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

This version: December 30, 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 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. The probability to repeat a grade level in the course of high school increases by more than 20 per cent, with a peak in the final two years at high school. However, the share of a cohort graduating from high school with university entrance qualifications is not affected. The results indicate the reform's success in reducing graduation age, though it stays behind its potential benefits for labour markets, pension schemes and fertility because of higher grade repetition rates.

Keywords: G12, G8, Graduation Age, Grade Repetition, Graduation Rates, Learning Intensity, Difference-in-Differences, Education Efficiency
JEL: I28, D04

☆Acknowledgements: We gratefully acknowledge Jörg Breitung, Nicolai Krstensen and Jan van Ours as well as participants of several conferences for valuable comments. Particular thanks go to Jakob Simonsen and Geraldine Frantz for excellent research assistance. Jan Marcus greatly appreciates the generous support of the College for Interdisciplinary Education Research (CIDER).

I Introduction

In designing education systems, policy-makers face a trade-off in the optimal allocation of the length of schooling. While more years of education carry advantages for individuals and for society, they also delay labour force participation.[1]

This trade-off is particularly crucial in light of the demographic changes that many industrialised countries are facing. Social security systems are increasingly 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.

Several European countries try to tackle the trade-off between high levels of human capital and early labour market entry in different ways. In 2008, Denmark lowered the school entry age from seven to six years. In 2009, Poland passed a law to reduce the school entry age from seven to six years until 2014. In 2013, Denmark reformed their public study grant scheme to incentivise students for fast college graduation. The reform reduces the period and the amount for financial support. Also, shifting exams has been made more difficult. After 2000, Finland, Norway, Poland and Portugal passed reforms to increase instruction time provided in a given number of years.

A fundamental reform of the German education system aims at resolving this trade-off by reducing the time at high school by one year, while leaving the overall instruction time and the educational input unchanged. Thereby, education efficiency is increased for the individual student. Between 2001 and 2007, almost all German states gradually passed laws that reduce the time to obtain the university entrance diploma from 13 to 12 years of schooling, redistributing the same number of lessons over the remaining fewer school years. It is referred to as G12, denoting graduation after 12 years.[2] We refer to the old regime with high school graduation after 13 years using the term G13.

We examine indicators of the overall effectiveness of the G12 reform by looking at the impact on the high school graduation age, grade repetitions and graduation rates. As the reform's principal objective was to reduce the age at which students graduate from high

[1] 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).

[2] In the German context the reform is often referred to as G8, for 8 years of high school. However, we deem the term G12 more appropriate as the tracking age into high schools varies across federal states

school, our first outcome directly measures to what extend the reform has been effective in this dimension. The reform's potential has been a reduction in mean graduation age by one year. An important explanation for any significant deviation from this potential year could arise through changes in grade repetitions, as every additional student who repeats a grade level because of the reform cannot leave high school earlier and mechanically increases the mean graduation age.

The effect on grade repetitions is interesting for two more reasons. First, grade repetitions impose significant costs to the education system while their effectiveness in helping students is debated. Second, it may inform how well students can cope with the increased learning intensity. The increase in daily school instruction time could lead to more focussed learning, less distraction and a longer exposure to their school peers. Consequently, student performance could improve and make it less likely for a student to repeat a grade level (Lavy, 2010). On the other hand, students might have less time to revise the discussed material or to engage in extracurricular activities. At the margin, they might experience difficulties in reaching the learning goals and grade repetitions would increase.

Finally, we investigate whether the increase in learning intensity and the shortening of high school time impacts high school graduation rates and society's human capital stock. As the overall time spent in school reduces, the reform allows the individual to enter the labour market earlier. This reduces the opportunity costs of schooling and could entail an increase in high school graduation rates. On the other hand, the reform might reduce graduation rates as individuals' contemporaneous utility from schooling reduces with more learning effort and less time for recreational activities. Higher grade repetitions might also impact high school graduation rates through their effect on school drop-out probabilities (Jacob and Lefgren, 2009).

Employing administrative data on all pupils in Germany, we exploit 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 that overall grade repetitions increase by more than one fifth. However, we cannot identify any adverse effect on the share of students graduating with university entrance qualifications. The results prove to be robust throughout a wide range of sensitivity tests.

We find that grade repetitions increase strongest in the final two years at high school, while there is no effect on repetitions in grade 7 through 9. This suggests that a substantial share of grade repeaters could repeat a grade level voluntarily to improve on their school performance and GPA. Boys show a higher absolute increase in grade repetition rates. Consequently, boys' reduction in the graduation age is smaller. We cannot find gender differential effects on graduation rates.

The remainder of this study is structured as follows. Section II discusses related studies. Section III provides additional information regarding the German education system and the G12 reform. Section IV introduces the data. The empirical strategy is outlined in section V. Section VI reports the average reform effects, followed by a broad range of robustness tests in section VII. Section VIII investigates heterogeneities of the reform effect, before discussing the findings and concluding in section IX.

II Related literature

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 increases education efficiency by keeping educational input in instruction time constant while redistributing the instruction hours over fewer years. This reduces the school leaving age at constant educational input. In the following, we revise empirical findings from related policies that decide to focus on one side of the trade-off between more years of education and earlier labour market entry, and from policies that try to accomplish both targets at the same time.

Many studies analysed the effect of increased 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). These policies target students that would leave school education as soon as possible. Over their time in school, these students are thought more hours of instruction and they receive more curricular input and the returns to one more year of education for this particular group are between 5 and 10 per cent.

Other reforms try to accelerate labour market access, for example through a reduction in the school entry age. However, many studies document adverse effects of early school entry. For example, Ponzo and Scoppa (2014) use exogenous variations in the birth month and school entry cut-off dates to examine the effect of age at school entry on school performance. They find that younger age at school entry negatively affects school performance and reduces the probability to be tracked in more academic school tracks.

McEwan and Shapiro (2008) show with a regression discontinuity design on birthday based school entry cut-offs that a one-year earlier school enrolment increases the probability of grade repetitions and impairs student test scores. These adverse effects on the human capital stock rather suggest a postponement of school entry.

Morin (2013) and Krashinsky (2014) look at a reduction of the high school track length from five to four years in the Canadian province of Ontario and find that it resulted in a significantly worse educational performance of affected cohorts. Webbink (2007) analyses a reduction of Dutch university duration from five to four years. He finds that wages decrease by 7-9 %. The G12 reform differs from these reforms as it does not change total instruction time.[3]

There is also an ongoing debate on the effect of term length on student achievements, which tries to find the optimal amount of instruction time. Fitzpatrick et al. (2011), for example, use the timing of assessment dates as exogenous variation in term length and find that more instruction time improves student performance. Parinduri (2014) exploit an arbitrary rule that assigned students to a longer school year in Indonesia in 1978-79 with a regression discontinuity design and finds a decrease in the probability of grade repetition, an increase in educational attainment and an increase in wages. He concludes that school years in Indonesia are not too long.

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 in which the material the regular 37 instruction weeks was compressed to 24 weeks of instruction without nominal changes in the curriculum. This also increased learning intensity. 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 fails to find effects on long-run labour market outcomes which could suggest that the German term length has been too long or educational input too little.

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 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

[3]This 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.

(2010) find that the G12 reform generally reduces final examination scores in Mathematics and for females in English. In their analysis of the reform effect on non-cognitive skills, Büttner et al. (2011) find no impact. Using the same set of data, Meyer and Thomsen (2012) identify a decrease in university enrolment among females.

