Moving up a Gear: The Impact of Compressing Instructional Time into Fewer Years of Schooling

Mathias Huebener and Jan Marcus

Opinions expressed in this paper are those of the author(s) and do not necessarily reflect views of the institute.

Moving up a gear: The impact of compressing instructional time into fewer years of schooling

This version: February 10, 2015

Mathias Huebener, Jan Marcus

DIW Berlin

Abstract

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 reducing high school by one year while leaving the total instructional time unchanged. Employing administrative data on all high school graduates in 2002-2013 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 compressing the high school track by one year reduces the mean high school graduation age by about 10 months. The probability to repeat a grade level in the course of high school increases by 21 percent (3 percentage points), peaking in the final three years before graduation. However, the high school graduation rate is not affected. The results indicate the reform's success in reducing the graduation age, though it stays behind its potential benefits for labour markets and social security schemes because of higher grade repetition rates.

Keywords: G12, G8, graduation age, grade repetition, grade retention, graduation rates, learning intensity, diff-in-diff, human capital, instructional time
JEL: I28, J18, D04

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

1 Introduction

Policy-makers face a trade-off in the optimal allocation of the length of schooling. While more years of education carry monetary and non-monetary advantages for individuals and for society (Card, 1999; Grossman, 2006; Lochner, 2011), they also delay labour force participation (Brugiavini and Peracchi, 2005). This trade-off is particularly crucial in light of population ageing in industrialised countries. The sustainability of social security systems is threatened by declining support ratios. Furthermore, ageing societies require policy-makers to respond to skilled worker shortages. Increases in the retirement age regularly appear on national policy agendas to mitigate the consequences of ageing populations, despite their unpopularity (Scheubel et al., 2013).

Alternatively, an earlier labour market entry could mitigate the demographic problem (Brugiavini and Peracchi, 2005; Börsch-Supan et al., 2014). Several industrialised countries try to achieve this by lowering the school starting age.[1] Similarly, the current European-wide harmonisation of university systems (*Bologna process*) aims at reducing the time spent in formal education (Bologna Declaration, 1999). Also, some states incentivise students for speedy college graduation by reforming public study grant schemes (e.g. Denmark in 2013) or by introducing additional tuition fees for long-term students (e.g. several German states).

German policy-makers suggested a novel policy action to allow for earlier labour market entry. Between 2001 and 2007, 13 out of 16 German states passed laws that reduce the high school track tenure by one year, while redistributing the same instructional time over the remaining fewer school years. Thereby, the formal learning intensity is increased for the individual student. The reform is referred to as G12, denoting high school graduation after 12 years, while the term G13 refers to the old regime with graduation after 13 years.[2]

We examine indicators of the overall effectiveness of the G12 reform by looking at the impact on high school graduation age, grade repetitions and graduation rates. Employing

---

[1]For instance, in 2008, Denmark lowered the school starting age from seven to six years. Poland passed a law in 2009 to reduce the school entry age from seven to six years by 2014. By 2015, Switzerland is integrating pre-school into primary school and reducing the school starting age to 4 years. In 2014, the Japanese Government started discussing a reduction in the school starting age from 6 to 5 years.

[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 and hence the years spent in high school vary across federal states. The term G12 is also more prevalent in the economic literature on high school tenure reductions in Canada and Germany (Morin, 2013; Büttner and Thomsen, 2015).

administrative data on all students from the graduation cohorts 2002-2013 in Germany, we exploit both temporal and regional variation in the implementation of the reform. Difference-in-differences estimates reveal that G12 reduces the mean graduation age by about 10 months, suggesting that the reform stays behind its potential of one full year. One possible explanation are increased grade repetitions, which we find to go up by 21 percent (3 percentage points) over the course of high school. However, there is no evidence for any adverse effect of the reform on completed education, measured by the share of students graduating with university entrance qualifications from high school. All findings prove to be robust throughout a wide range of sensitivity tests, including placebo tests and different sample restrictions.

Our results point to important effect heterogeneities. Boys show a higher increase in grade repetition rates and consequently a smaller reduction in the graduation age. We cannot find gender-specific effects on graduation rates. Grade repetitions increase mainly in the final years at high school, while there is no significant effect on repetitions in grades 7-9. The final two years are the ones in which the incentive is strongest to repeat a grade level voluntarily. There is no evidence that the effects are only of a temporary nature.

As the reform leaves the educational input the same, it constitutes an alternative for all policy-makers who aim at accelerating students' labour market entry without reducing educational input. However, little is known about the G12 reform impact. This study contributes to the literature in several dimensions. We provide the first analyses on the effect of this fundamental education reform on its major objective, the reduction in the school leaving age. We add an estimate of the G12 reform effect on educational attainments by looking at graduation rates. Furthermore, the analyses of grade repetitions generate important insights about students' ability to cope with the increased learning intensity, about additional costs of the reform, and about group compositional changes, which are important for future research that analyses the G12 reform.

The remainder of this study is structured as follows. Section 2 provides essential information regarding the German education system and the G12 reform. Section 3 discusses the related economic literature. Section 4 introduces the data, while the empirical strategy is outlined in section 5. Section 6 reports the average reform effects, followed by a broad range of robustness tests in section 7. Section 8 investigates heterogeneities of the reform effect, before discussing the findings and concluding in section 9.

2 Institutional background and the G12 high school reform

2.1 The German school system

Generally, education policy in Germany is in the domain of the 16 federal states. Still, the education system exhibits strong similarities across states. All states have in common that students are tracked into different school types according to their ability, after joint primary schooling.[3] Across all states, the highest school track *Gymnasium* intends to prepare students for university education. About one third of a cohort enters *Gymnasium* after primary school.[4] This is the only track that is directly affected by the G12 reform we are examining. 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, the *Abitur*. 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 tracks, *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). In the following, we use the terms *Gymnasium* and *high school* interchangeably.

High school students in Germany need to fulfil certain learning goals to move on to the next grade level. Otherwise, they are required to repeat the same grade level or to change into a lower school track. This ought to circumvent increasing performance heterogeneities within cohorts. 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 (GPA) for their school leaving certificate.[5]

2.2 The G12 reform compresses instructional time

Between 2001 and 2007, 13 out of 16 German federal states gradually implemented a reform that reduces the high school track tenure by one year. Earning *Abitur* then requires

---

[3]Switching between different school tracks is possible at any grade, but very few students do so. In the school year 2012-2013, 2.4 percent of all students switched between school types (Autorengruppe Bildungsberichterstattung, 2014).

