Courant Research Centre

‘Poverty, Equity and Growth in Developing and Transition Countries: Statistical Methods and Empirical Analysis’

Georg-August-Universität Göttingen
(founded in 1737)

Discussion Papers

No. 163

The Effects of Delayed Tracking: Evidence from German States

Simon Lange, Marten von Werder

December 2014

Wilhelm-Weber-Str. 2 · 37073 Goettingen · Germany
Phone: +49-(0)551-3914066 · Fax: +49-(0)551-3914059

Email: crc-peg@uni-goettingen.de Web: http://www.uni-goettingen.de/crc-peg

The Effects of Delayed Tracking: Evidence from German States*

Simon Lange† Marten von Werder‡

December 19, 2014

Abstract

Germany's education system stands out among OECD countries for early tracking: students are tracked into different secondary school types at the age of ten in most German states. In this paper we estimate the effects on educational outcomes of a reform that delayed tracking by two years. While our findings suggest that the reform had no effect on educational outcomes on average, we find a positive effect on male students with uneducated parents and a negative effect on males with educated parents. The reform thus increased equality of opportunity among males, yet not among females. We argue that the gendered pattern is best explained by developmental differences between boys and girls at the relevant age.

Keywords: tracking; educational institutions, intergenerational mobility.

JEL Classification Numbers: I21, I24, I28, J62.

*The authors would like to thank Ronny Freier, Stephan Klasen, Malte Reimers, Ramona Rischke, Viktor Steiner, Sebastian Vollmer, and seminar participants in Berlin and Göttingen for valuable comments on earlier versions of this paper.

†Corresponding author. Economics Department, University of Göttingen. Platz der Göttinger Sieben 3, 37073 Göttingen, Germany. E-mail: simon.lange@wiwi.uni-goettingen.de.

‡Economics Department, Free University of Berlin.

1 Introduction

The term *tracking* in the context of education systems refers to the practice of grouping students in some way by ability (Betts, 2010). Countries differ widely in the way they track students but almost all education systems in practice involve some degree of tracking. For instance, countries differ in the age at which students are tracked and in whether tracking occurs within schools (i.e. sorting students into different classrooms as in the United States and Canada) or across schools (i.e. sending students to different types of schools as in some European countries). Proponents of tracking argue that the creation of more homogeneous classes increases efficiency by allowing educators to tailor lessons to students' specific needs. Opponents, on the other hand, fear that misclassification of students is often rife—especially when students are tracked at an early age—and that tracking aggravates initial differences.

In comparison to other OECD countries, Germany's education system mostly tracks students across schools and at an early age. Primary school completion usually takes only four years, after which students are sorted into three different secondary tracks. At the same time, parents' socio-economic status is an important predictor of educational achievement in Germany—more important than in other OECD countries (e.g. Baumert and Schümer, 2001; Baumert et al., 2003b; OECD, 2003; Dustmann, 2004; Schütz et al., 2008). This observation has resulted in an on-going public debate about equity and efficiency within Germany's education system.[1]

In this paper we exploit a policy experiment that delayed tracking by two years in Lower Saxony, one of Germany's federal states, to study the effect of delayed tracking on educational outcomes. Between 1972 and 1982, Lower Saxony's government gradually introduced an independent type of school, the "Orientierungsstufe" (English: *orientation stage*; henceforth, OS), that would serve as an intermediate school between primary and lower secondary. While students were sent on to a different school upon completing fourth grade, tracking occurred only after sixth grade. We investigate the effects of this policy on educational outcomes based on a differences-in-differences (DD)-framework, comparing changes in outcomes between cohorts across states. The dataset we employ, the German Socio-Economic Panel (GSOEP), contains information on parents' educational attainment, allowing us to investigate the effects separately by parental background.

Our findings indicate that the reform has had a positive effect on the number of years in education for males with uneducated parents and a negative effect for males with educated parents. Both effects appear important economically but only the former result is statistically significant and robust to changing the econometric specification and the definition of our parental background variable. A similar pattern of effects is observed when we investigate the impact on the probability of attaining *Abitur*, the highest school leaving-certificate in the German system, and the probability of obtaining a university degree, although our estimates are only significant

[1] See, for instance, recent efforts to reform the school system in the city-state of Hamburg that eventually failed in a referendum in 2010: http://www.spiegel.de/schulspiegel/wissen/schwarz-gruenes-hamburg-saegen-an-der-schulreform-a-599904.html (in German).

and robust in the latter case for males with uneducated parents. There is also some weak evidence that the initially negative effect on outcomes for males with educated parents dissipates over time.

We find little evidence for an effect on females: while our estimates of the effect of the reform on the probability of attaining a tertiary degree for females with educated parents is negative and economically meaningful, it is not significant at conventional levels. Our findings thus suggest an adverse effect of early tracking on equality of opportunities, mainly among males. While this finding is consistent with both peer effects *and* an improvement in the allocation of students to tracks associated with de-tracking, we argue that the latter better explains the gendered pattern in our findings.

There is only a limited number of studies that investigate the effects of actual reforms relying on variation across time and space, notably Meghir and Palme (2005), Pekkarinen et al. (2009), Malamud and Pop-Eleches (2011), Hall (2012), and Kerr et al. (2013).[2] As argued by Betts and Shkolnik (2000) Pischke and Manning (2006), and Waldinger (2007), alternative identification strategies that rely on cross-sectional variation often plausibly suffer from bias resulting from selection-on-unobservables or omitted variables.

The German reform we study offers a particularly well-suited environment to study the effects of tracking: education policies in Germany are traditionally the responsibility of the states while the federal government is in charge of most other policy areas that may affect educational outcomes. Hence, our analysis is less likely to be subject to differences in trends in socio-economic and institutional factors that potentially confound the analysis.[3] Because of Germany's system of equalization payments between states and sharing of tax revenues, there is very limited potential for different trends in resources allocated to public education between states. Also, private education institutions that would potentially mitigate the effect of policy reforms aimed at improving inter-generational mobility traditionally play a negligible role in Germany's education system.

The remainder of the paper is organized as follows. The next section describes the German education system and the process that finally led to the introduction of OS schools in Lower Saxony during the 1970s. Section 3 reviews the literature on tracking and thus lays out some plausible channels through which tracking may affect educational outcomes. Section 4 describes the data and details our identification strategy. Section 5 presents results from estimating the effect of the reform on educational outcomes of different subgroups. Section 6 offers an interpretation of our results. Section 7 concludes.

[2] While useful, evidence from randomized control trials (RCTs) (e.g. Duflo et al., 2011) may have little external validity compared to observational studies of actual policy reforms in this context. Tracking policies typically involve exposure of students to very different curricula, an aspect that is difficult to implement in RCTs (Betts, 2010).

[3] As argued by Waldinger (2007), this may often be an issue in cross-country comparisons such as Hanushek and Woessmann (2006). See, however, Woessmann (2010).

2 Background: Germany's education system

2.1 Tracking in Germany's education system

Most states in Germany track students into three different types of lower secondary schools at the end of fourth grade when students are about ten years old:[4] the *Hauptschule* is a secondary school track for low-achieving primary students. The leaving certificate, awarded after five years of schooling, qualifies graduates to enroll in upper secondary vocational training courses, that is, apprenticeships in the dual-system (*Ausbildung*). Students with about average marks from primary education usually attend the *Realschule*, the intermediate secondary track. After another six years of schooling students are eligible to choose from an extended set of apprenticeships within the dual-system of vocational training. The *Gymnasium* is the third secondary school type and enrollment is recommended only to highly-achieving students. This school type is the only secondary school track that awards after eight to nine years of schooling the *Abitur*, the most prestigious school leaving certificate, which permits students to enroll at a university. Henceforth, we will refer to these three tracks as lower and upper vocational track and the academic track, respectively.

The streaming procedure varies across states but usually involves teachers formally recommending a secondary school track to parents based on their child's performance. In ten out of 16 states parents have the final say about the placement of their child, whereas in the remaining six states recommendations are binding yet parents have the right to let their child take an entry exam or attend test lessons.

Tracks differ in several respects in terms of the quality and quantity of inputs (see Dustmann et al., 2014): teachers in the academic track are better paid. Their university degree differs in terms of requirements and is more subject-oriented. It also typically takes one additional year to complete. Students in the academic track cover more topics and more advanced topics each year and they are often required to attend more hours per week.

Tracking across schools may be inconsequential if the education system would exhibit a high level of permeability. In principle, German students are allowed to switch between tracks at any time if their academic records justify such a step. However, research on the topic suggests that switching between tracks prior to completion is rare. Mühlenweg (2008), for instance, examines administrative data from Hesse and reports for the school years 2003/2004 and 2005/2006 that more than 96 percent of students in grades five through seven remain in their initial track. Similar results are reported by Dustmann et al. (2014) who find that only two percent of students in the states of Bavaria and Hesse change track. Avenarius et al. (2001) present similar numbers for Lower Saxony. Moreover, Schnepf (2002) points out that students in the academic or upper vocational track are more likely to shift to the lower track than vice versa.