Driven by the data, these studies have several limitations. First, potential reform effects cannot be distinguished from general cohort effects as only two different school entry cohorts are compared in a single state. Furthermore, there remains the risk that findings are biased by selection effects. If the policy change actually increases grade repetitions, these studies cannot account for changes in the sample composition. Certain students would drop out of the sample of the first analysed treatment unit.

Finally, 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 redistribute 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 in university places and vocational training slots. Therefore, it is questionable whether the 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 identification strategy. Using data on from the German 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.

III Institutional background and the G12 high school reform

A. The German school system

Generally, education policy in Germany is in the political domain of the 16 federal states. Still, the education system exhibits similarities across states. All states have in common that after the joint schooling in primary school, students are tracked into different school types according to their ability. Although switching between different school tracks in possible at any grade, very few students do so.[4]

[4] In the school year 2012-13, 2.4 per cent of all students switched across the school types throughout middle school (Autorengruppe Bildungsberichterstattung, 2014).

Across all states, the highest school track *Gymnasium* (comparable to the traditional British grammar school) intends to prepare pupils for university education. This is the track that is directly affected by the G12 reform we are examining.[5] It is distinct from other school tracks in at least three respects.

First, the *Gymnasium* is the main track to earn the general university entrance qualification *Abitur*. Students holding *Abitur* are allowed to study at any higher education institution in Germany. Every state provides options to earn entrance qualifications to technical colleges. These alternatives do not directly qualify for university education. Second, compared to other school types, *Gymnasiums* exhibit higher teaching intensity and more ambitious learning goals. Third, the quality of teachers and the peer environment is considered to be better (Dustmann et al., 2014).

The decision about the school track chosen after primary school depends on state regulations. Either, admission to *Gymnasium* requires the primary teacher's recommendation, or the decision is left to the parents. The age at which students are tracked also depends on federal state regulations and is either grade 5 or grade 7. In our sample period, two states reformed the tracking age. Details are provided in table A.1 in the appendix. Across all federal states, students are tracked in 7th grade.

About one third of a cohort enter *Gymnasium* after primary school.[6] In the following, we refer to *Gymnasium* as *high school*.

In the course of their school career, students in the German system need to fulfil certain learning goals to move on to the next grade level. This is common across all school tracks and circumvents increasing performance heterogeneities within cohorts. If students fail to fulfil the learning goals, they are required to repeat the same grade level or to change into a lower school track. In case of grade repetitions, all courses have to be resit and re-examined. Next to these performance based criteria, students can also repeat a grade level voluntarily with the same consequence of resitting all courses. They may want to do so in order to better master the material or to improve on their grade point average for the school leaving certificate.

[5]The structure of lower ability tracks exhibits variations across federal states. Some states further segregate pupils by there ability types in different school forms, while others teach them together.

[6]Own calculations based on 2002-2013 cohort data from the Federal Statistical Office, *Fachserie 11 Reihe 1*.

B. The G12 reform shortens high school track tenure

In international comparisons, German pupils traditionally entered the labour market at comparably high ages (OECD, 2001).

Between 2001 and 2007, the majority of German federal states gradually implemented a reform that reduces the high school track tenure by one year. Earning *Abitur* then requires 12 rather than 13 years of schooling. We shortly refer to the new regime with graduation after 12 years by G12, and to graduation after 13 years by G13. The primary policy objective of the G12 reform has been to reduce the high school graduation age to accelerate students' labour market access, to improve their international labour market competitiveness, and to reply to demographic changes which cause increasing shortages of qualified workers and public pension scheme contributions.

Table 1 provides an overview of the timing of policy implementation across different states. The first G12 cohort graduated in 2007 in Saxony-Anhalt. Year by year, the states of Mecklenburg-Vorpommern, Saarland, Hamburg and Bavaria followed next. By 2013, pupils in 14 out of 16 federal states require 12 years of schooling to graduate with the general university entrance qualification from high school.

While the reform reduces the overall number of years spent in school, it leaves the minimum required instruction time as well as the amount of holiday unchanged (KMK, 2013). From grade level 5 onwards, at least 265 hours per week have to be distributed over the years spent at high school, i.e. the instruction hours of the year dropped had to be redistributed over the remaining eight rather than nine years. This increases the learning intensity through an increase in the hourly workload per week.

The redistribution of instruction hours varies across federal states, and can even vary across schools in the same federal state. The average change in instruction hours is poorly documented, but figure 1 provides an idea of the average increase in weekly hourly workload for each grade level in 10 out of 12 reform states. The increase in hourly workload varies across grade levels. Grades 8-10 experienced the highest increase in additional instruction hours.[7]

The increased learning intensity fuels a vital, public debate on adverse effects of the G12 reform on students' development. Parents fear that their children cannot cope with the new requirements and that they lack time for extracurricular activities. Some states

[7] 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 young pupils from too much workload.

already reverse the reform and return to the former G13 regime, though sound academic insights are scarce.[8] Other states only allowed for a selective implementation.[9] For our analysis, neither reversion nor selective implementation constitute a problem as none of the affected cohorts or states are part of our sample.

However, the first cohorts that are affected by G12 are special. As the G12 reform cuts one year of schooling, it leads to a simultaneous high school graduation of the first G12 cohort and the last G13 pre-treatment cohort. The double graduation cohort is roughly of double size and is therefore prevalently competing for places in vocational trainings, volunteering activities or for resources at universities (e.g. available places, student-teacher-ratio). This might generate incentives for affected or preceding cohorts which are discussed in the robustness section.

IV Data

A. General data

Throughout our analyses, we employ administrative data from the Federal Statistical Office for cohorts that graduate between 2002 and 2013 from high school in 14 out of 16 federal states.[10] 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 high school gradually over a period of three years. In the aggregated data, we are unable to distinguish between treated and non-treated students. The federal state of Lower Saxony does not provide information on grade repetitions for the three final years at high school. To make results comparable in terms of the sample selection, we discard Lower Saxony from the main sample.

[8]In North Rhine-Westphalia, treatment state, 13 out of 630 high schools return to G13 and graduate in 2020. Lower Saxony, not in the main sample, returns to G13 with first cohorts graduating in 2024. It will allow for an optional fast track. Baden-Württemberg, treatment state, returns to G13 in 44 out of 455 high schools with cohorts graduating in 2022.

[9]Rhineland-Palatinate, a control state, allowed high schools to apply with an elaborate concept for high school shortening. The first G12 cohorts graduate in 2016 in 9 out of 149 high schools. By 2019, 19 out of 149 high schools (12.8%) offer the fast track. Schleswig Holstein, a control state, has first selective G12 cohorts graduating in 2016.

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

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.[11]

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 is not possible to link individual pupils over time. Second, it prevents from investigating effect heterogeneities across children's socio-economic background and from investigating changes in group compositions.

We comment separately on each of the three analysed outcome variables.

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 156 observations. The potential 168 state-year observations (14 states over 12 years) reduce to 156 observations as we lack information of the age at graduation for the years 2002-2005 in Brandenburg, Hamburg and Baden-Württemberg. As all missing information lie at least two years prior to the first reform observation, we are not concerned with an impact of missing observations on our estimation results. The remaining 156 year-state observations contain information on 1.9 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. 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

[11]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.

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 or the same school type.

Pupils who do not repeat the grade at the same high school but in the same state do 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. We consider these sources of bias very minor as mobility between different school tracks is very low.

In our main sample, we observe grade repeaters at high school in 14 states for each grade level starting in grade 7 over the period 1995 through 2013. For each graduation cohort, we aggregate the number of grade repeaters from grade 7 until graduation and divide this by the cohort size in grade 7. The overall sample covers the graduation cohorts 2002 through 2013 and contains 168 state-cohort observations. These observations contain information from 2 million students.