[4]Own calculations based on cohorts graduating 2002 through 2013 (German Federal Statistical Office, 2015a).

[5]The GPA is important for admission to university. For the winter term 2013/14, 51.5 percent of all bachelor study courses in Germany restricted admissions by GPA (Herdin and Hachmeister, 2014).

12 rather than 13 years of schooling, while the instructional time is redistributed over the remaining fewer years. The primary policy objectives of the G12 reform was to reduce the high school graduation age in order i) to accelerate students' labour market access as German students entered the labour market at comparably high ages (OECD, 2000); ii) to improve their international labour market competitiveness; and iii) to respond to population ageing that is causing increasing shortages of qualified workers and declining support ratios to social security schemes.

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

While the reform reduces the overall number of years spent in school, it leaves the minimum required instructional time and the amount of holiday unchanged. The redistribution of instructional hours of the dropped year varies across federal states and across schools within federal states. On average, the hourly workload increases by 3.7 instruction hours per week, or 12.5 percent, and thereby raises the learning intensity for high school students. Grades 8-10 experienced the highest increase in instructional time. Not much of the workload was shifted to the 5th and 6th grade in order to prevent young students from overworking (KMK, 2013).

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 have already reversed the reform, returning to the former G13 regime,[6] even though sound academic insights on the reform's impact are scarce. These reversions do not affect the cohorts in our sample.

The first G12 cohort is 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 co-

---

[6]In North Rhine-Westphalia, a treatment state, 13 out of 630 high schools return to G13 with first graduates in 2020. Lower Saxony, not in the main sample, returns to G13 with first cohorts graduating in 2021. It will allow for an optional fast track. Baden-Württemberg, a treatment state, returns to G13 in 44 out of 455 high schools with first cohorts graduating in 2022. Schleswig Holstein, a control state, has first G12 cohorts graduating in 2016 and offers both G12 and G13 for cohorts graduating after 2018.

hort. This double graduation cohort is roughly of double size and is therefore prevalently competing for resources at universities (e.g. available places, student-teacher-ratio), and for places in vocational trainings and volunteering activities. This might generate exceptional incentives for affected and preceding cohorts. We discuss this issue in the robustness section.

3 Related literature

Policy-makers face a trade-off in the provision of high levels of schooling and earlier labour market entries. One way to tackle this trade-off is to reduce the labour market entry age via a reduction in the school leaving age, denoted age^l. This age is determined by the school entry age, age^e, and the number of years spent in school to earn a specific certificate, S:

$$age^l = age^e + S \tag{1}$$

At the same time policy-makers want to preserve or even increase the level of schooling. For illustration, think of schooling to be solely determined by the number of instruction hours, with the hourly educational input set to unity for simplicity. The overall number of received instruction hours for a certain school certificate, H, is the product of instructional time per day, h, school days per week, d, and school weeks per year, w, summed over all years spent in school to earn this degree, S:

$$H = \sum_{t=1}^{S} h_t \cdot d_t \cdot w_t \tag{2}$$

While both, overall instructional time H and school leaving age age^l, increase in the years of schooling S, policy-makers aim at increasing H while reducing age^l. The German G12 reform can also be illustrated with these two equations. While age^l should be decreased through a reduction in S by one year, H remains constant as increases in daily instructional time, h, compensate for the reduction in S (d and w remain unchanged).[7] Several policy reforms that change one or more parameters in equations (1) and (2) are related to the G12 reform. In the following we discuss findings on these related policies

---

[7] More specifically, the states' education ministers agreed to require $H/w \geq 265$ hours for graduation (from grade 5 to graduation), independent of whether students graduate under the G12 or the G13 regime (KMK, 2013).

along the parameters in the two equations and identify differences and similarities to the G12 reform.

One of the most obvious reforms that change the school leaving age, age^l, are changes in compulsory schooling laws. Any legislative change in minimum compulsory schooling alters age^l of complying students through S, but they also change the total number of received instruction hours H. The consensus from this literature is that increases in minimum compulsory schooling provide monetary and non-monetary benefits (see e.g. Card, 1999; Grossman, 2006; Lochner, 2011; Machin et al., 2012; Carneiro et al., 2013). These reforms mostly affect low achieving students. Reductions in minimum compulsory schooling would consequently reduce their human capital and conflict with policy-makers' goal of providing high levels of education. Alternatively, policy-makers may want to change the required number of years to earn a specific degree for higher achieving students. Webbink (2007), Morin (2013) and Krashinsky (2014) evaluate policies that reduce the years to degree S for high school and university students, what also carries reductions in the total instructional time H. They find reform-induced poorer performances at university and on the labour market.

Another option to reduce age^l in equation (1) are changes in the school starting age age^e, which allow altering the school leaving age for all student, while leaving the educational input H constant. There is a lasting debate on the impact of the school starting age on student outcomes. The literature has mostly focussed on the impact of the relative school starting age for individuals, exploiting birth date variations around given starting age rules (Bedard and Dhuey, 2006; Puhani and Weber, 2007; McEwan and Shapiro, 2008; Black et al., 2011; Fredriksson and Ockert, 2014). Policy-makers, however, can only shift the absolute age at school entry. To our knowledge, Bedard and Dhuey (2012) are the only who address the question from a policy-maker's perspective and analyse the overall impact of school-entry policies. They find that later school entry has significant positive effects on the cohort's overall earnings and suggest policy-makers to back-up school entry at the cost of later labour market entry.

Additionally, policies that alter parameters of equation (2) are related to the G12 reform. For instance, there is an ongoing debate on the effect of term length (i.e. changes in w or d) on student achievements (e.g. Woessmann, 2003; Bellei, 2009; Fitzpatrick et al., 2011). Pischke (2007) studies the effects of reductions in term length, analysing a German 1966-1967 reform that aimed at harmonising the school year nationally. The reform introduced two short school years that taught the material of the regular 37 instruction

weeks within 24 weeks of instruction without nominal changes in the total educational input. Pischke finds that this increase in learning intensity increases grade repetition rates and decreases the number of students enrolled in higher secondary school tracks. He cannot find effects on long-run labour market outcomes.