Upgrading upon completion of one of the two vocational tracks is, however, quite common and

[4] Exceptions include Berlin and Brandenburg, in which elementary school comprises six grades, and Mecklenburg-West Pomerania, which tracks students after sixth grade but was part of the German Democratic Republic.

is often cited as proof for the permeability of Germany's education system (e.g. Dustmann et al., 2014). Students in the lower vocational track may switch to the upper vocational track or stay on for an additional year in order to obtain the school leaving-certificate awarded upon completion of the upper vocational track. Students in the upper vocational track, in turn, have the possibility to continue schooling in the academic track. Alternatively, they qualify to attend specialized academic track-schools that often have a special focus on a particular subject. Completion of such a specialized academic-track school allows them to apply to universities, although the range of subjects from which they may choose may be limited.

2.2 The introduction of OS schools in Lower Saxony

The roots of Germany's three-tiered education system can be traced back at least to developments in the 19th century. Efforts to reform this system were evident in the early 20th century, but were not realized until the post-war decades. In the mid-1950s, both Lower Saxony and Hesse started experimenting with less stringent forms of tracking in few selected schools. The goal at the time was to improve the selection of students into academic careers. These efforts, however, were terminated in Lower Saxony in 1964 (Schuchart, 2006). 1953 saw the creation of the *Deutscher Ausschuss für das Erziehungs- und Bildungswesen* (*engl.: German Advisory Board for the Educational System*) as an independent body tasked with giving advice on educational reform. Six years later, the committee formally recommended a prolongation of comprehensive schooling until sixth grade for most students that would also involve a degree of within-class tracking in major subjects (Deutscher Ausschuss, 1959). This recommendation, however, had no effect on policies at the time (von Friedeburg, 1992). Only Hesse introduced schools that closely resembled these recommendations but would also retain schools in the traditional three-tiered system.

A follow-up body of the *Ausschuss*, the *Deutscher Bildungsrat* (*engl.: German Education Council*), was created in 1965. The Council presented a blueprint for structural reforms that first mentioned the introduction of an *Orientierungsstufe* in 1970 (Deutscher Bildungsrat, 1970). In comparison to earlier plans, the focus was on pronounced tracking within schools (i.e. across subjects) and the dissemination of information about possible future careers to students and their parents (Schuchart, 2006). OS schools were supposed to be completely independent of schools in the three-tiered system.

In 1974, however, it became clear that this proposal would not be approved by a majority of states. In particular, states that were at the time governed by the center-right *Christian Democratic Union* were reluctant to de-couple the two grades from the three-tiered system. While the final document was entitled *Vereinbarung zur Orientierungsstufe* (*engl.: Agreement on the Orientation Stage*) and all states indeed agreed in principle to changes to the school system that were supposed to take effect in 1976, the compromise would leave the decision over whether or not to delay tracking to the states. Therefore, in subsequent years, the term *Orientierungsstufe* was applied to what were in effect very different school systems and the

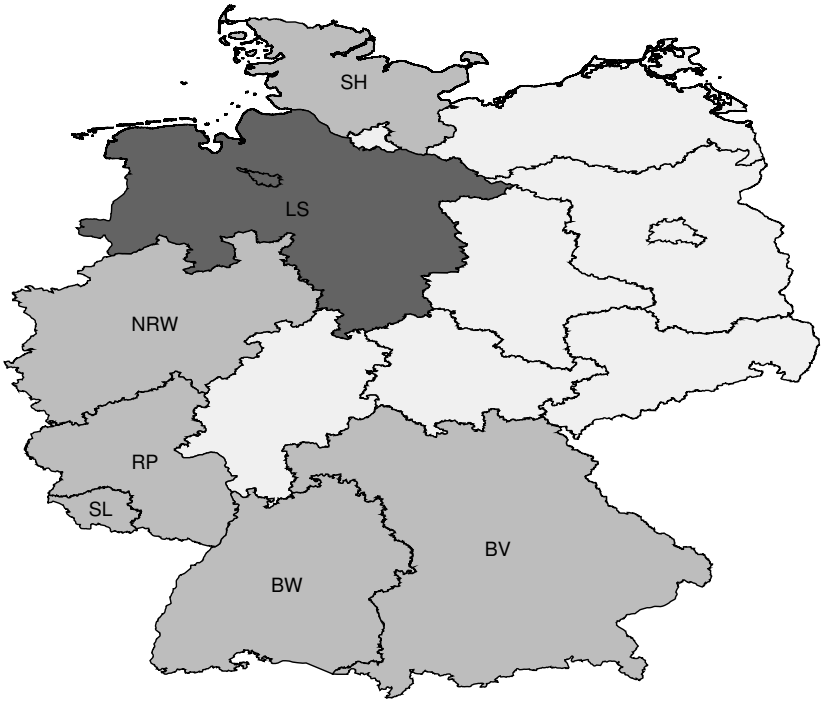

Figure 1: Treatment (dark gray) and control states (light gray). From North to South: SH (Schleswig-Holstein), LS (Lower Saxony), NRW (Northrhine-Westphalia), RP (Rhineland-Palatinate), SL (Saarland), BW (Baden-Württemberg), and BV (Bavaria).

speed at which changes occurred differed considerably (e.g. Ziegenspeck, 2000, p. 81). While initial trials with OS-type schools were evident in several states, ultimately, the OS as a school *independent of the three-tiered system* was fully introduced only in Lower Saxony and the small city state of Bremen.[5]

Our analysis below compares changes in educational outcomes in students that received schooling in Lower Saxony to changes in students that received schooling in other states of West Germany. We exclude the city states of Bremen, (West-)Berlin, and Hamburg as they differ in many important ways. We also exclude the state of Hesse which introduced an OS-style school, the *Förderstufe*, but also adhered to the old system. The new school type would thus merely be an alternative, not a replacement. Figure 1 depicts the locations of the states we focus on within Germany.

Out of the six remaining states that will serve as controls in our analysis below, two, Northrhine-Westphalia and Saarland, did not introduce OS schools of any stripe by 1975. In both states, however, the introduction was planned for the second half of the 1970s (Haenisch and Ziegenspeck, 1977, pp. 40ff). Ultimately, reform efforts would run out of steam or would

[5]Bremen established six years of primary education after the war. However, from 1957 onward, there was an option to switch to the academic track already upon completing fourth grade (Schuchart, 2006, p. 70).

be prevented through referenda (Rösner, 1981). There were experiments with OS-type schools in the two southern states of Bavaria and Baden-Württemberg by 1975. However, they would eventually also fail to abolish tracking after fourth grade.[6] Rhineland-Palatinate and Schleswig-Holstein re-labeled fifth and sixth grade (as did Northrine-Westphalia eventually), but tracking would continue to be conducted after fourth grade.

OS schools would be abolished again abruptly in Lower Saxony in 2004 in what was at the time a controversial political decision. While the OS was still judged favorably overall by a majority of parents, students, and teachers, a 2001 study found that these groups also took a critical view of the OS (Avenarius et al., 2001). Interestingly, the study found that the OS did not mitigate the effect of parental background on track choice and that teachers' recommendations were often influenced by factors other than academic potential (ibid.).

It is important to understand several aspects of the design of the OS in Lower Saxony over and above delayed tracking: first, the administration of OS schools was independent of other secondary schools. This was not the case, for instance, in states such as Schleswig-Holstein and Rhineland-Palatinate in which the OS schools were affiliated with secondary schools in the old three-tiered system.

Second, the reform entailed some degree of within-class tracking and ability grouping: teachers were expected to adjust tasks and demands within each class according to students' interests and abilities (Avenarius et al., 2001). Moreover, students were grouped into three different levels of academic achievement during English and math lectures, where the three levels were to roughly reflect the demand levels of secondary school tracks (Ziegenspeck, 2000, pp. 262ff.). One could thus argue that we are comparing tracking across schools to a system with a strong element of tracking within schools.

Third, OS schools employed teachers from all secondary school tracks on a part-time basis. As every class was to be taught by teachers of all secondary school tracks, students were exposed to teachers with profound knowledge of the everyday practice at the secondary schools in order to ensure a suitable tracking. Fourth, teachers at OS schools were asked to regularly document students' behavior and academic progress on an observation sheet according to fixed criteria. This was aimed at nudging teachers towards a more objective judgment of children's' academic potential (Hornich, 1976, p.110).

Finally, there was an emphasis on frequent consultations with parents in order to inform them about their child's future perspectives and to explain and to convince them of the ensuing recommendation. This was important since from 1979 onward, parents had the final say about whether or not to heed the schools' recommendation. There is anecdotal evidence that parents at the time had a tendency to overrule recommendations in favor of choices typical of their social class. There was an understanding among educators that working class parents had to be convinced that their child did not necessarily have to attend one of the two vocational tracks (Ziegenspeck, 2000, pp. 146ff.). To summarize, the reform intended to balance the advantages

[6] Bavaria would delay the decision between the two lower tracks by two years until the early 2000s but tracking into academic and vocational track would always be conducted after fourth grade. See Piopiunik (2013).

associated with creating homogeneous groups with an effort to dilute the influence of parental background.

3 Literature review

3.1 The effect of tracking on educational outcomes: mechanisms

There are two channels through which the timing of tracking potentially affects educational outcomes: first, peer effects are likely to play an important role.[7] Second, tracking may proceed on the basis of observables other than academic potential. We discuss both points in turn.