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.[12]. For this purpose, we use information on the state's number of graduates and divide it 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. different age specifications and cohort size at school entry).

Our data set consists of 168 state-cohort observations, comprising information on 2.11 million graduates.

Across all main outcome variables, we have one observation per state and graduation cohort.

[12]This standardisation is not only relevant due large difference between the size of the states, but also because the sharp fertility drop in East Germany after the fall of the wall. For details, see Goldstein and Kreyenfeld (2011)

V Empirical strategy

Our empirical strategy makes use of the institutional peculiarity of statehood in educational affairs. This state sovereignty can be treated as a quasi-experimental setting in which exogenous time and regional variations allow for the estimation of difference-in-differences type regressions. We estimate the effect of shortening high school on mean graduation age, aggregated grade repetition rates and *Abitur* graduation rates as dependent variables y for state s and cohort c with the following two-way fixed effects model:

$$y_{sc} = \beta \cdot G12_{sc} + \mu_s + \kappa_c + \delta_{sc} + X'_{sc} \cdot \lambda + P'_{sc} \cdot \phi + \varepsilon_{sc}, \tag{1}$$

$G12_{st}$ is a binary variable that identifies the treatment status of state s and cohort c. β is the coefficient of core interest and identifies the reform effect. μ_s captures state fixed effects and accounts for time-invariant differences in the outcome variable between different states. κ_c captures general changes in the outcome variables across cohorts that are common across states. Both, state and cohort fixed effects enter the model with a set of indicator variables. We keep the double graduation cohorts in the data to avoid structural breaks in the error term series. However, we include a dummy indicator δ_{sc} for the double graduation cohorts, as (i) we cannot distinguish between treatment and non-treatment cohorts and (ii) they are distinct in several ways as described before.

An unbiased estimation of the reform effect requires the common trend assumption to hold. It says that the dynamic change of the outcome variables would have been the same in treatment and control states if the reform was not in place. In our second main specification, we allow for state and time varying control variables X_{sc} and thereby relax the common trend assumption. Unbiased reform effect estimates now require the common trend assumption to hold conditional on these control variables. For each federal state and graduation cohort, we include GDP growth, the general unemployment rate and the youth unemployment rate of 20-25 years old measured in the year prior to graduation to account for changes in states' economic environment which might impact the dynamic path of grade repetitions, graduation rates and graduation ages across states.[13]. Furthermore, children born after the fall of the German wall are part of our sample period. These cohorts

[13]Several studies document the impact of business cycles on educational decisions, such as grade retention, college enrolment and college graduation. For details, see Edwards (e.g. 1976); Betts and McFarland (e.g. 1995); Messer and Wolter (e.g. 2010); Gaini et al. (e.g. 2013)

experienced a historical decline in overall fertility rates (e.g. Goldstein and Kreyenfeld, 2011). Chevalier and Marie (2014) find evidence for adverse parental selection of these cohorts. We indicate cohorts born up to two years after the fall of the German wall with a dummy variable.

In our third specification, we also include a vector P_{sc} for other education reforms that certain states passed in the sampling period and that could impact our outcome variables. These other reforms raise the concern that the estimated reform effect also captures co-treatment effects. One relevant change constitutes the introduction of centralised school exit examinations, in which exit exams are designed by federal institutions rather than by high schools themselves. Central exit examinations change the incentive structure for teachers and students and are impacting education outcomes.[14]

Also, the age at which students are tracked has been changed in two treatment states during the sample period. Furthermore, some states combined the two alternative middle and low tracks in the German three-tier system to a single alternative school track. Both reforms could impact the school track choice and the students composition within tracks.[15]

Finally, the subject choices in the qualification phase have been reduced in some states. This alters the scope to which students can chose subjects, in which they are tested in the school exit exams, according to their abilities. This constraint might ultimately impact student performance, their probability of grade repetition or just their desire to improve on the grades when they count toward the GPA for university applications.

The described reforms are documented in table A.1 in the appendix. As none of the other policy changes is collinear to the G12 reform, there is sufficient variation in the data to distinguish the G12 effect from other policy changes. In our third specification, we consider each of these reforms with a separate dummy variable in the vector X_{st}, indicating the cohorts that are affected by the respective reform in the respective state.

The remaining variation in the data is captured by ε. We assume the error term to be autocorrelated and allow for heteroskedasticity and contemporaneous correlations across

[14]The direction of the effect may depend on the institutional context. Bishop (1997) and Jürges et al. (2005) document positive effects of central exit examinations on student achievements for Canada and Germany. Other studies find negative effects on graduation rates and student achievement for US high schools (e.g. Holme et al., 2010).

[15]Bauer and Riphahn (2006) and Pekkarinen et al. (2009) show that the age of tracking determines educational and economic mobility between generations. They suggest that earlier tracking strengthens the role of the family background in determining education outcomes.

panels.[16] Estimators on the coefficient of interest assume a Prais-Winsten AR(1) error structure, and are obtained using feasible Generalized-Least-Squares methods.

Our empirical strategy employs GLS with panel adjusted standard errors.

VI 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.

A. 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.

Before turning to the estimation results, in panel A of figure 2 we present a descriptive illustration of the mean graduation age by graduation cohort for the five treatment states that introduced the reform first.[17], and for the control group of states that never change their policy in place[18] A first descriptive inspection of figure reveals that the graduation age has decreased by one year from about 19.7 to about 18.7 years in the treatment states in the observation period.

At the same time, there is a general time trend in the mean graduation age in the control states as well. From 2002 to 2013, graduates leave school on average 0.15 years earlier. From eye-balling, this suggests a reduction of mean graduation age of about 0.85 years, or 10 months.

The regression results obtained from estimating equation (1) support this finding (see panel A of table 2). The baseline DiD specification suggests that the reform reduces the mean graduation age by 0.82 years. In model (2), as outlined in the empirical strategy, we control for factors that characterise the economic environment in which students obtain their education. This specification yields a very similar point estimate. In model (3), we

[16] In the robustness section, we experiment with different assumptions on the error term structure. This does not change our findings.

[17] Saxony-Anhalt (ST), Mecklenburg-Vorpommern (MV), Saarland (SL), Hamburg (HH) and Bavaria (BY).

[18] Saxony and Thuringia has graduation after 12 years, Schleswig-Holstein and Rhineland-Palatinate after 13 years, over the entire sample period.

further include dummy indicators for other reforms that have been implemented in the federal states during our sample period. The point estimate decreases slightly. Across all specifications, 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.

But 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.

B. Grade repetition rates

There are several reasons to believe that shortening high school tenure has an impact on grade repetitions. First, students experience a higher learning intensity because of the G12 reform. This could deteriorate pupils' capability of mastering the material. Furthermore, students may also want to repeat a grade level for strategic reasons. They might want to trade the gained one year for a better performance in school. Both motives are captured by our data and would lead to a reduction in the overall effectiveness of the reform in reducing graduation age.

Panel B of figure 2 again presents descriptive results for the first five treatment states compared to never-changing control states. In the period prior to the reform introductions, the time trend of the treatment and the control group appears to be parallel. When the reform is gradually introduced after 2006, it appears that grade repetition rates diverge between the treatment and control group.