The short school years in 1966-1967 analysed by Pischke (2007) are conceptionally most similar to the G12 reform, as both reforms compress the instructional time into fewer years of schooling. However, not much is known about the impact of G12 on student outcomes. Initial research investigates the reform effect on student test scores, non-cognitive skills and post-schooling decisions. Büttner and Thomsen (2015) use a survey of 12 schools in two cities in the federal state of Saxony-Anhalt to estimate the reform effect on final exam scores. The authors find that the reform reduces scores in mathematics, while there is no effect on German literature scores. Based on the same data, Meyer and Thomsen (2012) identify decreased university enrolment among females because of G12, and Thiel et al. (2014) do not find a reform effect on measures of non-cognitive skills.

Despite their valuable contributions, these studies experience data driven limitations. The generalisability of the findings to overall G12 effects is challenged by the data that covers high school graduates of the double graduation cohort in one federal state. As the treatment status only varies by school entry cohort, it prevents from accounting for confounders such as birth cohort effects. Also, the analyses focus on a specific, special, cohort. The double graduation cohort is roughly of double size. Students compete for the same resources in post-secondary education, which might alter performance incentives in school compared to other cohorts. Furthermore, the first treated cohort in this particular state is anomalous in that students were only informed in grade 9 that they graduate one year earlier. This caused an extraordinary and exceptional increase in workload, which may not be informative about the policy impact on later cohorts or in other states. Finally, the studies need to rely on information collected two years after graduation of at most $N = 724$ observations. As grade repeaters are excluded from the samples, there dataset may exhibit compositional differences in student characteristics between treatment and control units caused by the reform itself.

The only study that accounts for these problems in a difference-in-differences approach similar to our study is by Dahmann and Anger (2014). They analyse the reform effect on non-cognitive skills, using 2005-2012 data from the German Socio-Economic Panel Study (SOEP) on $N = 1,058$ individuals from 13 different states. Unlike Thiel et al. (2014), they find some evidence for reform effects on specific personality traits.

4 Data

4.1 General data

Throughout our analyses, we rely on administrative data from the Federal Statistical Office for all students that graduated from high school between 2002 and 2013 (German Federal Statistical Office, 2015a).[8] The data cover the universe of all students in these federal states and contain the relevant information aggregated by year, gender and school type.

The rich data have not been used much by economists, with Pischke (2007) among the exceptions. The data set has three main advantages, which help to overcome some of the data limitations of previous studies on the G12 reform. First, it is a full population survey. Second, information about graduation and grade repetitions are not self-reported by the individuals. 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 required by law to provide the requested information.[9] However, it is not possible to link individuals over time.

We cannot consider the federal states of Hesse and Lower Saxony in our main analyses. Hesse introduced the compression of instructional time gradually over a period of three years. In the aggregated data, we are unable to distinguish between treated and non-treated students in the implementation period. Lower Saxony lacks information on grade repetitions for the three final years at high school for most of the years. In robustness checks we include the available information on Lower Saxony.

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

4.2 Graduation age

For each cohort of *Abitur* graduates from high school, the Federal Statistical Office provides information about the distribution of graduates' birth years. From this information, we first calculate the age, a, by subtracting the birth year from the graduation year. Then, we compute the mean graduation age y^{age} for each state s and each graduation cohort c according to the following equation:

---

[8]This data is available electronically as school year specific Excel files for recent years, while earlier information was published only as printed version. The prepared data are available from the authors upon request.

[9]The schools provide the information to the statistical offices of the federal states. The Federal Statistical Office then harmonises the state level information and makes them publicly available.

$$y_{sc}^{age} = \frac{\sum\limits_{a} a \cdot N_{sc,a}}{\sum\limits_{a} N_{sc,a}},$$

where $N_{sc,a}$ denotes the number of students graduating at age a in state s and cohort c.

The potential 168 state-cohort observations (14 states over 12 years) reduce to 156 observations as we lack information about the age at graduation for the pre-treatment years 2002-2005 in Brandenburg, Hamburg and Baden-Württemberg. As all missing information lies 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 2.20 million students.

4.3 Grade repetitions

The data on grade repetitions are further disaggregated at the grade level. We observe grade repeaters at high school for each grade level over the school years 1994/95 through 2012/13. For each graduation cohort c in state s, we obtain the aggregate grade repetition probability y^{rep} by summing over the number of grade repeaters R starting in grade $g = 7$ until graduation and dividing this by the cohort size N in grade 7:[10]

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

We multiply the fraction with 100 for the sake of clarity in the tables and for an immediate percentage point interpretation. Note that for G12 repetitions are summed over six grades, while for G13 they are summed over seven grades. For the main analysis, we use information on the graduation cohorts 2002-2013. This leaves us with 168 state-cohort observations, containing information about 2.35 million graduates.

Note that the nature of the data recording introduces a source of potential measurement error. At the beginning of the new school year (usually in September), it is recorded how many students repeated the respective grade level. The vast majority of students who did not pass a grade repeats the grade at the same high school or the same school type. Students who repeat a grade level at a lower school type or in another state are potential sources of measurement error. However, we consider these sources of potential bias very

---

[10]The age at which students are tracked depends on federal state regulations and is either grade 5 or grade 7. We start aggregating in grade 7 to circumvent these differences and the change in tracking age policies in two states in our sample period. Details are provided in table 1.

minor for two reasons. First, mobility between different school tracks is generally very low (see section 2). Second, measurement errors in the outcome variable bias the reform effect estimates only if they correlate with the introduction of G12. A systematic increase in the ratio of repeaters who repeat the grade in another state or another school track will most likely result in a downward bias of the reform's effect on grade repetitions. Random measurement errors would only affect the precision of our estimates and result in larger standard errors.

4.4 Graduation rate

The Federal Statistical Office also reports the total number of students who earned 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 normalise it. 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, which is provided by the German Federal Statistical Office (2015b). Again, we multiply the share with 100. Our data set consists of 168 state-cohort observations, containing information on 2.35 million graduates.