Peer effects While the early empirical literature on direct peer effects is beset with econometric problems (e.g. Manski, 1993; Sacerdote, 2001), recent studies suggest that students benefit from the presence of higher-achieving schoolmates (see, *inter alia,* Sacerdote, 2001; Ding and Lehrer, 2007; Lavy et al., 2011). If tracking effectively matches peers of similar quality, one would thus expect that it has a positive effect on the variance in achievement.

At the same time, it is often argued that tracking may allow for efficiency gains. Tracking will result in more homogeneous groups which, in turn, may allow teachers to tailor lessons more specifically to students' needs. Tracking may thus benefit *all* students. However, empirical studies on this channel are ambiguous (e.g. Epple and Romano, 2011). While Ding and Lehrer (2007) and Duflo et al. (2011) find a positive effect of increasing peer homogeneity on achievement, Lyle (2009), for instance, finds that a *higher* variance in peer SAT math scores benefits students.

Misallocation of students to tracks The second point relates to the precision with which students' are tracked. It is usually argued that academic potential, on which track choice at the relevant age should arguably be based, is difficult to observe initially yet that the signal becomes stronger over time (e.g. Brunello et al., 2007). Early tracking may thus be associated with a misallocation of students to tracks. In particular, it may be the case that non-cognitive skills such as attentiveness become more important for track choice at an early age when cognitive skills are still difficult to observe. Non-cognitive skills, in turn, have been shown to be related to parental background variables such as parents' educational attainment (Segal, 2008).

There is indeed evidence that early tracking fails in separating students effectively by academic potential, particularly in Germany. First, there is often considerable overlap in test scores between different school tracks (Baumert et al., 2003a). Lehmann and Peek (1997) find that in Hamburg, which tracks at the age of ten, students with uneducated parents have to score higher in standardized tests in order to receive an academic track recommendation with the same probability as students with educated parents.

Second, track choice is often found to be determined by variables that are arguably unrelated to academic potential. A large literature, for instance, documents relative age-effects in educa-

[7] Epple and Romano (2011) provide a comprehensive review of the literature on peer effects.

tion: Puhani and Weber (2007), Mühlenweg (2010), Jürges and Schneider (2011) and Dustmann et al. (2014) all find that students' exact birthday relative to an arbitrary enrollment cut-off date predicts track choice in Germany.[8] While such findings are indicative of inefficiencies associated with tracking, both Jürges and Schneider (2011) and Dustmann et al. (2014) find no evidence for a persistent effect of relative age on educational outcomes. Moreover, Jürges and Schneider (2011) find no evidence that the age at which states track affects the strength of the relative age effect based on variation across states.

There is also an increasing interest in gender differences in student achievement, the evolution of such differences over time, and their interaction with tracking. Bedard and Cho (2010) report that tracking is pro-female in Germany in that females are placed in classes with higher average ability. Both Lehmann and Peek (1997) and Jürges and Schneider (2011) report that boys are less likely to be recommended to the academic track in Germany *conditional on academic achievement*, suggesting that girls outperform boys in other relevant dimensions.[9] Interestingly, Jürges and Schneider find no evidence that the gender effect varies across states with different tracking procedures and while the relative age-effect seems to fade over time, the number of female students in the academic track is still greater in ninth grade. The authors conclude that delaying tracking by two years would not reduce gender bias in track attendance.

3.2 Related work

There is a large literature that investigates the effects of tracking but evidence is mixed (Betts, 2010, provides a review). Betts and Shkolnik (2000) review an early literature that relies on variation across US schools with regard to the decision to group students by achievement within schools (i.e. within-school tracking) (e.g. Hoffer, 1992; Argys et al., 1996). These studies suggest that tracking benefits high-achieving students and hurts low-achieving ones. However, Betts and Shkolnik argue that these studies overstate the effect of ability grouping on inequality because of omitted ability bias.[10] In their own empirical analysis, in which they compare similar students in tracking and non-tracking schools, they find almost no effect of formal ability grouping.

A different strand of the literature studies differences in the degree of tracking across countries. Based on a DD estimation strategy that replaces changes within countries with changes across grades, Hanushek and Woessmann (2006) find a positive effect of tracking on performance inequality and no effect on performance levels. Waldinger (2007), however, shows that while parental background is important in countries that track early, it is not more important after tracking takes place. His findings thus hint at the presence of an omitted variable that affects both the inter-generational transmission of education and the likelihood of a country adopting early tracking.

[8] Schneeweis and Zweimüller (2014) present evidence for Austria in which tracking also occurs at the age of ten.

[9] Jürges and Schneider relate this finding to "differences in verbal and non-cognitive skills at age 10."

[10] Another problem in this literature relates to difficulties in identifying schools that use ability grouping as this could be done both formally or informally.

Brunello and Checchi (2007) and Schütz et al. (2008) rely on cross-country variation in design features of education systems in order to estimate their effects on the importance of family background characteristics. The former authors focus on long-term outcomes such as earnings and employability and are thus closer to the present study. Both studies find that early tracking accentuates the importance of family background characteristics.

To date, there are only few studies that provide convincing evidence on the effects of reforms that affect tracking within countries. Meghir and Palme (2005) assess the effects of a major school reform in Sweden in the late 1940s on education and earnings in a set-up similar to our own. The reform entailed the abolition of ability tracking after grade six. While the authors find a small positive impact of the reform on wages, the effects they report for sub-groups stratified by ability and parental education are more pronounced: their estimates suggest that children of fathers with low levels of education experienced an increase in their annual earnings by about 3.4 percent while earnings of children of educated fathers decreased by 5.6 percent. The quantity of education accumulated appears to account only partly for these changes: the former group witnesses a significant increase in years of education, while the later is not affected.

Pekkarinen et al. (2009) exploit a major Finnish school reform which replaced a two-track school system with a single-track system in order to analyze the effect of tracking on the correlation of earnings between fathers and sons. The reform effectively postponed tracking from age eleven to age 16. The authors find that the inter-generational income elasticity decreases by 0.07 percentage points. Allowing the effect of the reform to vary across quintiles of the distribution of fathers' earnings, they find that the likely channel is a weakening of the statistical association between earnings within wealthy families. Analyzing the same data, Kerr et al. (2013) find that the reform increased cognitive skills among Finnish army recruits whose parents had less than a high school education.

As one would expect, country-wide reforms in Scandinavian countries usually involved other elements and it is often all but impossible to disentangle the effects of de-tracking from other components of reforms. Sweden's reform, for instance, entailed in addition to de-tracking an increase in compulsory years of schooling and cash transfers to compensate families for foregone earnings. Finland's reform entailed the abolition of a vast network of private schools which were placed under municipal authority. Lower Saxony's reform, in contrast, entailed 'only' adjustments to curricula in addition to delayed tracking. Our results below suggest that the pro-equality character of Scandinavian reforms is very similar to what we find for Germany.

4 Estimating the impact of the reform on educational outcomes

4.1 Data and descriptives

In this section, we describe some key variables of our analysis and explain coding decisions. The dataset we use is the German Socio-Economic Panel (GSOEP v29), a nationally representative survey that covers 77,927 individuals in 34,416 households in Germany. GSOEP data has been collected on an annual basis between 1984 and 2012 and includes a wide range of background variables. See Wagner et al. (2007) for a description of the dataset.

Key variables in our analysis are total years of education and two binary variables indicating whether an individual has attained *Abitur* and a tertiary degree. Means are reported in panel A of table 1. Years of education in this dataset refer to the number of years usually required to obtain certain degrees, not to the time spent in education, and includes all stages from primary to tertiary education.[11] Grade repetition will thus not be reflected in this measure of educational attainment.

Information on parents' educational attainment is available for almost all individuals in the data. There are two variables for each mother and father that relate to the school-leaving certificate and the type of tertiary or vocational training completed. We code a binary variable, $educ^p$, equal to unity if either the mother or father has (i) attained *Abitur*, has (ii) completed the upper vocational track of secondary (*Realschule*) and at the same time has completed vocational education (*Ausbildung*), or has (iii) completed training as a clerk, a public health worker, a civil servant, an engineer, or holds a degree from a tertiary education institution.[12] This results in about 35 percent of our observations being classified as having educated parents. Note that our definition allows for either the father or the mother (or both) to have attained this level of education. Hence, the presence of only one educated parent is assumed sufficient to generate the relevant externality at the household-level (Basu and Foster, 1998). Since this definition of having educated parents is somewhat arbitrary, we also test in section 5.2 whether our results below are robust to using an alternative proxy for parents' socio-economic status (based on parents' occupations at the age the respondent was 15).

Other variables considered in our analysis capture socio-demographic characteristics such as year of birth, gender and migrant status. We also include in our analysis below a complete set of indicator variables that indicate the size of the respondent's locality during childhood.[13] Means of these variables are reported in panel A of table 1.

The GSEOP data do not provide direct information about which type of school individuals

[11] The maximum this variable takes is 18 years, 13 years until the Abitur plus five years in order to obtain a diploma. Obtaining a BA degree or a degree from a university of applied sciences (*Fachhochschule*) usually takes three years. Depending on the type of job, vocational training adds 1.5 or two years to the total. Completing the lower and upper secondary track takes nine and ten years, respectively.

[12] This could be either a regular university (including foreign institutions) or a university of applied sciences.