Panel B of table 2 reports the difference-in-differences results for aggregated grade repetition rates. Column (1) constitutes the baseline difference-in-differences model and again includes state and year specific effects only. The specification in column (2) further incorporates the aforementioned time-varying economic control variables, capturing state differential developments in the economic environment. The final model in column (3) additionally accounts for other policy reforms that have been implemented in our sampling period. We find that even though treatment cohorts are intended to spend one year less at high school, their aggregated probability of repeating a grade level at high school from grade 7 until graduation increases by 3.12 percentage points, or 22 %. This effect is highly significant and very similar across all specifications.

These higher rates of grade repetition provide an explanation why the reform stays behind its potential in reducing the high school graduation age.

C. 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 the G12 reform increases grade repetition rates. Several other studies relate grade repetitions and graduation rates. However, there is no general agreement in the literature on the effect of grade repetitions on 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 an empirical question.

Figure 2 describes the development of the share of *Abitur* graduates on the 18-20 year old population in five treatment states and never-changing control states over time. There seems to be a positive impact on the share of graduates.

In table 2, we report the estimation results for the impact of the G12 reform on the graduates share based on our statistical model. The baseline DiD specification in column (1) yields a negative point estimate of -0.13 that is statistically insignificant. The same picture emerges when we control for changes in the economic environment in column (2). Further including controls for other policy reforms, the point estimate turns positive and suggests an increase in the *Abitur* graduation share of 0.46 percentage points. This effect remains statistically insignificant.

We conclude that we cannot identify a statistically and economically significant effect of shortening high school tenure on the share of *Abitur* graduates.

VII Robustness analyses

In this section, we perform various robustness checks for our three outcome variables. Panels A to C of table 3 summarise the results for different definitions of the treatment group and for different model specifications.

The first set of robustness checks in columns (2)-(5) investigates the sensitivity of the findings to different restrictions of the treatment and control group. First, we test whether the reform effect is distorted by a form of Ashenfelter's dip. Students in the penultimate pre-treatment cohort may have an incentive not to repeat a grade level to avoid allocation to the double graduation cohort, which is associated with several drawbacks for post-schooling decisions. Consequently, through lower grade repetitions in the penultimate cohort, the effect on graduation age and graduation share would be underestimated; the effect on grade repetitions would be overestimated. Therefore, we include a dummy variable in column (2) that indicates the penultimate cohort of the old regime, which removes these observations from the control group. We cannot find evidence for Ashenfelter's dip.

Further, we account for exceptional effects of the first two treatment cohorts in specification (3), which could arise through institutional adjustments and lack of experience with the redistributed curriculum and the new learning environment.

The next specification in column (4) incorporates a dummy for the first three treatment cohorts in Saxony-Anhalt and the first two treatment cohorts in Mecklenburg-Vorpommern as they were already in grade 7 to 9 and grade 7 to 8, respectively, when they were assigned to the shortened high school time. Consequently, the additional workload has been distributed over fewer remaining years for these surprised cohorts, which could affect grade repetition rates and school drop outs exceptionally.

In our main results, we excluded the state of Lower Saxony, as we lack complete information for the sampling period across all outcome variables.[19] In column(5), we incorporate the available observations.

Column (6) and column (7) examine the sensitivity of the results to a different account of the double graduation cohort. First, we exclude these completely. We lose ten observations for one cohort in each of the ten treatment states. Second, we interact the double graduation cohort with the state for which they occur.

[19]One observation for the mean graduation age and nine observations for grade repetition rates are missing.

The treatment effect identification in these quasi-experimental analyses rests on the common time trend of the outcome variables between treatment and control states in the absence of treatment. It cannot be tested directly. In the main specification we condition the outcome variables on the state's economic environment and other policy reforms to make the common trend assumption more likely to hold. The visual inspection of treatment and control groups' trend in the pretreatment period suggests a deviation from parallel time trends for the year of 2002. The specification in column (8) excludes the year of 2002 from the estimations. In column (9), we substitute the control variables with linear state specific time trends to check whether the estimated reform effect is driven by generally differing time trends between the states.

The final two specifications in table 3 deal with alternative assumptions about the error term. In the main specification, we assume the error term to be heteroskedastic and contemporaneously correlated across panels which we estimate with GLS, assuming an AR(1) process of the error term with the correlation coefficient ρ to be common across states. In column (10), the autocorrelated process of the error term is allowed to vary by state such that each correlation coefficient ρ_s is estimated at the state level. Column (11) employs a standard OLS procedure in which the standard errors are clustered at the state level.[20] The grouped errors strongly reduce the statistical power. Standard errors increase considerably in panels B through D. This is in line with evidence from Monte-Carlo simulations provided by Brewer et al. (2013).

We also run a series of placebo reform tests to further check for violations of the crucial common time trend assumption between treatment and control states. We pretend that the policy change took place one to five years before the actual reform and analyse the effects of these placebo treatments.

For placebo analyses, we include placebo policy indicators in our main regression model which pretend that the reform was implemented one to five years before its actual implementation. The results for the respective placebo treatments are reported in table 4. All coefficients on placebo reforms are insignificant and close to zero across all outcomes. This supports the main identification assumption of a similar time trend in treatment and control states in the absence of treatment.

Another important assumption of our difference-in-differences estimation strategy is

[20] We also calculated the standard errors based on wild cluster bootstrapping. The p-values are very similar to the p-values obtained from clustered standard errors.

the exclusion of compositional changes in the treatment and control groups, induced by cross-border migration or changes in the school track choice because of the reform. The school that an individual visits is partly determined by the individuals themselves.

We analyse the compositional change by analysing the share of 7-graders enrolled to the high school track. This test would also capture changes in the number of high school students because of cross-border migration. The results and all robustness checks are reported in panel D of tables 3 and 4. Throughout all specifications there is no statistical evidence for the G12 reform to alter the share of a cohort that enrols to high school rather than other school types and there is also no evidence for cross border migration. This is in line with our findings on the zero-reform impact on the graduation share. It also accords with the G12 study of Dahmann and Anger (2014) who also cannot find signs of selective migration in micro-level data.

While in our main specification we normalise the number of high school graduates with the average cohort size of the 18-20 year old, in table 5 we further normalise the number of graduates by (i) the average cohort size of the 18-19 year old, (ii) of the 18-21 year old (as employed by the Federal Statistical Office) and (iii) of the 17-19 year old. Furthermore, in column (4) we normalise the number of high school graduates by the number of students of this cohort that entered school. For the 2002 cohort, we lack information on school entries. The sample size drops by 14 state observations. Finally, 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, adverse effect of the G12 reform on graduation rates.

In summary, the real policy effects on all outcome variables are strikingly robust and support the main findings. The G12 reform effect stays behind its potential in reducing the mean graduation age. One reason is the significant increase in grade repetition rates. This increase, however, does not translate into changes in graduation rates as there is no evidence for adverse effects of the G12 reform on *Abitur* graduation rates. The findings are further supported by the absence of evidence for reform induced changes in the school track choice and the selection into high school.

VIII Effect heterogeneities

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

A. By grade level

As pointed out earlier, there are at least two motives why students repeat a grade level. First, students may have to repeat a grade level if they miss to fulfil the required performance criteria. Second, students may want to repeat a grade level voluntarily, e.g. to trade the gained year for a better performance in school which then potentially improves their grade point average for university applications or to delay labour market entry.

In the data, we cannot distinguish between the different motives and we cannot observe the final grades at the end of the school year to infer the motive. Even if this information would be observed, it is under the control of the individual student to under-perform in school to force a desired grade repetition.

Still, if we distinguish between the grade level in which the reform effect hits strongest, we can develop an idea about the underlying motives. If the additional workload is driving the increase in grade repetitions, 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).