5   Empirical strategy

Our empirical strategy makes use of the fact that the G12 reform was implemented at different points in time in different states. This quasi-experimental setting allows to implement a difference-in-differences (DiD) type approach. We estimate the effect of the G12 reform on graduation age, grade repetitions and graduation rates as dependent variables y with the following two-way fixed effects model:

$$y_{sc} = \beta \cdot G12_{sc} + \mu_s + \kappa_c + \delta \cdot DGC_{sc} + X'_{sc} \cdot \lambda + P'_{sc} \cdot \phi + \varepsilon_{sc} \tag{3}$$

$G12_{sc}$ is a binary variable that identifies the treatment status of state s and graduation cohort c. β is the coefficient of core interest and identifies the reform effect. State fixed effects (μ_s) account for cohort-invariant differences in the outcome variables between different states. κ_c captures general changes in the outcome variables across cohorts. We include a dummy indicator DGC_{sc} for the double graduation cohorts, as we cannot

distinguish between the last G13 and the first G12 cohort in the data.[11]

We allow for state and time varying control variables X_{sc} and thereby relax the common trend assumption. 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 (German Federal Statistical Office, 2015b) measured in the year prior to graduation to account for changes in states' economic environment that might impact the dynamic path of grade repetitions, graduation rates and graduation ages across states.[12] Furthermore, also children born in East Germany shortly after the fall of the wall are included in our sample. These cohorts experienced a historical decline in overall fertility rates (e.g. Goldstein and Kreyenfeld, 2011). Chevalier and Marie (2013) find evidence for adverse parental selection of these cohorts. Hence, we indicate observations of East German cohorts born up to two years after the fall of the wall with a separate dummy variable, with birth calculations based on the regular school duration until graduation.

We also include a vector P_{sc} of binary variables for other education reforms certain states passed during the sampling period, which could impact our point estimates as potential co-treatments. None of the considered policy changes is collinear to the G12 reform, as documented in table 1. The first relevant change constitutes the introduction of centralised school exit examinations, in which exit exams are designed by federal institutions rather than by high schools. Central exit examinations change the incentive structure for teachers and students, thus impacting education outcomes (e.g. Brahm, 1997; Jürges et al., 2005). Additionally, the age at which students are tracked changed in two treatment states during the sampling period. Furthermore, some states combined the middle and low track in the German three-tier system into a single alternative school track. Both reforms could impact the school track choice and the student composition within tracks (e.g. Bauer and Riphahn, 2006; Pekkarinen et al., 2009). Finally, the subject choices in the qualification phase were reduced in some states, which might also affect our outcomes (see e.g. Görlitz and Gravert, 2015).

The remaining variation in the data is captured by the error term ε_{sc}. A common problem in difference-in-differences settings is that error terms might be serially correlated

---

[11]In our main analysis, we refrain from excluding the double graduation cohorts to avoid breaks in the time series of the error term, which would complicate estimations of its autocorrelation structure. In the robustness section, we show that our results are insensitive to dropping the double graduation cohorts.

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

within states, also after controlling for state and time fixed effects (Bertrand et al., 2004). We account for this problem by allowing the error terms to be autocorrelated of first order within states, i.e. we assume the error terms to follow a common AR(1) process:[13]

$$\varepsilon_{sc} = \rho \cdot \varepsilon_{s,c-1} + v_{sc}, \tag{4}$$

where v_{sc} is an idiosyncratic error term. Additionally, we allow for panel-level heteroskedasticity (i.e. each state s may have an individual error term variance; $Var(\varepsilon_{sc}) = \sigma_s^2$) and contemporaneously correlated errors across states (i.e. each pair of states may have its own error term covariance; $Cov(\varepsilon_{ic}, \varepsilon_{jc}) = \sigma_{ij}, \forall i, j \in s$). We estimate the regression equation with feasible Generalised Least Squares (FGLS) applying the Prais-Winsten transformation in order to use the first cohort in each state as well. In the robustness section, we relax the assumption of the specific autocorrelation structure and work with different assumptions about the error term structure (e.g., panel-specific AR(1) processes, clustered standard errors).

6   Results

6.1 Graphical results

We first present descriptive evidence on the treatment effects. Figure 1 plots the evolution of the outcomes for cohorts in the first policy adopting states and for cohorts in states that never change their policy in place.[14]

Panel A of figure 1 plots the mean graduation age for both groups. Before the first state implements the reform in 2007, the graphs of first-adopting and never-changing states move parallel, providing some confidence for the crucial common trend assumption in our empirical estimations. In 2007, the first G12 students graduate in Saxony-Anhalt as part of the double graduation cohort, for which we cannot distinguish the treatment status in the data. As more and more states introduce G12, the mean graduation age for the first-adopting states diverges from the parallel movement prior to the G12 reform. The

---

[13]The notion of autocorrelation in our data is also supported by Wooldridge's (2002) test for serial correlation in panel-data.

[14]The first adopting states are Saxony-Anhalt (ST), Mecklenburg-Vorpommern (MV), Saarland (SL), Hamburg (HH) and Bavaria (BY). Never-changers are Saxony and Thuringia with graduation after 12 years as well as Schleswig-Holstein and Rhineland-Palatinate with graduation after 13 years.

reform's principal objective was to reduce the age at which students graduate from school. The descriptive inspection reveals that the graduation age has evidently decreased.

Panel B presents descriptive results for grade repetition rates. Prior to the reform, the time trend of first-adopting and never-changing states appears to be parallel, with the exception of 2002.[15] When the reform is gradually introduced, grade repetition rates diverge between first-adopting and never-changing states. This indicates an increase in overall grade repetition rates with the introduction of G12.

Panel C shows how the share of *Abitur* graduates evolves over time. Again, it can be seen that both first-adopting and never-changing states follow the same time trend before the reform introductions. As the reform is gradually introduced in the first-adopting states, there seems to be no adverse effect of the G12 reform on the share of high school graduates.

6.2 Estimation results

The graphical findings are supported by the estimation results of the G12 reform effect, reported in table 2. Panel A provides the G12 reform effect estimates on the graduation age. The baseline difference-in-differences specification in column (1) only includes state, time and double graduation cohort fixed effects and suggests a reduction in the mean graduation age by 0.82 years. In model (2), we control for factors that characterise the economic environment in which students obtain their education (X_{sc}). This specification yields a very similar point estimate. In column (3), we further include dummy indicators for other reforms (P_{sc}). Across all specifications, we test whether the reform achieves its potential to reduce the graduation age by a full year. There is strong statistical evidence against the hypothesis that $\beta = -1$, suggesting that the reform stays behind its potential of one year.

One possible explanation why the reform stays behind is increased grade repetition. Panel B summarises the results for grade repetition rates across the three different models. In our preferred specification (3), the G12 reform increases the probability of repeating a grade at high school by 3.08 percentage points, or 21 percent.[16] This effect is highly significant and very similar across all three specifications and can provide an explanation why the reform stays behind its potential in reducing the high school graduation age.

---

[15]In a robustness analysis, we remove the graduation cohort 2002 from our sample. The main results are identical.