[13] Categories are "no information available", "city", "large town", "small town", and "rural area."

Table 1: Means of key variables by state.

	LS	SH	NRW	RP	BW	BY
Panel A. Outcomes						
Years of education	12.62	12.55	12.89	12.33	12.69	12.49
Abitur	0.39	0.39	0.45	0.36	0.41	0.37
University degree	0.23	0.21	0.26	0.22	0.25	0.24
Panel B. Individual background characteristics						
Panel B1. Basic characteristics						
Age	45.20	44.81	45.44	46.45	44.59	45.39
Male	0.48	0.48	0.47	0.47	0.49	0.48
Educated parent	0.37	0.45	0.36	0.30	0.35	0.35
Migrant	0.07	0.06	0.11	0.10	0.18	0.11
Panel B2. Childhood place of residence was mostly...						
...city.	0.12	0.15	0.30	0.10	0.12	0.16
...large town.	0.16	0.21	0.23	0.15	0.19	0.12
...small town.	0.22	0.22	0.21	0.22	0.21	0.22
...rural area.	0.46	0.39	0.20	0.49	0.42	0.46
Panel B3. Sample information						
Last observed in 2012	0.56	0.49	0.56	0.60	0.53	0.57
Schooling state is...						
...state at age ten.	0.03	0.02	0.04	0.02	0.05	0.04
...state of last school visit.	0.71	0.73	0.68	0.68	0.67	0.70
...childhood state.	0.15	0.14	0.16	0.19	0.14	0.14
...first state in which observed.	0.10	0.11	0.12	0.10	0.13	0.12
Observations	1,593	488	3,484	981	1,977	2,430

SH: Schleswig-Holstein; LS: Lower Saxony; NRW: Northrhine-Westphalia; RP: Rhineland-Palatinate; SL: Saarland; BW: Baden-Württemberg; BV: Bavaria. Based on GSOEP v29-data.

attended at the age of ten. We therefore supplement the data with information from the statistical reports on schooling in Lower Saxony (Landesamt für Statistik Niedersachen). These reports tabulate the number of students by grade and birth year in different school types. Combining these data allows us to calculate the percentage of students of one cohort that attended a OS school in a given school year.[14] Enrollment at the time occurred at the age of six and school years start after the end of the summer in July or August. For instance, an individual born in the first half of 1960 was supposed to start schooling in school year 1966/1967 and somebody born in the second half of that year was supposed to start schooling in school year 1967/1968. We therefore associate individuals of a given cohort in Lower Saxony with the share of students in OS schools with one of two subsequent school years depending on whether the individual was born during the first or second half of the year. Figure 2 plots the variable OS against birth years for individuals that attended schools in Lower Saxony.

There is no direct information in the GSOEP dataset on where exactly an individual resided

[14] We ignore private schools which play a negligible role in terms of student in-take in Germany.

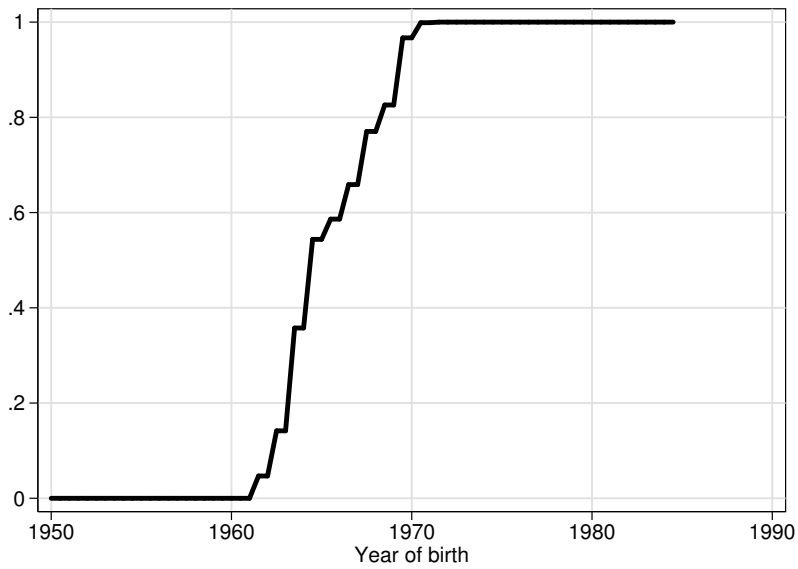

Figure 2: OS share against birth year in Lower Saxony.

at the age of ten. A second challenge is thus to infer the state in which individuals went to school at that age. Clearly, imputing the current state would confound treatment and control due to inner-German migration. There are several variables in the dataset that allow us to close in on the required information such as whether the current place of residence was also the childhood place of residence, the location of an individual in 1989, and, importantly, the state of the last school visit. We proceed as follows:

1. We impute the state of the last school visit whenever the exact state cannot be inferred. This information was only gathered in one year, so there are many individuals in the dataset for which it is not available.

2. If neither the exact state nor the state of the last school visit is available, we resort to information from a question about the childhood state of residence. All individuals were asked whether they still live in the same place in which they lived during their childhood. While 'childhood' does not seem particularly well defined, we impute the first state in which an individual was encountered whenever the answer is 'yes.'

3. If all of the above fails, we impute the current state.

4. Finally, we remove individuals from our sample that match one of the following criteria: first, all individuals were asked about the place in which they lived in 1989, shortly before Germany's re-unification. If somebody states that she lived in East Germany and was born before 1978, it is nearly impossible that she received her schooling in one of the Western

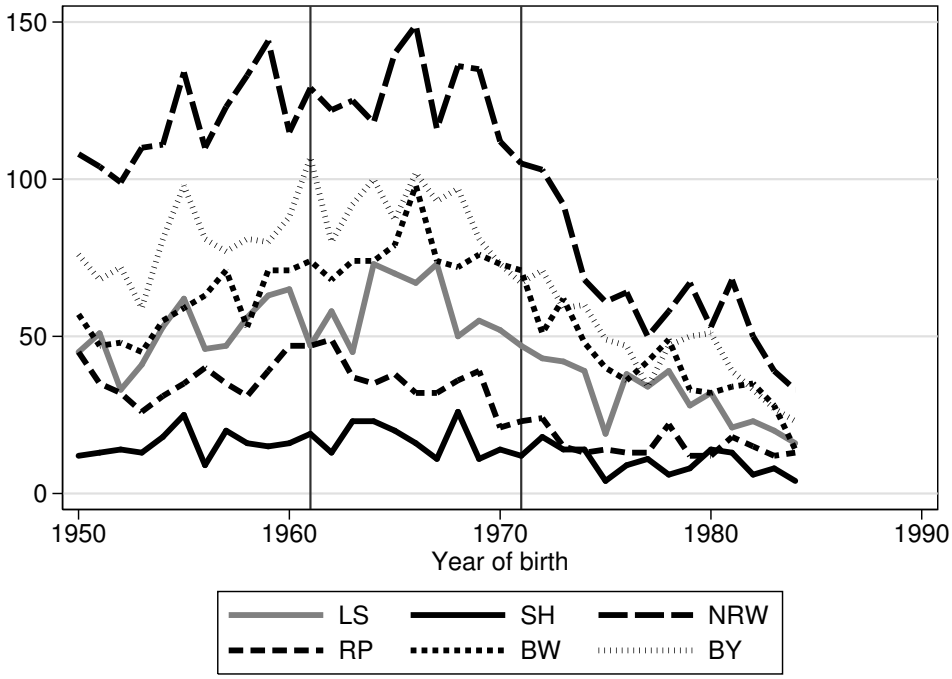

Figure 3: Number of individuals against year of birth. SH: Schleswig-Holstein; LS: Lower Saxony; NRW: Northrhine-Westphalia; RP: Rhineland-Palatinate; SL: Saarland; BW: Baden-Württemberg; BV: Bavaria. Based on GSOEP v29-data.

German states. We thus drop all these observations from our sample. Second, the dataset contains information about the year in which an individual migrated to Germany. Together with information on birth years, we can thus calculate the age at which an individual migrated. We drop all migrants that moved to Germany only after they turned eleven.

Panel B2 of table 1 reports the frequencies with which individuals were classified according to the above steps. It shows that for the majority of individuals, we make the assumption that the state of last school visit is also the state at the age of ten. Only in ten to 15 percent of individuals are classified based on the first state in which they are observed. While we believe that these steps are all reasonable, there will still be misclassification error. It is thus important to remember that confounding treatment and control would tend to bias coefficients towards zero. Our estimates should thus be interpreted as lower bounds of the true effect.

Since our interest is exclusively in variables that do not change over time, we discard all observations except the most recent for each individual. This is 2012 for more than half of the observations in our sample but sometimes an earlier year (see panel B3 of table 1).

As was discussed in section 2, we retain only observations on individuals that are likely to have received their schooling in one of seven states. Two states, Rhineland-Palatinate and Saarland,

in this dataset were until recently treated as one entity due to low case counts in the latter and associated privacy issues. We further retain only observations on individuals above the age of 28 as these are likely to have completed their education.[15] In consequence, all individuals in our sample were born before 1985. We also exclude individuals born before 1950. Figure 3 depicts the distribution of observations by birth year for each state.