If voluntary grade repetitions are dominant, the effect could be most pronounced within the final three years of high school. They consists of the *introductory phase* and the *qualification phase*. Students can make certain subject choices that will be at the core of their final examinations. Moreover, they can drop certain courses and specialise in others. While the *introductory phase* provides pupils with the time to acclimatise to this new system, the *qualification phase* is important as earned grades count towards the grade point average (GPA) of the *Abitur*. The GPA is important for most universities in the admission process to certain subjects that are equipped with study quotas.[21] With the reform, the institutionally distinct 2-years qualification phase has been shifted one year ahead. The role of the introductory phase (grade 11 under G13 and grade 10 under G12) is not as clear. It exhibits important elements of the original grade 10 under the G13 regime, as high school students still earn their intermediate school leaving certificate in grade 10. It no longer constitutes a year in which students can get used to selective course schemes and revise material that will be important for the qualification phase to come.

[21] 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).

For the identification of effect heterogeneities by grade levels, we adjust 1 in the following way:

$$y_{gsc} = (G12_{sc} \cdot grade_g)' \cdot \beta_g + \gamma_g + \mu_s + \kappa_c + \delta_{sc} + \pi_{gs} + \nu_{gc} + X'_{sc} \cdot \lambda + P'_{sc} \cdot \phi + \varepsilon_{gsc}. \quad (2)$$

Now, y_{gsc} denotes the fraction of pupils repeating the grade level g in state s in cohort c. We interact the $G12$ indicator with dummies $grade_g$ for each grade level such that β_g is now capturing the reform effect estimate for grade level g. In a first specification, we compare grade levels 7 through 10 and the 2-year qualification phase Q1 and Q2. We thereby account for the shift in the 2-year qualification phase and compare grade 12 and 13 under G13 to grade 11 and 12 under G12. Grade 11 is dropped under G13 and treats the introductory phase like the former grade 10. In a second specification, we instead assume a shift of the introductory phases as well and consequently drop grade 10 under G13.

δ_s, κ_c are defined as before. In the second specification, we again include economic control variables X_{sc} to relax the restrictive common trend assumption and account for the economic environment in each federal state in the year of grade repetition using GDP growth, the general unemployment rate and the youth unemployment of 20-25 years old. Again we include a dummy indicator for the cohorts born up to two years after the fall of the wall. Furthermore, we include a vector P_{sc} for other policy reforms that have been implemented in the sample period (introduction of centralised school exit examinations, change in tracking age, merger of middle tracks and low tracks, subject choice restrictions).

We also 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_{gc} controls for each grade-cohort interaction using binary variables.

The sample size increases to $168 \cdot 6 = 1008$ observations, as each we use information on six different grade levels for each cohort in each state we have separate information for grade 7 until high school graduation.

Table 6 reports the G12 reform effects on grade repetition rates for each grade level, as outlined in model (2). We account for different interpretations of the introductory phase

I. In each case, the first model accounts for state, year and grade level variations as well as state specific and time specific grade level variations. The specifications in column (2) includes the control variables.

In column (3), we use all available information across grade levels, while before we conditioned on cohorts for which we completely observe grade repetitions until graduation. The sample size increases for two reasons. First, we consider cohorts that are in higher grade levels at the beginning of our sampling period. For them, we cannot observe grade repetitions in lower grade levels. Second, we consider cohorts in lower grade levels that have not graduated from high school yet at the end of our sampling period. For them, we consequently cannot observe grade repetitions in higher grade levels.

Across all specifications, there are no significant effects on grade repetition rates in grade 7-9.

If we abstract from the introductory phase and drop grade 11 under G13, we find highly significant effects of G12 on grade repetition rates in the final three years at high school. They amount to an increase between 60 and 80 per cent. The strongest effect in absolute terms can be observed in Q1, the first year of the qualification phase, with repetition rates increasing by 2.48 percentage points. Compared to the original grade repetition level, the increase is strongest in the final year at high school. Grade repetitions increase by roughly one percentage point, equalling an increase of about 80 per cent.

The pattern is the same if we also assume a shift in the introductory phase, for which we drop grade 10 under G13. The reform effect for the introductory phase reduces in levels and significance, whereas the effects are almost identical for the 2-year qualification phase.

Notice that the effect on grade repetition rates does not correspond one-to-one to the shift in work load. Rather, the increase in grade repetitions is strongest in the final years, where examination results count towards the final grade point average with which students apply to university. Combining this with the findings of a zero-effect on *Abitur* graduation rates, this could indicate that the effect is dominated by individual, strategic decisions rather than enforced grade repetitions due to students' inability in coping with the new requirements.

It is further possible that our data partly captures students that spend a school year abroad. This was previously done in the introductory phase in grade 11, and students were previously able to move on to the qualification phase in grade 12 after their return. Now, the qualification phase starts in grade 11 and students need to repeat grade 11. We consider this a form of voluntary grade repetition.

B. By gender

A broad, interdisciplinary literature documents gender gaps in school achievements.[22] For two reasons, we expect gender differences in the G12 reform effect on the analysed outcomes. First, as boys perform on average worse in school than girls, they might experience a higher probability to fail a grade level because of the reform's higher learning intensity. Second, girls may be better equipped with non-cognitive skills that might be relevant for coping with the new learning environment, such as paying attention in class, self-discipline and organizing the work schedule (Spinath, 2014).

Since changes in grade repetition rates mechanically translate into changes in mean graduation ages, we also expect that the reduction in mean graduation age is smaller for boys in absolute terms (i.e. further away from the potential reduction of one year). It remains unclear whether the zero-effect on graduation rates masks gender heterogeneities in the reform effect, which is positive for one gender and negative for the other.

Table 7 reports the estimated G12 effects separately for females and males. The specifications differ from the main specifications only with respect to the outcome variable. The new outcomes now refer to females and males, respectively. As expected, males exhibit higher absolute increases in grade repetition rates than females. The reduction in mean graduation age is stronger for females than for males, which is in line with the hypothesis that grade repetitions attenuate the reform effect on the mean graduation age.

Finally, table 7 shows that there is no adverse effect on the graduation rates neither for females nor for males. The absence of a significant reform effect does not stem from heterogeneous treatment effects by gender.

IX Discussion and conclusion

Demographic changes in developed countries put policy-makers into a dilemma to provide individuals with high levels of education, while there need to early enter the labour market. Germany tackles the trade-off by increasing education efficiency: The

[22]Several sources for this observation have been identified, such as biological differences in brain structures, brain functioning and hormonal levels (see e.g. Cahill, 2006), differences in socialisation (see e.g. Bertrand and Pan, 2013), differences in personality and non-cognitive skills (see e.g. Spinath, 2014; Cornwell et al., 2013; Poropat, 2009).

time in high school is compressed by one year while the total number of instruction hours is left unchanged and redistributed over the remaining, fewer years at high school.

Using data from the German Federal Statistical Office that covers all pupils of the 2002 through 2013 graduation classes, we study indicators the overall effectiveness of this reform by looking at (i) the mean age at which students graduate from high school, (ii) high school grade repetition rates and (iii) high school graduation rates.

Exploiting the regional and time variation in the implementation of the reform with difference-in-differences estimations, we find that the policy change reduces the graduation age by about 10 months. The reduction stays behind its potential of one full year, which can at least partly be explained by increased grade repetition rates. We find that the probability of repeating a grade level at high school increases by more than 20 per cent. The effects are concentrated in the final three years prior to graduation with increases up to 85 per cent. Lower grade levels are seemingly unaffected by the reform. We find that the increase in grade repetitions is stronger for males than for females, translating into a smaller reduction in the graduation age in response to the G12 reform. However, we cannot find evidence for adverse effects of the G12 reform on *Abitur* graduation rates.