[16]Calculations of percentage changes are based on the following equation: %-change$= \frac{\hat{\beta}}{E(y|G12=1)-\hat{\beta}} \cdot 100$

The estimation results presented in panel C suggest that G12 leaves the graduation share unaffected. Across all three model specifications, the same picture emerges. The effect estimates are very small in magnitude and statistically insignificant.

7   Robustness analyses

In this section, we perform various robustness checks, which we summarise for our three outcome variables in panels A to C of table 3. All regressions are based on our main specification described by equation (3). 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 (Ashenfelter, 1978). 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 including more intense competition for post-schooling options. 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.

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. Again, we include a dummy variable to remove these cohorts from the treatment group.

The next specification in column (4) incorporates a dummy for the first three G12 cohorts in Saxony-Anhalt and the first two G12 cohorts in Mecklenburg-Vorpommern as they were already in grade 7 to 9 and grade 7 to 8, respectively, when they were assigned to G12. Consequently, the additional workload has been distributed over even fewer remaining years for these surprised cohorts, which could affect grade repetition rates and school drop outs exceptionally. In column (5), we incorporate the available observations from the state of Lower Saxony, which we excluded from our main sample as we lack complete information for all outcome variables. Column (6) and column (7) examine the sensitivity of the results to different ways of accounting for the double graduation cohort. First, we exclude them completely. We lose one cohort in each of the ten treatment states. Second, we alternatively include a set of state-specific double graduation cohort dummies.

The treatment effect identification in our quasi-experimental analyses rests on the assumption of a common time trend in the outcome variables between treatment and

control states in the absence of treatment, conditional on the control variables. It cannot be tested directly. The visual inspection of treatment and control groups' trends in the pretreatment period suggests a deviation from parallel time trends in 2002. The specification in column (8) excludes the year 2002 from the estimations. In column (9), we substitute the economic and education reform 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. Neither redefinitions of the treatment and control groups, nor different ways of dealing with the double graduation cohorts and with the common trend assumption change the drawn conclusions.

The final two specifications in table 3 deal with alternative assumptions about the error term. In column (10), the AR(1) process of the error term is now allowed to vary by state s, such that in equation (4) ρ is replaced by ρ_s. In column (11), we employ a standard OLS procedure with standard errors clustered at the state level. Our preferred FGLS estimation strategy is more efficient than OLS with clustered standard errors if the error terms follow indeed an AR(1) process. Consistent with this argument, standard errors increase considerably in panels B through D, but the effects on graduation age and grade repetitions are of similar magnitude and remain statistically significant. This confirms evidence from Monte-Carlo simulations provided by Brewer et al. (2013), who argue that the main problem of clustered standard errors in difference-in-differences settings is the low statistical power to detect existing effects.[17]

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. We include respective placebo policy indicators one at a time in our main regression model. The results for placebo treatments are reported in table 4. All coefficients 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 the absence of compositional changes in treatment and control groups. We check for

---

[17]We also performed wild cluster bootstrap procedures as there is a discussion on the appropriateness of clustered standard errors when the number of clusters is small (Donald and Lang, 2007; Cameron et al., 2008). These procedures provide p-values that are very similar to OLS results with clustered standard errors and, hence, confirm our conclusions.

changes in the group composition by looking at reform effects on the share of 7-graders enrolled to the high school track. This captures changes in the school track choice and cross-border migration prior to grade 7 induced by G12. 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 and, hence, there is no evidence for compositional changes.

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 work with alternative normalisations. However, none of the alternative normalisations provides evidence for an adverse effect of G12 on graduation rates. In summary, the policy effects on all outcome variables are strikingly robust and support the main findings.

8   Effect heterogeneities

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

8.1 By gender

A broad, interdisciplinary literature documents gender gaps in school achievements.[18] We expect such gender differences in the G12 treatment effects for two reasons. First, as boys perform more poorly in school than girls on average, 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 organising the tighter time schedule (Spinath, 2014).

Table 6 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 refer separately to females and males. As expected, males exhibit higher absolute increases in grade repetition rates than females. The reduction in the mean graduation age is stronger for females than for males, which is in line with the

---

[18]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. Poropat, 2009; Cornwell et al., 2013; Spinath, 2014).

hypothesis that grade repetitions attenuate the reform effect on the mean graduation age. Finally, there is no adverse effect on the graduation rates for neither females nor males.

8.2 Over time

In the evaluation of the G12 reform effect it is important to distinguish between temporary and lasting reform effects. Table 7 reports the reform effect separately by years passed since the implementation of the reform for the three main outcome variables.

The estimations are based on the main regression equation (3), in which we now interact the G12 indicator with dummy variables for each year after the reform implementation. The estimates reported in previous sections are weighted averages across treatment states and across different treatment durations. States that implemented the reform earlier receive a higher weight as they have more post-treatment observations. To isolate the over-time changes in the treatment effect from potential effect heterogeneities across states, we additionally impose different sample restrictions.

In the first column of table 7, the treatment effect calculations only consider treatment observations for the three states that implemented the reform first. For these states, we have complete information for four graduation cohorts following the double cohort such that we can identify separate treatment effects for each of these four cohorts. Further treatment observations are removed from the sample. Column (2) adds the state that introduced G12 one year later. As fewer treatment cohorts are observed for this state, we can only identify separate treatment effects for up to three cohorts following the double graduation cohort. In the same vein, column (3) considers another state that introduced the reform next.

The results of the effect on the graduation age suggest some convergence of the point estimates towards −1, i.e. towards the reform's potential to reduce the mean graduation age by one full year. However, none of the treatment estimates is statistically different over time and all estimates are statistically different from −1. There is also no evidence for a decline of the effect on repetition rates over time. For the first three treatment states in column (1), the effects are imprecisely estimated and rather suggest an increase of the effect over time, which is not statistically significant. Adding more states to the sample increases the precision, but there is still no sign of a fading out of the treatment effect. For graduation rates, there is no evidence that the average effect estimated in the previous section hides heterogeneities over time. Taken together, table 7 shows that the estimated reform effects are not only temporary.

8.3 By grade level

The identification strategy also allows us to look at effect heterogeneities for grade repetitions across different grade levels. We adjust equation (3) in the following way:

$$y_{scg} = (G12_{sc} \cdot grade_g)' \cdot \beta_g + \mu_s + \kappa_c + \gamma_g + \pi_{sg} + \nu_{cg} + \delta \cdot DGC_{sc} + X'_{scg} \cdot \lambda + P'_{scg} \cdot \phi + \varepsilon_{scg} \quad (5)$$

Now, y_{scg} denotes the fraction of students repeating grade level g in state s in cohort c. We interact the $G12$ indicator with dummies for each grade level, $grade_g$, such that β_g captures the reform effect estimate for grade level g. μ_s, κ_c, DGC_{sc}, X_{scg}, and P_{scg} are defined as before. We further include grade fixed effects (γ_g), grade-state fixed effects (π_{sg}) and grade-cohort fixed effects (ν_{cg}).