4.2 Identification strategy

We estimate the causal effect of the reform on educational outcomes, splitting the sample successively by gender and parental education. Our analysis is retrospective: individuals are observed only after they have acquired their education. We employ a DD estimator, where the first difference is across cohorts and the second difference is across individuals in different states. The control group is composed of individuals that received schooling in states that did not (or not fully) introduce comprehensive schools in grade five and six. Our main specification is

$$educ_{isc} = \beta(OS_{sc} \times LS_s) + \mathbf{x}'_{isc}\gamma + \lambda_s + \tau_c + \epsilon_{isc}. \tag{4.1}$$

$educ_{isc}$ denotes the educational outcome of an individual i who received schooling at the relevant age in state s, and is a member of cohort c. As the dependent variable $educ_{isc}$ we consider either years of education (as defined above) or binary indicators of having attained *Abitur* or a degree from a university. In the latter case, the model is thus a linear probability model (LPM).

OS_c is the share of students in Lower Saxony of cohort c that have attended an OS school and LS_s is a binary indicator for having received schooling at the age of ten in that state. The parameter of interest is thus β which measures the effect of the reform on educational outcomes. $\mathbf{x}_{isc}$ is a matrix of time-invariant demographic and socio-economic variables. We include indicators for the respondent's gender, migrant status, and size of the locality during childhood. λ_s and and τ_c denote complete sets of state- and cohort-fixed effects, respectively. When the sample is pooled across males and females, λ_s, τ_c, and all variables in $\mathbf{x}_{isc}$ are also interacted with the respondent's gender. Finally, ϵ_{isc} is the usual white noise-error term.

We also estimate an alternative model which allows for a trend-break, a linear change in the treatment effect over time:

$$educ_{isc} = \beta_1(OS_{sc} \times LS_s) + \beta_2(OS_{sc} \times LS_s \times c.\,birth\,year_c) + \mathbf{x}'_{isc}\gamma + \lambda_s + \tau_c + \epsilon_{isc}. \tag{4.2}$$

This model includes in addition to all right hand-side variables in (4.1) an interaction between the treatment variable, $(OS_{sc} \times LS_s)$, and a continuous birth year variable centered on 1972, the year of birth of the first cohort in which (nearly) every student in Lower Saxony attended an OS

[15] Schooling usually starts at the age of six and completing secondary education with the *Abitur*, the highest school-leaving certificate in Germany, requires 13 years of schooling over the time period that we study. Mandatory military service never exceeded two years. If we add five years required to obtain a university degree, the age at which one would complete university is 26.

school. The coefficient on this variable, β_2, can be interpreted as the annual change in Lower Saxony in the outcome variable in each year after the introduction of the OS due to the reform.

In section 5.2, we address possible violations of the common trend-assumption and plausible self-selection into OS schools in partially treated cohorts. One general concern one may have in estimating equations (4.1) and (4.2) is sample selection: if families decided to move from or to Lower Saxony in response to the reform and if this decision is correlated with parental education, our estimate of β would be biased. There are, however, no reports of such responses that we know of. Note also that this would have required parents to move to another state which, in most cases, would imply leaving one's job. It is implausible that a reform that only affected two grades would have induced this behavior.

If there would have been an important migratory response we would observe that the reform was associated with a change in the probability of having educated parents. To check that this is not the case, we regress the binary indicator of having educated parents on $(OS_{sc} \times LS_s)$ as well as a complete set of state- and cohort-fixed effects (results not reported). We also include all controls that we include when we estimate (4.1) and run this regression for the pooled sample and males and females separately. We find no evidence that the reform had an effect on the probability of having educated parents: all coefficients are close to zero and statistically insignificant. It seems that strategic migration does not affect our results.

Before we turn to regression results, figure 4 offers a preview of our main results. We plot the average number of years of education for males with educated (dashed lines) and uneducated parents (solid) in Lower Saxony (gray) and in all other states (black) against their year of birth. Three features of these series are of interest: first, we see that there is no strong trend in years of education in these series. While there is a minor increase in this variable over time, from 12.8 years for individuals born in the 1950s to 13.2 years for those born in the 1980s, this increase is mostly due to secular drift, that is, parents becoming more educated. Second, we see that the gap in educational attainment in Lower Saxony is large initially yet narrows considerably during the early 1970s. The un-adjusted gap is 2.8 years in Lower Saxony before the introduction of the OS and shrinks to 2.1 years after the end of the roll-out period. This does not seem to be the case in other states, where the un-adjusted gap increases from 2.2 years to about 2.3. Finally, it may seem that the initial decrease in the gap that coincides with the end of the implementation period dissipates over time. It should be noted, however, that the number of observations available from all states falls over the course of the 1970s—see figure 3. The point estimates depicted towards the end of the series therefore have larger confidence bands.

Table 2: Impact of the reform on years of education: OLS estimates.

	All			Females			Males		
Parents' education	All	High	Low	All	High	Low	All	High	Low
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A1. Baseline results									
$OS_{sc} \times LS_s$	0.05	−0.37	0.25	−0.06	−0.07	−0.05	0.16	−0.74	0.57**
	(0.17)	(0.31)	(0.19)	(0.23)	(0.41)	(0.25)	(0.26)	(0.46)	(0.29)
Panel A2. Trend break model									
$OS_{sc} \times LS_s$	−0.02	−0.57*	0.24	−0.13	−0.19	−0.05	0.11	−1.01**	0.54*
	(0.18)	(0.33)	(0.20)	(0.25)	(0.45)	(0.26)	(0.27)	(0.49)	(0.29)
$OS_{sc} \times LS_s \times c.birth\,year$	0.03	0.05	0.01	0.03	0.03	0.00	0.02	0.08	0.02
	(0.02)	(0.03)	(0.03)	(0.03)	(0.05)	(0.04)	(0.03)	(0.05)	(0.04)
Observations	10,953	3,920	7,033	5,698	2,045	3,653	5,255	1,875	3,380
Panel B1. Linear, state-specific year-of-birth-trends									
$OS_{sc} \times LS_s$	0.06	−0.33	0.26	−0.06	−0.01	−0.07	0.19	−0.71	0.62**
	(0.17)	(0.30)	(0.19)	(0.23)	(0.41)	(0.25)	(0.26)	(0.46)	(0.29)
Panel B2. Excluding partially treated cohorts									
$OS_{sc} \times LS_s$	0.07	−0.53	0.35	0.01	−0.15	0.04	0.14	−0.97*	0.68**
	(0.19)	(0.34)	(0.22)	(0.26)	(0.46)	(0.28)	(0.29)	(0.51)	(0.34)
Panel B3. Sample split by prestige of parents' occupation									
$OS_{sc} \times LS_s$	0.05	0.00	0.15	−0.06	0.46	−0.18	0.16	−0.57	0.50*
	(0.17)	(0.34)	(0.18)	(0.23)	(0.45)	(0.23)	(0.26)	(0.51)	(0.28)

Standard errors reported in parentheses. *, **, and *** denote significance at the ten-, five-, and one-percent level, respectively. All regressions include state-of-school-visit- and year-of-birth-fixed effects. Further controls include a full set of dummies capturing the size of respondents' childhood place of residence and a dummy for migrant status. When we use the pooled sample, we also include a dummy variable for the respondent's sex and interact it with all other controls (including fixed effects). Based on GSOEP v29-data.

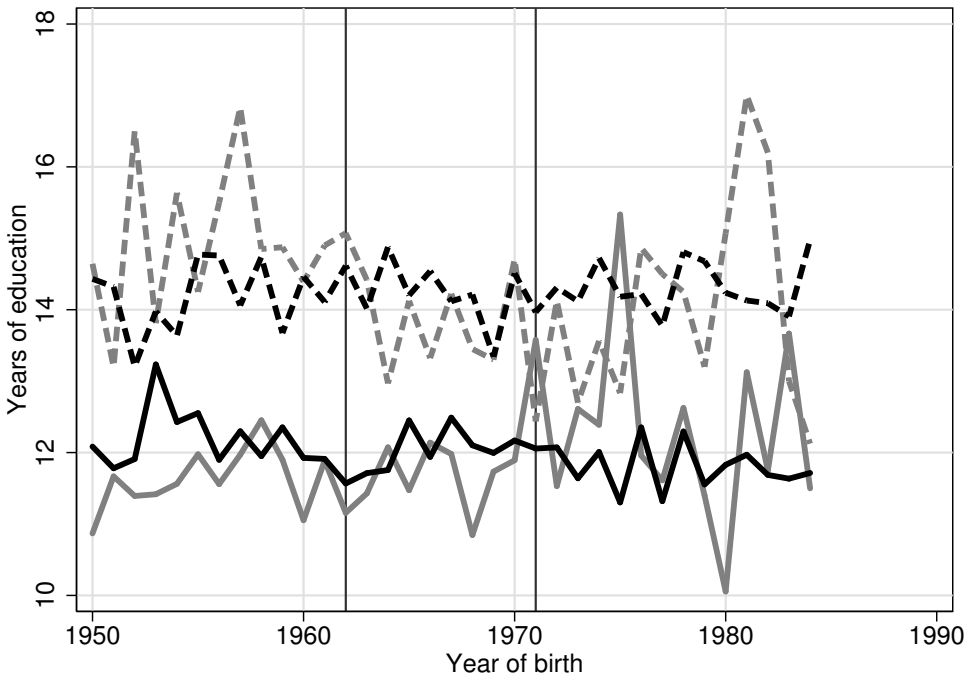

Figure 4: Average years of education of males against birth year. The gray lines depict observation for Lower Saxony, black lines observations for all other states; dashed lines depict observations for individuals with educated parents, and solid lines observations for individuals with uneducated parents. Based on GSOEP v29-data.