The estimates prove to be stable in a broad range of robustness checks and hold 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.

Our results point out that shortening high school tenure can decrease school leaving age without adverse effects on the economy's endowment with high school graduates. However, we need to distinguish our findings, which focus on the quantitative composition of the human capital stock, from reform effects on other characteristics of the human capital stock. There are two reasons to believe that the reform impacts other dimensions of the human capital stock. For example, the increased workload may lead to a substitution of extracurricular activities for school activities. Recent analyses find some evidence for a G12 reform impact on non-cognitive skills ((Dahmann and Anger, 2014)), which in turn determine labour market outcomes (Heckman et al., 2006).

Grade repetitions generate significant costs for the education system. 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). If grade repetitions are induced by under-performance, there are also considerable costs at the individual level (Eide and Goldhaber, 2005).

Individual costs would not incur if grades are repeated voluntarily. It is hard to distinguish the underlying motives with observational data. However, two things need to be noticed. First, the increase in grade repetition rates does not match the shift in the work load across grade levels. We cannot identify any effect of the G12 reform on grade repetitions in grade 7 to 9.

Second, the increase in grade repetitions is highest in the final two years of high school. This is where the incentive for voluntary grade repetition is highest, as grades count towards the GPA required for university applications. Combining this with the lack of an adverse effect on graduation rates, it favours the view that students may trade the gained year from the reform for a grade repetition with the chance to improve on their GPA. In the final years, students can make certain subject choices. If they mismatched before the reform, they could revise their choice when entering the final two years. Now, revising certain subject choices requires grade repetition.

Also, students that decided for a school year abroad in grade 11 were previously able to skip the introductory phase. Under G12, they would need to repeat the grade level. This form of grade repetition is also not induced by adverse effects on student performance and students incapability to cope with the new, more intense learning environment.

However, the results could also be consistent with a deferment effect of grade repetitions. It is conceivable that students accumulate learning deficiencies while they are still just able to cope with the higher learning intensity to pass the current grade level. Depending on how long students can delay grade repetition, it may occur in higher grade levels. We believe that the observed effect on grade repetitions is a mixture of both motives. Whatever the reason is for repeating a grade, any grade repetition induced by the reform hampers the overall gain in time of the reform.

The findings are important 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 certain ability types of students select out of the first treatment cohort. Assuming that low-ability types are now more likely to repeat a grade, any test performance comparisons of the last pre-treatment and the first treatment cohort would naturally lead to an increase in average test scores for the G12 cohort due to selection processes, even in the absence of real treatment effects. Upcoming evaluations of the G12 reform should be aware of the fact that G12 increases grade repetitions and that the student compositions between the last pre-treatment and first treatment cohort might ultimately differ.

Our findings also inform policy makers in other industrialised countries that need to resolve the trade-off between high levels of human capital and early labour market entries that in turn help deal with skilled worker shortages, deficient social security schemes or declining fertility rates.

References

Autorengruppe Bildungsberichterstattung (2014). *Bildung in Deutschland 2014*. Bertelsmann W., Bielefeld, 1st edition.

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.

Bauer, P. and Riphahn, R. T. (2006). Timing of school tracking as a determinant of intergenerational transmission of education. *Economics Letters*, 91(1):90–97.

Bertrand, M. and 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.

Betts, J. R. and McFarland, L. L. (1995). Safe Port in a Storm: The Impact of Labor Market Conditions on Community College Enrollments. *The Journal of Human Resources*, 30(4):pp. 741–765.

Bishop, J. H. (1997). The Effect of National Standards and Curriculum-Based Exams on Achievement. *The American Economic Review*, 87(2):pp. 260–264.

Brewer, M., Crossley, T. F., and Joyce, R. (2013). Inference with Difference-in-Differences Revisited. IZA Discussion Papers 7742, Institute for the Study of Labor (IZA).

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.

Cahill, L. (2006). Why sex matters for neuroscience. *Nature Reviews Neuroscience*, 7(6):477–484.

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

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.

Chevalier, A. and Marie, O. (2014). Economic Uncertainty, Parental Selection, and the Criminal Activity of the 'Children of the Wall'. *Working Paper*.

Cornwell, C., Mustard, D. B., and Parys, J. V. (2013). Noncognitive Skills and the Gender Disparities in Test Scores and Teacher Assessments: Evidence from Primary School. *Journal of Human Resources*, 48(1):236–264.

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.

Dustmann, C., Puhani, P. A., and Schönberg, U. (2014). The Long-Term Effects of Early Track Choice. IZA Discussion Papers 7897, Institute for the Study of Labor (IZA).

Edwards, L. N. (1976). School Retention of Teenagers Over the Business Cycle. *The Journal of Human Resources*, 11(2): 200–208.

Eide, E. R. and Goldhaber, D. D. (2005). Grade Retention: What Are the Costs and Benefits?. *Journal of Education Finance*, 31(2):195–214.

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.

Fitzpatrick, M. D., Grissmer, D., and 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., and 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.

Goldstein, J. R. and Kreyenfeld, M. (2011). Has East Germany Overtaken West Germany? Recent Trends in Order-Specific Fertility. *Population and Development Review*, 37(3):453–472.

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.

Heckman, J. J., Stixrud, J., and Urzua, S. (2006). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics*, 24(3):411–482.

Holme, J. J., Richards, M. P., Jimerson, J. B., and Cohen, R. W. (2010). Assessing the Effects of High School Exit Examinations. *Review of Educational Research*, 80(4):476–526.

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.

Jürges, H., Schneider, K., and Büchel, F. (2005). The effect of central exit examinations on student achievement: Quasi-experimental evidence from TIMSS Germany. *Journal of the European Economic Association*, 3(5):1134–1155.

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. Technical report, Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (Secretariat of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany), Bonn/Berlin.

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. (2010). Do Differences in Schools' Instruction Time Explain International Achievement Gaps? Evidence from Developed and Developing Countries. *NBER Working Paper*, (16227).

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.

McEwan, P. J. and Shapiro, J. S. (2008). The Benefits of Delayed Primary School Enrollment: Discontinuity Estimates Using Exact Birth Dates. *Journal of Human Resources*, 43(1).

Messer, D. and Wolter, S. (2010). Time-to-degree and the business cycle. *Education Economics*, 18(1):111–123.

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*, 46(2):441–468.

OECD (2001). *Education at a Glance 2001 OECD Indicators - OECD Indicators*. OECD Publishing, Paris.

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*.

Parinduri, R. A. (2014). Do children spend too much time in schools? Evidence from a longer school year in Indonesia. *Economics of Education Review*, 41:89–104.

Pekkarinen, T., Uusitalo, R., and Kerr, S. (2009). School tracking and intergenerational income mobility: Evidence from the Finnish comprehensive school reform. *Journal of Public Economics*, 93(7-8):965–973.

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.

Ponzo, M. and Scoppa, V. (2014). The long-lasting effects of school entry age: Evidence from Italian students. *Journal of Policy Modeling*, 36(3):578–599.

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological bulletin*, 135(2):322–338.

Spinath, B. (2014). The roles of intelligence, personality and motivation in girls' outperforming boys at school. *Personality and Individual Differences*, 60(Supplement):S45.

