effect heterogeneity by gender

	pooled (1)	females (2)	males (3)
Panel A: Abitur graduation age			
G12 reform effect	-0.86***	-0.87***	-0.84***
	(0.04)	(0.04)	(0.04)
N	142	142	142
Panel B: Grade repetition rates			
G12 · grade 7	-0.14	-0.23	-0.12
	(0.29)	(0.23)	(0.38)
G12 · grade 8	0.30	0.21	0.36
	(0.28)	(0.24)	(0.37)
G12 · grade 9	0.06	-0.01	0.11
	(0.34)	(0.29)	(0.44)
G12 · grade 10	1.32*	1.45	1.99*
	(0.70)	(0.82)	(1.03)
G12 · grade Q1	3.39***	2.55**	4.36***
	(0.99)	(1.00)	(1.09)
G12 · grade Q2	1.28**	0.94**	1.72***
	(0.44)	(0.43)	(0.48)
N	826	825	825
%-change grade 7	-8.18	-16.30	-5.64
%-change grade 8	14.41	13.35	13.49
%-change grade 9	2.51	-0.68	3.79
%-change grade 10	69.94	107.39	92.84
%-change grade Q1	112.40	105.44	117.86
%-change grade Q2	120.38	105.56	140.98
Panel C: Abitur graduates			
G12 reform effect	-0.42	-0.46	-0.39
	(1.61)	(1.94)	(1.34)
N	139	139	139
%-change	-1.64	-1.58	-1.71

Notes: Effect of the G12 reform on mean graduation age, grade repetitions and overall graduation rates by gender.* $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 9: Reform effect over time on grade repetition rates

	unbalanced	balanced sample (post)			
	All years	1-5 years	1-4 years	1-3 years	1-2 years
	(1)	(2)	(3)	(4)	(5)
1 years after	3.13***	7.37***	4.45**	4.27***	3.43**
	(0.87)	(0.65)	(1.76)	(1.29)	(1.30)
2 years after	3.69***	7.23***	4.10*	4.51**	3.71**
	(1.22)	(1.06)	(2.05)	(1.49)	(1.35)
3 years after	3.18**	5.68***	2.74	3.18*	—
	(1.26)	(0.56)	(1.88)	(1.49)	
4 years after	2.78**	4.88***	2.55	—	—
	(1.01)	(0.63)	(1.56)		
5 years after	3.43***	5.01***	—	—	—
	(1.07)	(0.85)			
N	143	119	121	121	119
Effect identifying states					
Saxony-Anhalt		Yes	Yes	Yes	Yes
Mecklenburg-Vorpommern		Yes	Yes	Yes	Yes
Saarland		No	Yes	Yes	Yes
Hamburg		No	No	Yes	Yes
Bavaria		No	No	No	Yes

Note: The table reports the effect of the G12 reform on the share of grade repetitions (in %) two years before graduation for different years after the implementation of the reform. Parentheses show standard errors clustered on the state level. All models include fixed effects for state, year and time. Additionally, several time-varying control variables on the state level (GDP growth, general and youth unemployment rate, population density, and the share of *Gymnasium*-aged children in the total population) are included. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Appendix

Table A.1: First graduation cohorts affected by new policies

	G12	Reduction in subject choice options	Central exams	Tuition fees first	Tuition fees last
Baden-Württemberg	2012	2004	1952	2007	2011
Bavaria	2011	2011	1946	2007	2012
Berlin	2012	-	2007	-	-
Brandenburg	2012	2014	2005	-	-
Bremen	2012	-	2007	-	-
Hamburg	2010	2011	2005	2007	2011
Hesse	2012-2014	2005	2007	2007	2007
Lower Saxony	2011	2008	2006	2006	2013
Mecklenburg-Vorpommern	2008	2008	1990	-	-
North Rhine-Westphalia	2013		2007	2007	2010
Rhineland-Palatinate	-	-	-	-	-
Saarland	2009	2010	1946	2007	2009
Saxony	1990	2010	1990	-	-
Saxony-Anhalt	2007	2005	1990	-	-
Schleswig-Holstein	2016	2011	2008	-	-
Thuringia	1990	2011	1990	-	-

Note: The table presents for each state the first graduation cohort that was affected by the policy indicated in the column header.

Table A.2: Reform effect over time on grade repetition rates - Q2

	unbalanced	balanced sample (post)			
	All years	1-5 years	1-4 years	1-3 years	1-2 years
	(1)	(2)	(3)	(4)	(5)
1 years after	1.68***	2.35***	2.07***	1.25**	1.40***
	(0.44)	(0.26)	(0.28)	(0.55)	(0.40)
2 years after	1.55***	2.48***	1.82***	1.11*	1.50**
	(0.46)	(0.31)	(0.52)	(0.52)	(0.53)
3 years after	1.32	2.95***	1.78*	1.15	—
	(0.76)	(0.49)	(0.89)	(0.71)	
4 years after	0.97*	0.89	1.17*	—	—
	(0.53)	(0.59)	(0.56)		
5 years after	1.01*	1.66**	—	—	—
	(0.57)	(0.62)			
Adj. R-Squared	0.71	0.75	0.73	0.71	0.72
N	145	124	127	128	127
Effect identifying states					
Saxony-Anhalt		Yes	Yes	Yes	Yes
Mecklenburg-Vorpommern		No	Yes	Yes	Yes
Saarland		No	No	Yes	Yes
Hamburg		No	No	No	Yes

Note: The table reports the effect of the G12 reform on the share of grade repetitions (in %) one year before graduation for different years after the implementation of the reform. Parentheses show standard errors clustered on the state level. All models include fixed effects for state, year and time. Additionally, several time-varying control variables on the state level (GDP growth, general and youth unemployment rate, population density, and the share of *Gymnasium*-aged children in the total population) are included. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.