Analysing grade level effect heterogeneities requires to take a decision regarding the institutional particularity that between grade 7 and graduation there are seven grades for G13 students but only six grades for G12 students. For a grade-to-grade comparison we therefore have to drop one grade under the G13 regime. We report grade-specific results for two different grade comparisons due to the following considerations. Grade 11 played a special role in the G13 regime (*introductory phase*, denoted by I). It introduced students to selective course schemes, allowed for time to revise material and aimed at preparing students for the last two years, the *qualification phase*. Grade 10 in the G12 regime only partly fulfils these functions. It is therefore unclear whether it should be compared to grade 10 or grade 11 under G13. Hence, we exercise the former comparison in first two columns (grade 11 is dropped under G13) and the latter in the last two columns (grade 10 is dropped under G13).

Grades 7-9 under G12 are always compared to grades 7-9 under G13. We also always compare grade 12 under G13 to grade 11 under G12 (the first year of the *qualification phase*, denoted by Q1) and grade 13 under G13 to grade 12 under G12 (the second year of the *qualification phase*, Q2). In these two years, students can drop certain courses and specialise in others. Earned grades count towards the final GPA.

Table 8 reports the G12 reform effects on grade repetition rates for each grade level, as outlined in equation (5). In columns (1) and (3), we consider cohorts graduating between 2002-2013 with separate information for six grade levels. The previous sample size of $N = 168$ increases to $N \cdot 6 = 1008$ observations. We find no significant effects of the G12 reform on grade repetition rates in grades 7-9, and highly significant effects in the final three years at high school. Compared to the share of students that would have repeated

these grade levels in the absence of the reform, the effects amount to an increase in grade repetitions in the final two years by 69 to 85 percent. The relative increase in grade 10 varies between 27 and 67 percent, depending on the comparison group.

Our data on the school years 1994/95-2012/13 also includes information on cohorts that have not been fully observed, i.e. grade repetitions in higher grade levels for cohorts graduating before 2002 and grade repetitions in lower grade levels for cohorts graduating after 2013. In columns (2) and (4), we use also this information. The sample size increases, while the findings are almost identical.

Note that the grade level effects on grade repetition rates do not correspond to the shift in workload. Rather, the increase is strongest in the final years in which exam results count towards the final GPA. Together with the zero-effect on graduation rates, it seems that some of the effect on grade repetitions is also driven by individual decisions to repeat a grade level voluntarily, i.e. to either revise subject choices, to spend a school year abroad or to improve on the GPA.

9 Discussion and conclusion

Policy-makers in industrialised countries face a dilemma in designing education systems. Equipping individuals with high levels of education takes time, while ageing populations require early labour market entries in order to fund social security systems and to mitigate the resulting skilled worker shortages. Lowering the school starting age is one possible way to resolve this problem. Another alternative is to compress the total instructional time into fewer years of schooling. Germany chose the latter, reducing the time spent in high school by one year such that children now take 12 years to earn the university entrance qualification, the *Abitur*. Little is known about the reform impact in the literature.

Using data from the German Federal Statistical Office that covers all students of the 2002-2013 graduation cohorts, we study indicators of the overall effectiveness of this reform and extend the literature in several ways. First, we pioneer in studying the effect of this fundamental reform on its major objective, the reduction in the school leaving age. Second, we investigate the reform effect on the quantitative composition of the human capital stock, examining high school graduation rates. Third, we analyse the impact on grade repetitions and generate important insights about students' ability to cope with the increased learning intensity, about additional costs of the reform and about group

20

compositional changes that are important for future analyses of the G12 reform.

We exploit the regional and temporal variation in the implementation of the reform with difference-in-differences estimations, finding that the policy change reduces the graduation age by about 10 months. The reform stays behind its potential of reducing the graduation age by a 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 about 3 percentage points, or 21 percent. The effects are concentrated in the final years prior to graduation, where repetition rates increase by 69 to 85 percent. 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 for males in response to the G12 reform. The differential effect on grade repetitions across grade levels does not match the shift in workload, which focuses on lower grade levels. Grade repetitions may also capture individual considerations to repeat a grade level voluntarily and may therefore imperfectly proxy for students' ability to cope with the increased learning intensity. There is no evidence that the effects on grade repetitions and graduation age quickly fade out over time. Also, we cannot find evidence for adverse effects of the G12 reform on *Abitur* graduation rates. Finally, we do not find evidence for a reform impact on the number of students enrolling to high school in grade 7. All estimates prove to be stable in a broad range of robustness checks. Our results on grade repetitions are qualitatively similar to Pischke's (2007) findings on compressed instructional time due to short school years in Germany in 1966-1967.

The increase in grade repetitions caused by G12 mitigates the reform effect on lowering the graduation age and induces additional costs to the education system that policy-makers need to consider.[19] To provide a rough estimate of the magnitude of these additional expenditures, consider the following back-of-the-envelop calculations: Taking our point estimate of 3 percentage points at face value and assuming a homogeneous treatment effect for the G12 students in the double graduation cohort and for the two excluded states Lower-Saxony and Hesse, our estimates suggest that the G12 reform induced about 7,500 students in the graduation cohort 2013 to repeat a grade. Over all reform states, this translates into total additional expenditures of 41 to 51 million Euros for the 2013 graduation cohort, depending on whether we assume the same annual expenditures per

---

[19]Note that the reform keeps the number of instruction hours constant. Therefore, the states' expenditures for teachers are assumed to be constant.

grade repeater as for the average high school student (6,800 Euros, German Federal Statistical Office, 2014) or whether we consider grade repeaters with only 80 percent of that value, as suggested by Klemm (2009). Assuming constant effects over time and across federal states, the increased rate of grade repetitions suggests an increase in public expenditure for German high schools by about 2.4-3 percent per year. Further costs of grade repetitions may accrue at the individual level (Eide and Goldhaber, 2005).