5 Results

5.1 Main results

Results from estimating (4.1) using years of education as the outcome variable are reported in panel A1 of table 2. Following recommendations by Angrist and Pischke (2009, p. 307),[16] we report conventional standard errors throughout as these turned out to be the most conservative option.[17] Column (1) reports results from using the entire sample. The coefficient suggests the reform had no effect on years of education on average. In columns (2) and (3) we report results for using the subsamples of individuals with educated and uneducated parents, respectively. While the estimates suggest that the reform had a negative effect on years of education for individuals with educated parents and a positive effect for individuals with uneducated parents, we cannot

[16] Angrist and Pischke (2009) show that robust standard errors may be smaller than conventional standard errors due to small sample bias and/or higher sampling variance.

[17] We also experimented with robust 'sandwich'-type standard errors as well as standard errors clustered at the level of cohort-states and states. We found that conventional standard errors were the most conservative and not very different from robust standard errors and standard errors clustered at the level of cohort-states. Clustering at the state-level, however, resulted in much smaller errors.

reject that the reform had no effect for either group.

In the remaining columns of panel A1, we report results from estimating equation (4.1) separately for females and males (columns (4)–(6) and (7)–(9), respectively). We find no evidence for an effect of the reform on females: the coefficients are small and insignificant. The pattern for males, however, is interesting: while the average effect for males is about one-sixth of a year yet insignificant, we find a negative effect of about three-fourths of a year for males with educated parents. The coefficient is significant at the ten-percent level. For individuals with uneducated parents, on the other hand, we find a negative effect of slightly more than one-half of a year of education (column (9)). The coefficient is statistically significant at the five-percent level. We conclude that the effect of the reform on years of schooling depends on parents' educational attainment for males.

Panel A2 of table 2 reports results from estimating equation (4.2). We report both coefficients of interest. We find that estimates of β_2, the coefficient on the three-way interaction term, are always greater than zero. This would suggest a modest acceleration in progress towards higher levels of educational attainment in the years following the roll-out of OS schools in Lower Saxony across subgroups. We find the largest effect, an increase by 0.08 years of education per year, for males with educated parents. This would imply that the negative effect of the reform for this group dissipated after 12.5 years. However, all estimates are insignificant at conventional levels.

In panel A1 of table 3, we report results from estimating (4.1) and (4.2) using the binary indicator of having attained *Abitur* as one's school leaving-certificate as the outcome variable. As mentioned above, the model is now an LPM. Coefficient estimates and standard errors are all multiplied by one hundred in this case and should thus be interpreted as percentage point changes.

Results from estimating (4.1) are reported in panel A1 and suggest that we cannot reject that the reform had no impact on the probability of attaining *Abitur* for any subgroup. Some of the coefficient estimates are substantial in economic terms, however, and the pattern is similar to the one observed for years of education. In particular, we find economically meaningful coefficients for males once we consider individuals with educated and uneducated parents separately (columns (8) and (9), respectively). The pattern of effects changes when we include the three-way interaction term (panel A2): we find a negative effect of the reform on males with educated parents of -15.5 percentage points that is statistically significant at the five-percent level. While this seems very large by all standards, the estimate on the three-way interaction term suggests that the effect vanishes at a speed of 1.7 percentage points per year.

Finally, table 4 presents estimates of the reform's effect on the probability of obtaining a university degree. Results are well in line with our findings concerning years of education: we find no effect on average (column (1)) and no statistically significant effects on individuals when we split the sample by parental education (columns (2) and (3), respectively). Both estimates are economically significant, however, with a pattern mirroring the one for years of education. Comparing estimates for females and males, it becomes clear that this is again

Table 3: Impact of the reform on probability of attaining Abitur: OLS estimates.

	All			Females			Males		
Parents' education	All	High	Low	All	High	Low	All	High	Low
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A1. Baseline results									
$OS_{sc} \times LS_s$	−0.9	−4.9	0.6	−2.9	−0.6	−4.1	1.3	−10.2	5.7
	(3.0)	(5.0)	(3.6)	(4.0)	(6.9)	(4.7)	(4.5)	(7.3)	(5.5)
Panel A2. Trend break model									
$OS_{sc} \times LS_s$	−2.3	−8.9	0.5	−4.4	−3.0	−4.3	0.1	−15.5**	5.5
	(3.2)	(5.4)	(3.7)	(4.3)	(7.6)	(4.8)	(4.7)	(7.8)	(5.6)
$OS_{sc} \times LS_s \times c.\,birth\,year$	0.5	1.1*	0.1	0.5	0.6	0.1	0.6	1.7*	0.1
	(0.4)	(0.6)	(0.5)	(0.5)	(0.8)	(0.6)	(0.6)	(0.9)	(0.8)
Observations	10,953	3,920	7,033	5,698	2,045	3,653	5,255	1,875	3,380
Panel B1. Linear, state-specific year-of-birth-trends									
$OS_{sc} \times LS_s$	−0.7	−5.2	0.9	−2.8	−1.1	−4.1	1.6	−10.0	6.3
	(3.0)	(5.0)	(3.6)	(4.0)	(6.9)	(4.7)	(4.5)	(7.2)	(5.5)
Panel B2. Excluding partially treated cohorts									
$OS_{sc} \times LS_s$	−0.8	−7.6	0.8	−2.6	−2.5	−4.0	1.1	−13.5*	6.0
	(3.4)	(5.6)	(4.0)	(4.5)	(7.8)	(5.2)	(5.0)	(8.1)	(6.0)
Panel B3. Sample split by prestige of parents' occupation									
$OS_{sc} \times LS_s$	−0.9	0.7	−0.5	−2.9	11.9	−8.1*	1.3	−13.3*	7.7
	(3.0)	(5.5)	(3.4)	(4.0)	(7.5)	(4.4)	(4.5)	(7.9)	(5.2)

Standard errors reported in parentheses. *, **, and *** denote significance at the ten-, five-, and one-percent level, respectively. All coefficient estimates and standard errors are multiplied by one hundred. All regressions include state-of-school-visit- and year-of-birth-fixed effects. Further controls include a full set of dummies capturing the size of respondents' childhood place of residence and a dummy for migrant status. When we use the pooled sample, we also include a dummy variable for the respondent's sex and interact it with all other controls (including fixed effects). Based on GSOEP v29-data.

Table 4: Impact of the reform on the probability of obtaining a tertiary degree: OLS estimates.

	All			Females			Males		
Parents' education	All	High	Low	All	High	Low	All	High	Low
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A1. Baseline results									
$OS_{sc} \times LS_s$	−0.0	−8.2	4.1	−3.4	−6.0	−1.7	3.8	−10.9	10.2**
	(2.6)	(5.1)	(2.8)	(3.4)	(6.7)	(3.5)	(4.0)	(7.6)	(4.4)
Panel A2. Trend break model									
$OS_{sc} \times LS_s$	−0.9	−10.4*	3.6	−4.3	−7.0	−2.1	2.9	−14.1*	9.7**
	(2.8)	(5.5)	(2.9)	(3.6)	(7.4)	(3.6)	(4.2)	(8.2)	(4.5)
$OS_{sc} \times LS_s \times c.\,birth\,year$	0.3	0.6	0.3	0.3	0.3	0.2	0.4	1.0	0.3
	(0.3)	(0.6)	(0.4)	(0.4)	(0.7)	(0.5)	(0.5)	(0.9)	(0.6)
Observations	10,953	3,920	7,033	5,698	2,045	3,653	5,255	1,875	3,380
Panel B1. Linear, state-specific year-of-birth-trends									
$OS_{sc} \times LS_s$	0.1	−8.0	4.2	−3.5	−5.4	−2.1	4.1	−11.1	11.0**
	(2.6)	(5.0)	(2.8)	(3.4)	(6.7)	(3.5)	(4.0)	(7.6)	(4.4)
Panel B2. Excluding partially treated cohorts									
$OS_{sc} \times LS_s$	−1.5	−14.3**	4.5	−5.1	−10.9	−2.3	2.6	−18.2**	12.0**
	(3.0)	(5.7)	(3.1)	(3.8)	(7.6)	(3.7)	(4.6)	(8.4)	(5.0)
Panel B3. Sample split by prestige of parents' occupation									
$OS_{sc} \times LS_s$	−0.0	0.6	1.0	−3.4	6.7	−5.6*	3.8	−7.0	8.2*
	(2.6)	(5.6)	(2.7)	(3.4)	(7.6)	(3.3)	(4.0)	(8.4)	(4.3)

Standard errors reported in parentheses. *, **, and *** denote significance at the ten-, five-, and one-percent level, respectively. All coefficient estimates and standard errors are multiplied by one hundred. All regressions include state-of-school-visit- and year-of-birth-fixed effects. Further controls include a full set of dummies capturing the size of respondents' childhood place of residence and a dummy for migrant status. When we use the pooled sample, we also include a dummy variable for the respondent's sex and interact it with all other controls (including fixed effects). Based on GSOEP v29-data.

driven by males. The reform is associated with a decline in the probability of obtaining a university degree by 10.9 percentage points for males with educated parents and an increase in the probability by 10.2 percentage points for males with uneducated parents. Only the latter effect is statistically significant at the five-percent level, however. For females, the coefficient estimates are negative and insignificant throughout but greater in absolute terms for those with educated parents (columns(4)–(6)).