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

Figures

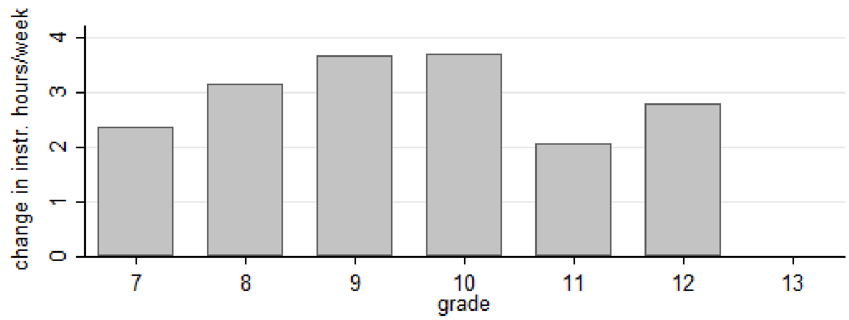

Figure 1: G12 reform changes the number of weekly instruction hours

Notes: 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. Calculations are based on 561 state-grade-year combinations for 2002-2012, thereof 298 post treatment observations. Grade 13 has been removed by the reform. Data has been provided by the *Kultusministerkonferenz* ('standing conference of the ministers of education and cultural affairs').

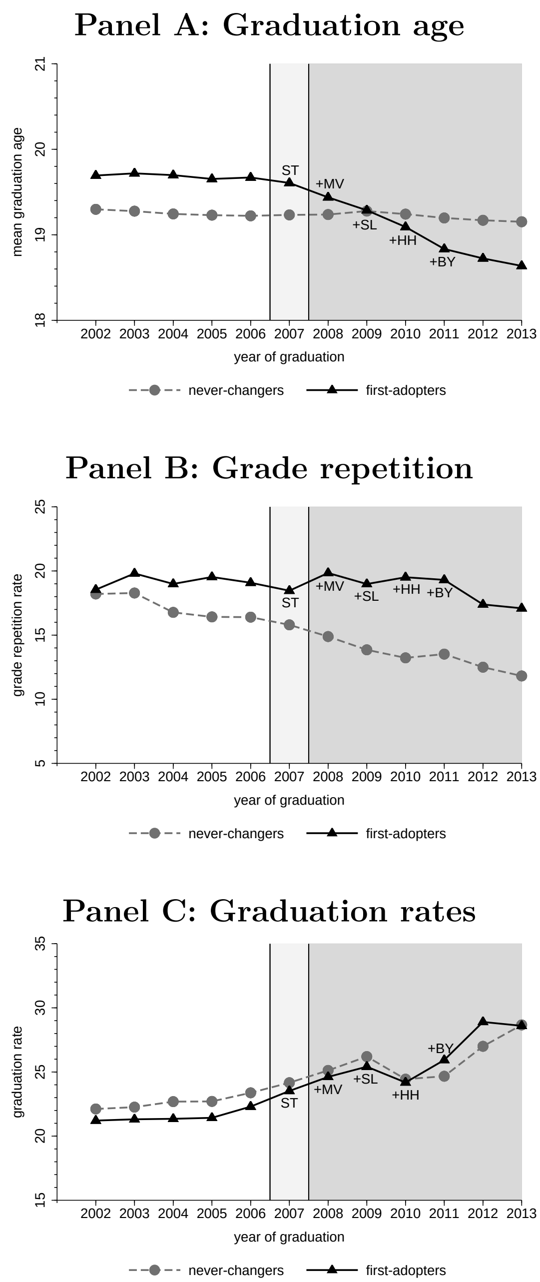

Figure 2: Trends in outcome variables in first-adopting G12 treatment states and never-changing control states

Notes: The graphs plot the means of outcome variables for first-adopters and never-changers over time. The group of never-changers consists of Saxony, Thuringia, Rhineland-Palatinate and Schleswig-Holstein. The group of first-adopters consists of Saxony-Anhalt (ST), Mecklenburg-Vorpommern (MV), Saarland (SL), Hamburg (HH) and Bavaria (BY). The year of the double graduation cohort in the respective treatment state is smoothed by the last pre-treatment and first treatment observation. The grey area constitutes the treatment period. The light-grey area does only contain a smoothed observation for the first treatment state of Saxony-Anhalt.

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
Always G12		
Saxony		
Thuringia		
Always G13		
Rhineland-Palatinate		
Schleswig-Holstein		
Excluded from main sample		
Lower-Saxony	2011	2004
Hesse	2012-14	2004

Notes: The table specifies for each federal state the first graduation cohort affected by the G12 reform and the year in which the reform was implemented at high school.

Table 2: Reform effects - main results

	baseline DiD (1)	economic controls (2)	all controls (3)
Panel A: graduation age			
G12 reform effect	-0.82***	-0.83***	-0.86***
	(0.03)	(0.03)	(0.04)
$p(\beta = -1)$	0.00	0.00	0.00
N	156	156	156
Panel B: grade repetition			
G12 reform effect	3.20***	3.25***	3.12***
	(1.17)	(0.77)	(0.94)
$\% - change$	22.31	22.74	21.64
N	168	168	168
Panel C: graduation rate			
G12 reform effect	-0.13	-0.51	0.46
	(0.95)	(0.86)	(0.81)
$\% - change$	-0.48	-1.79	1.69
N	168	168	168
State fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Economic controls	No	Yes	Yes
Policy controls	No	No	Yes

Notes: The table reports the effect of the G12 reform for the cohorts graduating 2002 through 2013 on the mean age at high school graduation, overall high school grade repetition rates and the high school graduation rate. All models include fixed effects for state and year. Economic control variables are the federal state's GDP growth, the general unemployment rate and the youth unemployment rate of 20-25 years old measured in the year prior to graduation and a dummy indicator for the post-wall generation. Policy controls are dummies for other education policies as described in table A.1. Standard errors are panel adjusted and allow the error term to be heteroskedastic and contemporaneously correlated across panels. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 3: Robustness I - Sensitivity checks

		redefining the treatment group				double cohort		common trend		standard errors	
	main (1)	w/o penultimate G13 cohort (2)	w/o first two treatment (3)	w/o surprised (4)	with LS (5)	w/o DGC (6)	interacted DGC (7)	w/o 2002 (8)	state time trend (9)	psar1 (10)	cluster (11)
Panel A: graduation age											
G12 reform effect	-0.86***	-0.87***	-0.91***	-0.89***	-0.85***	-0.85***	-0.85***	-0.86***	-0.76***	-0.85***	-0.88***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
N	156	156	156	156	167	146	156	145	156	156	156
Panel B: grade repetition											
G12 reform effect	3.12***	3.47***	3.65***	2.75***	3.00***	3.03***	2.90***	2.98***	2.62***	3.49***	3.60**
	(0.94)	(0.99)	(1.13)	(1.01)	(1.10)	(0.95)	(0.89)	(0.91)	(0.84)	(0.87)	(1.55)
N	168	168	168	168	171	158	168	154	168	168	168
Panel C: graduation rate											
G12 reform effect	0.46	0.82	-0.22	0.57	0.51	-0.10	-0.16	0.47	-1.35	0.22	0.31
	(0.81)	(0.87)	(1.01)	(0.81)	(0.73)	(0.51)	(0.59)	(0.85)	(1.23)	(0.71)	(1.44)
N	168	168	168	168	180	158	168	154	168	168	168
Panel D: grade 7 high school share											
G12 reform effect	0.70	0.70	-0.34	0.52	0.82	0.78	0.60	0.70	0.50	0.73	0.61
	(0.56)	(0.65)	(0.85)	(0.55)	(0.58)	(0.56)	(0.58)	(0.57)	(0.70)	(0.49)	(0.83)
N	168	168	168	168	180	158	168	154	168	168	168