Our results point out that compressing the instructional time into fewer years of schooling can decrease the school leaving age without adverse effects on the economy's endowment with high school graduates. However, further research is needed to investigate whether the reform impacts the quality of high school graduates. Our results also show that the vast majority of students graduates more quickly under G12. Some students cope with the increased learning intensity by repeating a grade, which might also happen voluntarily. However, taking the evidence on the three outcomes together, we conclude that for other countries policies like G12 can constitute an interesting alternative to reductions in the school starting age, when ageing populations require to deal with the important trade-off between high levels of education and early labour market entries.

References

Ashenfelter, O. (1978). Estimating the effect of training programs on earnings. *Review of Economics and Statistics*, 60(1):47–57.

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

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.

Bedard, K. and Dhuey, E. (2006). The persistence of early childhood maturity: International evidence of long-run age effects. *Quarterly Journal of Economics*, 121(4):1437–1472.

Bedard, K. and Dhuey, E. (2012). School-entry policies and skill accumulation across directly and indirectly affected individuals. *Journal of Human Resources*, 47(3):643–683.

Bellei, C. (2009). Does lengthening the school day increase students' academic achievement? Results from a natural experiment in Chile. *Economics of Education Review*, 28(5):629–640.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics*, 119(1):249–275.

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. *Journal of Human Resources*, 30(4):pp. 741–765.

Black, S. E., Devereux, P. J., and Salvanes, K. G. (2011). Too young to leave the nest? The effects of school starting age. *Review of Economics and Statistics*, 93(2):455–467.

Bologna Declaration (1999). The Bologna Declaration of 19 June 1999: Joint declaration of the European Ministers of Education. *http://www.magna-charta.org/resources/files/BOLOGNA_DECLARATION.pdf*.

Börsch-Supan, A., Härtl, K., and Ludwig, A. (2014). Aging in Europe: Reforms, international diversification, and behavioral reactions. *American Economic Review*, 104(5):224–229.

Brahm, G. (1997). The effect of national standards and curriculum-based exams on achievement. *American Economic Review*, 87(2):pp. 260–264.
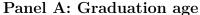
Brewer, M., Crossley, T. F., and Joyce, R. (2013). Inference with difference-in-differences revisited. *IZA Discussion Paper*, 7742.

Brugiavini, A. and Peracchi, F. (2005). The length of working lives in Europe. *Journal of the European Economic Association*, 3(2-3):477–486.

Büttner, B. and Thomsen, S. L. (2015). Are we spending too many years in school? Causal evidence of the impact of shortening secondary school duration. *German Economic Review*, 16(1):65–86.

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

Cameron, C., Gelbach, J. B., and Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, 90(3):414–427.

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. (2013). Economic uncertainty, parental selection, and the criminal activity of the 'Children of the Wall'. *IZA Discussion Paper*, 7712.

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.

Donald, S. G. and Lang, K. (2007). Inference with difference-in-differences and other panel data. *Review of Economics and Statistics*, 89(2):221–233.

Dustmann, C., Puhani, P. A., and Schönberg, U. (2014). The long-term effects of early track choice. *IZA Discussion Paper*, 7897.

Eide, E. R. and Goldhaber, D. D. (2005). Grade retention: What are the costs and benefits? *Journal of Education Finance*, 31(2):195–214.

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.

Fredriksson, P. and Ockert, B. (2014). Life-cycle effects of age at school start. *Economic Journal*, 124(579):977–1004.
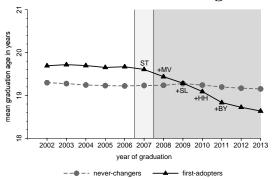
Gaini, M., Leduc, A., and Vicard, A. (2013). School as a shelter? School leaving-age and the business cycle in France. *Annales d'Economie et de Statistique*, 111-112:251–270.

German Federal Statistical Office (2014). *Bildungsausgaben: Ausgaben je Schülerin und Schüler 2011.* Statistisches Bundesamt, Wiesbaden.

German Federal Statistical Office (2015a). Allgemeinbildende Schulen: Fachserie 11, Reihe 1. *https://www.destatis.de/DE/Publikationen/Thematisch/BildungForschungKultur/ Schulen/BroschuereSchulenBlick.html.*

German Federal Statistical Office (2015b). Genesis-Online Datenbank. *https://www-genesis.destatis.de.*

Goldstein, J. R. and Kreyenfeld, M. (2011). Has East Germany overtaken West Germany? Recent trends in order-specific fertility. *Population and Development Review*, 37(3):453–472.

Görlitz, K. and Gravert, C. (2015). The effects of increasing the standards of the high school curriculum on school dropout. *Free University Berlin Discussion Paper, School of Business & Economics*, 1.

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.

Herdin, G. and Hachmeister, C.-D. (2014). Der CHE Numerus Clausus-Check 2013/2014. Eine Analyse des Anteils von NC-Studiengängen in den einzelnen Bundesländern. *CHE Arbeitspapiere*, 178:1–35.

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.

Klemm, K. (2009). *Klassenwiederholungen - teuer und unwirksam.* Bertelsmann Stiftung, Gütersloh.

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.

Lochner, L. (2011). Nonproduction benefits of education: Crime, health, and good citizenship. In *Handbook of the Economics of Education*, volume 4, pages 183–282. Elsevier B.V., Amsterdam.

Machin, S., Salvanes, K. G., and Pelkonen, P. (2012). Education and mobility. *Journal of the European Economic Association*, 10(2):417–450.

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).

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 (2000). *Education at a Glance 2000*. Education at a Glance. OECD Publishing, Paris.

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.

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

Puhani, P. A. and Weber, A. M. (2007). Does the early bird catch the worm? Instrumental variable estimates of early educational effects of age of school entry in Germany. *Empirical Economics*, 32(2-3):359–386.

Scheubel, B., Schunk, D., and Winter, J. (2013). Strategic responses: A survey experiment on opposition to pension reforms. *Scandinavian Journal of Economics*, 115(2):549–574.

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

Thiel, H., Thomsen, S. L., and Büttner, B. (2014). Variation of learning intensity in late adolescence and the effect on personality traits. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 177(4):861–892.

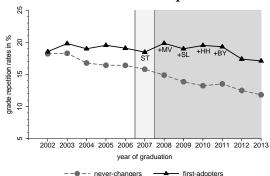Webbink, D. (2007). Returns to university education: Evidence from a Dutch institutional reform. *Economica*, 74(293):113–134.

Woessmann, L. (2003). Schooling resources, educational institutions and student performance: The international evidence. *Oxford Bulletin of Economics and Statistics*, 65(2):117–170.