5.2 Robustness checks

State-specific trends The identification assumption in estimating (4.1) is that, conditional on observables, changes in outcomes for individuals that went to school in Lower Saxony would have been the same on average as changes for individuals in other states. We observe several pre-reform cohorts in our dataset that we can use to estimate an alternative specification with state-specific linear time trends.

$$educ_{isc} = \beta_1(OS_{sc} \times LS_s) + \beta_2 OS_s + \mathbf{x}'_{isc}\gamma + \lambda_s + \eta^s c.\, birth\, year_c + \epsilon_{isc}, \tag{5.1}$$

where η^s is an estimate of the average annual change in years of education in state s that multiplies the birth year variable. We thus relax the common trend-assumption to some extent but impose linear trends in each state. Note that we also include in (5.1) the main effect of OS_s in order to capture variation that affects all states to the same extent at the time of the introduction of the OS in Lower Saxony.

Results are reported in panels B1 of tables 2, 3, and 4. Reassuringly, we find that both our estimates and their standard errors remain virtually unchanged. This suggests that our initial identification assumption of common trends across states holds.

Self-selection in partially treated cohorts An additional concern may be self-selection into OS schools. The introduction of OS schools in Lower Saxony took almost one decade and during this time period, sorting may have played an important rule. Given our findings, it seems plausible that educated parents, if given the opportunity, would have decided to send their children to schools in the old, three-tiered system. In any case, the co-existence of two tracking regimes potentially produces very different outcomes.[18] We therefore also estimate (4.1) including only cohorts that were either fully treated or fully untreated, i.e. cohorts born before 1962 or after 1971.

Results are reported in table in panels B2 of tables 2, 3, and 4. We find that omitting partially treated cohorts accentuates our previous findings. In particular, we find that the negative effect on years of education for males with educated parents (table 2, panel B2, column (8)) is now almost one year and significant at the ten-percent level. The positive effect for males with

[18] Recall that this was the reason for us to exclude Hesse, a state in which the two systems exists next to each other.

uneducated parents is now two-thirds of a year, about one-tenth of a year higher than when we use the entire sample. We also see particularly large heterogeneous effects on the probability of obtaining a tertiary degree for males (table 4, panel B2, columns (8) and (9)).

Alternative definition of parents' socio-economic status So far, we have relied on a single binary variable as a proxy parental education. Since the definition of such a variable will always be arbitrary to some extent, we consider an alternative measure of parental status, the prestige associated with parents' positions when the respondent was 15 years old. We define individuals as having parents which held a prestigious job when either the mother or the father were middle- or high-ranking public servants, employed as qualified laborers, or if they were self-employed academics. Applying this definition results in a total number of 3,190 individuals that are classified as having parents with prestigious jobs as opposed to 3.920 that have educated parents. As one would expect, there is limited overlap between these two variables: 8.5 percent of the sample are classified as having parents which have had a prestigious job yet are not educated while 15.2 percent are classified as having parents with non-prestigious jobs yet high levels of education. Neither the totals nor the percentages vary by gender.

Results are reported in panels B3 of tables 2, 3, and 4, respectively. Interestingly, we find that there is a positive effect on females with parents that have had prestigious jobs and a negative effect for females with parents that have had unprestigious jobs. However, both coefficients are not significantly different from zero. The pattern we observe for males is similar qualitatively to results obtained when splitting the sample by parental education. Estimates are, however, smaller in absolute terms (columns (8) and (9)) and only the estimate for males with parents that have had unprestigious jobs is significantly different from zero at the ten-percent level.

Results for the probability of obtaining Abitur, reported in panel B3 of table 3, show a similar pattern. For the former, for instance, we also find negative effects on females with parents that have had unprestigious jobs and males with parents that have had prestigious jobs that are statistically different from zero at the ten-percent level. The positive effect on the probability of obtaining a university degree that we document for disadvantaged males holds up (table 4, panel B3, column (9)).

6 Discussion

The previous sections show that the reform had no effect on educational outcomes on average. The effects on females that we find are often small and insignificantly different from zero. For males, on the other hand, we find that the effect varied by parental education with negative effects on individuals with uneducated parents and positive effects on individuals with educated parents. Overall then, it seems that for male students the reform achieved just what it was intended to achieve: it attenuated the importance of parental background for students' educational prospects.

The effect on total years in education required to attain a certain degree is comparatively

large given that the reform delayed tracking by only two years: the effect we estimate on males with uneducated parents was an additional 0.56–0.67 years of education on average while the the effect on males with educated parents was a negative 0.76–0.99. The returns of education for males in Germany are usually estimated to be around seven percent and often higher in tertiary education (Lauer and Steiner, 2000; Ammermüller and Weber, 2005). The implied changes in hourly wages would thus be neither very large nor negligible. They would also be well in line with findings reported by Meghir and Palme (2005).

As one would expect in the presence of selection effects for partially treated cohorts, our estimates show a tendency to increase in absolute terms when we exclude cohorts that may have been in a position to choose between school systems. This suggests that a partial reform that allows parental discretion in choosing among systems may have very different effects from a complete reform.

Both peer effects and uncertainty about academic potential are to an extent consistent with the above results. However, we argue that the latter explanation is more plausible given our results. Start with peer effects: if parental education captures own ability (i.e. because parental education depends on parental ability which, in turn, maybe partly hereditary), we would expect to find a negative effect of de-tracking on individuals with educated parents and vice versa. Unfortunately, there seems to be no possibility to control directly for own ability given our data.

However, there is no obvious way of reconciling the standard argument concerning tracking and peer effects with the gendered pattern of effects we find. A more plausible explanation of our findings is that tracking age interacts with gender differences in psycho-social and cognitive development. If behaviors and abilities that are mostly acquired during early childhood and depend on parental inputs are mistaken for academic potential and if educators are less likely to make this mistake at an older age, we would expect to find that later tracking is associated with a less prominent role of parental education. Differences in the levels and trajectories of cognitive development between boys and girls are well-established in the literature on cognitive development. Matthews et al. (2009), for instance, find evidence for gender differences in self-regulation, the ability to control behavior, cognitions, and emotions, for kindergarten children with females outperforming boys in self-control but not in academic achievement outcomes. In addition, Lenroot et al. (2007) find important differences in the trajectories of brain development between boys and girls with boys often trailing behind by several years.

A further point in support of this line of argumentation is that within-class tracking in major subjects was an important element of OS schools (as was discussed in section 2). Hence, while all students had the opportunity to interact during less salient classes, all students were likely in a position to benefit from greater peer homogeneity.

7 Conclusions

Recent research suggests that design features of education systems are an important determinant of the strength of the relationship between parental and own education. Systems that aim to track students early have often been found to affect positively the importance of parental background. On the other hand, tracking is often credited with increasing the efficiency of teaching through an increase in class homogeneity.

In this study we investigate whether and how the introduction of delayed tracking in Lower Saxony, one of Germany's federal states, in the 1970s affected educational outcomes of individuals differentiated by parental education. Based on a differences-in-differences estimation strategy, we find a positive effect of the reform on males with uneducated parents and a negative effect for males with educated parents. We cannot reject the hypothesis that the reform had no effect on females. Our findings are robust to the inclusion of state-specific linear trends for cohorts, the exclusion of partially treated cohorts, and alternating the definition of parental background. Our results thus suggest that the reform partly succeeded in increasing equality of opportunity in education.

Although we cannot rule out that the effects of the reform on educational outcomes operate through externalities through ability spill-overs and an increase in peer heterogeneity, we argue that the gender differences we observe are more likely to be explained by changes in the observability of academic potential over time for males *vis-à-vis* females. However, more research on the precise channels is clearly warranted.

References

Ammermüller, A. and Weber, A. M. (2005). Educational Attainment and Returns to Education in Germany—An Analysis by Subject of Degree, Gender and Region. *Centre for European Economic Research (ZEW) Discussion Paper 05-17*.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Princeton, New Jersey.

Argys, L., Rees, D., and Brewer, D. (1996). Detracking America's Schools: Equity at Zero Cost? *Journal of Policy Analysis and Management*, 15(4):623–645.

Avenarius, H., Döbert, H., Knauss, G., Weishaupt, H., and Weiß, M. (2001). Stand und Perspektiven der Orientierungsstufe in Niedersachsen. Technical report, Deutsches Institut für Internationale Pädagogische Forschung.

Basu, K. and Foster, J. E. (1998). On Effective Literacy. *Economic Journal*, 108(451):1733–1749.

Baumert, J. and Schümer, G. (2001). Schulformen als Selektionsbedingte Lernmilieus. In Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.-J., and Weiß, M., editors, *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im Internationalen Vergleich*, pages 454–467. Leske + Budrich, Opladen, Germany.