Notes: The table reports robustness tests for the effect of the G12 reform for the cohorts graduating 2002 through 2013 on the mean age at high school graduation, overall high school grade repetition rates, the high school graduation rate and the share of a cohort enrolled to high school in grade 7. All models are based on the main specification, see table 2 column (3). Standard errors are panel adjusted and allow the error term to be heteroskedastic and contemporaneously correlated across panels. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 4: Robustness II - Placebo tests

	Placebo reform in				
	$t-1$	$t-2$	$t-3$	$t-4$	$t-5$
	(1)	(2)	(3)	(4)	(5)
Panel A: graduation age					
G12 reform effect	-0.03	-0.01	-0.04	-0.02	-0.01
	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)
N	156	156	156	156	156
Panel B: grade repetition					
G12 reform effect	0.69	0.48	0.69	0.52	-0.57
	(1.00)	(0.75)	(0.69)	(0.69)	(0.80)
N	168	168	168	168	168
Panel C: graduation rate					
G12 reform effect	0.65	0.72	0.81	0.78	0.30
	(0.68)	(0.66)	(0.61)	(0.53)	(0.52)
N	168	168	168	168	168
Panel D: grade 7 high school share					
G12 reform effect	0.00	0.20	0.00	0.16	0.04
	(0.42)	(0.42)	(0.40)	(0.35)	(0.37)
N	168	168	168	168	168

Notes: The table reports various placebo tests for the effect of the G12 reform for the cohorts graduating 2002 through 2013 on the mean age at high school graduation, overall high school grade repetition rates, the high school graduation rate and the share of a cohort enrolled to high school in grade 7. All models are based on the main specification, see table 2 column (3). Additionally, each model includes a placebo reform dummy indicator for which the coefficient is reported in this table. Standard errors are panel adjusted and allow the error term to be heteroskedastic and contemporaneously correlated across panels. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 5: Robustness III - Graduation share normalisation

	18-20 (1)	18-19 (2)	18-21 (3)	17-19 (4)	school entry (5)	log(#) (6)
G12 reform effect	0.46	0.68	0.40	1.10	0.75	0.05
	(0.81)	(0.82)	(0.85)	(0.86)	(1.14)	(0.04)
N	168	168	168	168	154	168

Notes: The table reports the effect of the G12 reform on the share of graduates with different normalisations for the cohorts graduating 2002 through 2013. Column (1) reports the original estimate for the share of graduates on the average 18-20 years old population. Column (2)-(4) normalise the number of graduates by the average size of the 18-19, 18-21 and 17-19 years old population, respectively. Column (5) normalises by the size of the graduation cohort at school entry. Column (6) reports the results for the logarithm of the number of graduates. All models are based on the main specification, see table 2 column (3). Standard errors are panel adjusted and allow the error term to be heteroskedastic and contemporaneously correlated across panels. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 6: Reform effect - by grade level

	grade 11 dropped under G13			grade 10 dropped under G13		
	baseline DiD (1)	all controls (2)	all years (3)	baseline DiD (1)	all controls (2)	all years (3)
G12 · grade 7	0.11	0.01	-0.26	0.12	0.02	-0.24
	(0.29)	(0.27)	(0.24)	(0.28)	(0.26)	(0.23)
G12 · grade 8	0.44	0.29	0.11	0.46	0.34	0.14
	(0.32)	(0.31)	(0.27)	(0.31)	(0.31)	(0.26)
G12 · grade 9	0.33	0.14	0.03	0.35	0.17	0.05
	(0.31)	(0.26)	(0.24)	(0.30)	(0.26)	(0.23)
G12 · grade 10/I	1.50***	1.27***	1.15***	0.86**	0.68*	0.62*
	(0.35)	(0.34)	(0.30)	(0.38)	(0.37)	(0.33)
G12 · grade Q1	2.70***	2.48***	2.28***	2.73***	2.51***	2.32***
	(0.50)	(0.49)	(0.47)	(0.48)	(0.47)	(0.45)
G12 · grade Q2	1.25***	1.01***	0.98***	1.26***	1.02***	1.02***
	(0.28)	(0.32)	(0.33)	(0.28)	(0.30)	(0.30)
N	1008	1008	1451	1008	1008	1459
%-change grade 7	6.04	0.79	-14.44	6.84	1.29	-13.50
%-change grade 8	18.91	11.81	5.12	19.90	14.09	6.48
%-change grade 9	14.28	5.82	1.05	15.19	6.86	1.90
%-change grade 10/I	89.30	66.67	57.05	37.11	27.39	24.62
%-change grade Q1	80.18	68.85	62.32	81.37	70.22	64.20
%-change grade Q2	129.02	82.81	79.74	131.58	84.90	84.96

Notes: The table reports the effect of the G12 reform for the cohorts graduating 2002 through 2013 on grade repetition rates separately by grade levels. All models are based on the main specification, see table 2 column (3). Standard errors are panel adjusted and allow the error term to be heteroskedastic and contemporaneously correlated across panels. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 7: Reform effect - by gender

	female (1)	male (2)
Panel A: graduation age		
G12 reform effect	-0.87***	-0.84***
	(0.04)	(0.04)
N	156	156
Panel B: grade repetition		
G12 reform effect	2.40***	3.78***
	(0.88)	(1.14)
N	168	168
Panel C: graduation rate		
G12 reform effect	0.11	0.76
	(0.95)	(0.73)
N	168	168

Notes: The table reports the effect of the G12 reform for the cohorts graduating 2002 through 2013 on the mean age at high school graduation, overall high school grade repetition rates and the high school graduation share separately by gender. The dependent variable is gender specific. All models are based on the main specification, see table 2 column (3). Standard errors are panel adjusted and allow the error term to be heteroskedastic and contemporaneously correlated across panels. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Appendix

Table A.1: Reforms

	G12	central exit examination	tracking in grade 7	two-tier system	restricted upper-secondary subject choice
Change from G13 to G12					
Saxony-Anhalt	from 2007	all	2006-2009	from 2006	from 2005
Mecklenburg-Vorpommern	from 2008	all	none	from 2010	from 2008
Saarland	from 2009	all	none	from 2006	from 2010
Hamburg	from 2010	all	none	none	from 2011
Bavaria	from 2011	all	none	none	from 2011
Baden-Württemberg	from 2012	all	none	none	from 2004
Bremen	from 2012	from 2007	until 2011	from 2012	all
Berlin	from 2012	from 2007	all	none	all
Brandenburg	from 2012	from 2005	all	from 2012	none
North Rhine-Westphalia	from 2013	from 2007	none	none	all
Always G12					
Saxony	all	all	none	all	from 2010
Thuringia	all	all	none	all	from 2011
Always G13					
Rhineland-Palatinate	none	all	none	none	from 2011
Schleswig-Holstein	none	from 2008	none	none	from 2011
Excluded from main sample					
Lower-Saxony	from 2011	from 2006	until 2011	none	from 2008
Hesse	from 2012	from 2007	none	none	from 2005

Notes: The table reports the graduation cohorts in our sample period 2002-2013 that are affected by different education reforms and institutional changes. *Centralised school exit examinations* shift the design of exit exams from high schools to federal state institutions such that all students in the specific state sit the same exit exam. *Tracking in grade 7* indicates reforms that changed the age at which students are tracked. *Two-tier system* indicates reforms that combine the low and middle 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 high school has been restricted. Sources for the reform dates are available from the authors upon request.