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data.* MIT Press, Cambridge.

Figures

Panel A: Graduation age

Panel B: Grade repetition

Panel C: Graduation rates

Figure 1: Trends in outcome variables in first-adopting 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 combines information on the last pre-treatment and first treatment observation. The grey area constitutes the treatment period. The light-grey area indicates that none of the first-adopters fully moved to G12, as the first treatment state of Saxony-Anhalt is only included with an observation for the double graduation cohort.

Tables

Table 1: Implementation of G12 and other education reforms in the federal states

	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 (in sample period)					
Rhineland-Palatinate	none	all	none	none	from 2011
Schleswig-Holstein	from 2016	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 how the graduation cohorts in our sample period 2002-2013 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.

Table 2: Reform effects - main results

	Baseline DiD (1)	Economic controls (2)	Main specification (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	3.18***	3.22***	3.08***
	(1.16)	(0.75)	(0.94)
$\% - change$	22.11	22.49	21.32
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
Double graduation cohort dummy	Yes	Yes	Yes
Economic controls	No	Yes	Yes
Education reform controls	No	No	Yes

Notes: The table reports the effect of the G12 reform, which compresses high school instructional time by one year, for the cohorts graduating 2002 through 2013 on the mean age at high school graduation, the overall grade repetition rates and the high school graduation rate. All models include fixed effects for state, year and the double graduation cohort. Economic control variables are the federal state's GDP growth, the general unemployment rate and the youth unemployment rate of 20-25 years old, all measured in the year prior to graduation, and a dummy indicator for the generations born in East German shortly after the fall of the wall. Policy controls are dummies for other education policies as described in table 1. Standard errors allow for heteroskedasticity, contemporaneous correlations across states and within-state-autocorrelation of the error terms. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 3: Robustness I - Sensitivity checks

	Main	Redefining the treatment group				Double cohort		Common trend		Standard errors	
		w/o penult. G13 cohort	w/o first 2 treatment	w/o surprised	+ Lower Saxony	w/o DGC	Interacted DGC	w/o 2002	State time trend	Panel-spec. AR(1)	OLS + cluster
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(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	3.08***	3.47***	3.64***	2.72***	2.96***	2.98***	2.84***	2.95***	2.62***	3.47***	3.58**
	(0.94)	(0.98)	(1.11)	(1.00)	(1.10)	(0.96)	(0.88)	(0.90)	(0.84)	(0.87)	(1.53)
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, the overall 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). * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

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	0.76	0.49	0.68	0.50	-0.58
	(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, the overall 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 one placebo reform indicator for which the coefficient is reported in this table. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

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. 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. As the data lacks information on school entries for the 2002 cohort, the sample size drops by one observation per state. 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). * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 6: Reform effect by gender

	Females (1)	Males (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 on the mean age at high school graduation, the overall grade repetition rates and the high school graduation rate separately by gender. The dependent variable is gender specific. All models are based on the main specification, see table 2 column (3). * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 7: Reform effect over time

	Effect identifying states		
	First 3 states (1)	First 4 states (2)	First 5 states (3)
Panel A: Graduation age			
G12 · 1 year after	-0.85***	-0.87***	-0.88***
	(0.04)	(0.04)	(0.03)
G12 · 2 years after	-0.86***	-0.89***	-0.90***
	(0.05)	(0.04)	(0.04)
G12 · 3 years after	-0.84***	-0.90***	—
	(0.05)	(0.04)	
G12 · 4 years after	-0.92***	—	—
	(0.05)		
N	132	132	130
Panel B: Grade repetition			
G12 · 1 year after	1.75	3.78**	2.86**
	(1.61)	(1.51)	(1.21)
G12 · 2 years after	2.83	4.29***	3.42**
	(2.05)	(1.65)	(1.37)
G12 · 3 years after	2.95	4.25**	—
	(2.02)	(1.70)	
G12 · 4 years after	3.06	—	—
	(1.88)		
N	144	144	142
Panel C: Graduation rate			
G12 · 1 year after	-0.06	-0.16	0.40
	(1.00)	(1.12)	(0.70)
G12 · 2 years after	-1.80	-0.58	0.02
	(1.17)	(1.22)	(0.75)
G12 · 3 years after	-1.33	-0.96	—
	(1.10)	(1.38)	
G12 · 4 years after	0.74	—	—
	(1.10)		
N	144	144	142

Notes: The table reports the effect of the G12 reform on the various outcomes for different years after the implementation of the reform. The column headers indicate the states that contribute to the identification of the treatment effects, all other treated observations are excluded. All models are based on the main specification, see table 2 column (3). * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 8: Reform effect by grade level

| | Grade 11 dropped under G13 | | Grade 10 dropped under G13 | |
| | Cohorts 2002-13 | All years | Cohorts 2002-13 | All years |
	(1)	(2)	(3)	(4)
G12 · grade 7	0.01	-0.26	0.02	-0.24
	(0.27)	(0.24)	(0.26)	(0.23)
G12 · grade 8	0.29	0.11	0.34	0.14
	(0.31)	(0.27)	(0.31)	(0.26)
G12 · grade 9	0.14	0.03	0.17	0.05
	(0.26)	(0.24)	(0.26)	(0.23)
G12 · grade 10/I	1.27***	1.15***	0.68*	0.62*
	(0.34)	(0.30)	(0.37)	(0.33)
G12 · grade Q1	2.48***	2.28***	2.51***	2.32***
	(0.49)	(0.47)	(0.47)	(0.45)
G12 · grade Q2	1.01***	0.98***	1.02***	1.02***
	(0.32)	(0.33)	(0.30)	(0.30)
N	1008	1451	1008	1459
%-change grade 7	0.78	-14.44	1.29	-13.50
%-change grade 8	11.81	5.12	14.10	6.48
%-change grade 9	5.82	1.05	6.86	1.90
%-change grade 10/I	66.67	57.05	27.39	24.62
%-change grade Q1	68.85	62.32	70.22	64.20
%-change grade Q2	82.81	79.74	84.90	84.96

Notes: The table reports the effect of the G12 reform on grade repetition rates separately by grade levels. All models include fixed effects for state, cohort, grade, state-grade, cohort-grade, and double graduation cohort as well as time-varying economic control variables and education reform dummies. Standard errors allow for heteroskedasticity within state-grade panels, contemporaneous correlations across state-grade panels and within-state-grade-autocorrelation of the error terms.* $p<0.1$, ** $p<0.05$, *** $p<0.01$.