Baumert, J., Trautwein, U., and Artelt, C. (2003a). Schulumwelten—Institutionelle Bedingungen des Lehrens und Lernens. In Baumert, J., Artelt, C., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Tillmann, K.-J., and Weiß, M., editors, *PISA 2000: Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland*, pages 261–331. Leske + Budrich, Opladen, Germany.

Baumert, J., Watermann, R., and Schümer, G. (2003b). Disparitäten der Bildungsbeteiligung und des Kompetenberwerbs: Ein Institutionelles und Individuelles Modetationsmodell. *Zeitschrift für Erziehungswissenschaften*, 6(1):46–72.

Bedard, K. and Cho, I. (2010). Early Gender Test Score Gaps across OECD Countries. *Economics of Education Review*, 29:348–363.

Betts, J. R. (2010). The Economics of Tracking in Education. In Hanushek, E. A., Machin, S., and Woessmann, L., editors, *Handbook of the Economics of Education*, volume 3, pages 341–381. North Holland, Amsterdam.

Betts, J. R. and Shkolnik, J. R. (2000). Key Difficulties in in Identifying the Effects of Ability Grouping on Student Achievement. *Economics of Education Review*, 19(1):21–26.

Brunello, G. and Checchi, D. (2007). Does School Tracking Affect Equality of Opportunity? New International Evidence. *Economic Policy*, 22(52):781–861.

Brunello, G., Giannini, M., and Ariga, K. (2007). The Optimal Timing of School Tracking: A General Model with Calibration for Germany. In Woessmann, L. and Peterson, P., editors, *Schools and the Equal Opportunity Problem*, pages 129–156. The MIT Press, Cambridge, Mass.

Deutscher Ausschuss (1959). Empfehlungen zur Umgestaltung und Vereinheitlichung des Allgemeinbildenden Öffentlichen Schulwesens. In Bohnekamp, H., Dirks, W., and Knab, D., editors, *Empfehlungen und Gutachten des Deutschen Ausschusses für das Erziehungs- und Bildungswesen 1953–1965*, pages 59–115. Stuttgart, Germany.

Deutscher Bildungsrat (1970). Empfehlungen der Bildungskommission: Strukturplan für das Bildungswesen. Technical report, Deutscher Bildungsrat, Stuttgart.

Ding, W. and Lehrer, S. (2007). Do Peers Affect Student Achievement in China's Secondary Schools. *Review of Economics and Statistics*, 89(2):300–312.

Duflo, E., Dupas, P., and Kremer, M. (2011). Peer Effects, Teacher Incentives, and the Impact of Tracking. *American Economic Review*, 101:1739–1774.

Dustmann, C. (2004). Parental Background, Secondary School Track Choice, and Wages. *Oxford Economic Papers*, 56:209–230.

Dustmann, C., Puhani, P. A., and Schönberg, U. (2014). The Long-Term Effects of Early Track Choice. *IZA Discussion Paper No. 7897*.

Epple, D. and Romano, R. (2011). Peer Effects in Education: A Survey of the Theory and Evidence. In Benhabib, J., Jackson, M. O., and Bisin, A., editors, *Handbook of Social Economics*, volume 1B, pages 1053–1163. North Holland, Amsterdam.

Haenisch, H. and Ziegenspeck, J. W. (1977). *Die Orientierungsstufe*. Beltz Verlag, Weinheim, Germany, and Basel, Switzerland.

Hall, C. (2012). The Effects of Reducing Tracking in Upper Secondary School: Evidence from a Large-Scale Pilo Scheme. *Journal of Human Resources*, 47:237–269.

Hanushek, E. A. and Woessmann, L. (2006). Does Educational Tracking Affect Performance and Inequality? Differences-in-Differences Evidence Across Countries. *The Economic Journal*, 116(510):C63–C76.

Hoffer, T. (1992). Middle School Ability Grouping and Student Achievement in Science and Mathematics. *Educational Evaluation and Policy Analysis*, 14(3):205–227.

Hornich, K. (1976). *Praxis der Orientierungsstufe*. Verlag Herder.

Jürges, H. and Schneider, K. (2011). Why Young Boys Stumble: Early Tracking, Age and Gender Bias in the German School System. *German Economic Review*, 12(4):371–394.

Kerr, S. P., Pekkarinen, T., and Uusitalo, R. (2013). School Tracking and Development of Cognitive Skills. *Journal of Labor Economics*, 31(3):577–602.

Landesamt für Statistik Niedersachen (2014). Amtliche Schulstatistik. Various years.

Lauer, C. and Steiner, V. (2000). Returns to Education in West Germany - An Empirical Assessment. *ZWE Discussion Papers No. 00-04*.

Lavy, V., Paserman, D., and Schlosser, A. (2011). Inside the Black Box of Peer Effects: Evidence from Variation in the Proportion of Low Achievers in the Classroom. *The Economic Journal*, 122(559):208–237.

Lehmann, R. and Peek, R. (1997). Aspekte der Lernausgangslage von Schülerinnen und Schülern der fünften Jahrgangsstufe an Hamburger Schulen. Bericht über die Untersuchung im September 1996.

Lenroot, R. K., Gogrtay, N., Greenstein, D. K., Wells, E. M., Wallace, G. L., Clasen, L. S., Blumenthal, J. D., Lerch, J., Zijdenbos, A. P., Evans, A. C., Thompson, P. M., and Gieed, J. N. (2007). Sexual Dimorphism of Brain Developmental Trajectories During Childhood and Adolescence. *NeuroImage*, 36:1065–1073.

Lyle, D. S. (2009). The Effects of Peer Group Heterogeneity on the Production of Human Capital at West Point. *American Economic Journal: Applied Economics*, 1(4):69–84.

Malamud, O. and Pop-Eleches, C. (2011). School Tracking and Access to Higher Education Among Disadvantaged Groups. *Journal of Public Economics*, 95:1538–1549.

Manski, C. F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies*, 60:531–542.

Matthews, J., Ponitz, C. C., and Morrison, F. J. (2009). Early Gender Differences in Self-Regulation and Academic Achievement. *Journal of Educational Psychology*, 101(3):689–704.

Meghir, C. and Palme, M. (2005). Educational Reform, Ability, and Family Background. *American Economic Review*, 95(1):414–424.

Mühlenweg, A. M. (2008). Educational Effects of Alternative Secondary School Tracking Regimes in Germany. *Schmollers Jahrbuch: Journal of Applied Social Science Studies / Zeitschrift für Wirtschafts- und Sozialwissenschaften*, 128(3):351–379.

Mühlenweg, P. A. P. A. M. (2010). Persistence of the School Entry Age Effect in a System of Flexible Tracking. *Journal of Human Resources*, 45:407–435.

OECD (2003). *Learning for Tomorrow's World: First Results from Pisa 2003*. OECD (Organization for Economic Cooperation and Development), 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.

Piopiunik, M. (2013). The Effects of Early Tracking on Student Performance: Evidence from a School Reform in Bavaria. *Ifo Working Paper No. 153*.

Pischke, J.-S. and Manning, A. (2006). Comprehensive Versus Selective Schooling in England and Wales: What Do We Know? *NBER Working Paper No. 12176*.

Puhani, P. A. and Weber, A. M. (2007). Does the Early Bird Catch the Worm? Instrumental Variables Estimates of the Educational Effects of Age at School Entry in Germany. *Empirical Economics*, 32:359–386.

Rösner, E. (1981). *Schulpolitik durch Volksbegehren. Analyse eines gescheiterten Reformversuchs*. Beltz, Weinheim, Germany.

Sacerdote, B. (2001). Peer Effects with Random Assignment: Results for Dartmouth Roommates. *Quarterly Journal of Economics*, 116(2):681–704.

Schneeweis, N. and Zweimüller, M. (2014). Early Tracking and the Misfortune of Being Young. *Scandinavian Journal of Economics*, 116(2):394–428.

Schnepf, S. V. (2002). A Sorting that Fails? The Transition from Primary to Secondary School in Germany. *UNICEF Innocenti Working Paper No. 92*.

Schuchart, C. (2006). *Orientierungsstufe und Bildungschancen: Eine Evaluationsstudie*. Waxmann, Münster, Germany.

Schütz, G., Ursprung, H. W., and Wößmann, L. (2008). Education Policy and Equality of Opportunities. *Kyklos*, 61(2):279–308.

Segal, C. (2008). Classroom Behavior. *Journal of Human Resources*, 43(4):783–814.

von Friedeburg, L. (1992). *Bildungsreform in Deutschschland: Geschichte und Gesellschaftlicher Widerspruch*. Suhrkamp Verlag.

Wagner, G. G., Frick, J. R., and Schupp, J. (2007). The German Socio-Economic Panel Study (SOEP): Scope, Evolution and Enhancements. *Schmollers Jahrbuch*, 127(1):139–169.

Waldinger, F. (2007). Does Ability Tracking Exacerbate the Role of Family Background for Students' Test Scores?

Woessmann, L. (2010). Institutional Determinants of School Efficiency and Equity. *Jahrbücher für Nationalökonomie und Statistik*, 230(2):234–270.

Ziegenspeck, J. W. (2000). *Handbuch Orientierungsstufe: Sachstandsbericht und Zwischenbilanz*. Klinkhardt, Bad Heilbrunn, Germany.