Tracking and the Intergenerational Transmission of Education: Evidence from a Natural Experiment*

Simon Lange† Marten von Werder‡

May 21, 2015

Abstract

Proponents of tracking argue that the creation of more homogeneous classes increases efficiency while opponents fear that tracking aggravates initial differences between students. We estimate the effects on the intergenerational transmission of education of a reform that delayed tracking by two years in one of Germany's federal states. While the reform had no effect on educational outcomes *on average*, it increased educational attainment among individuals with uneducated parents and decreased attainment among individuals with educated parents. The reform thus lowered the gradient between parental education and own education. The effect is driven entirely by changes in the gradient for males.

Keywords: tracking; educational institutions; educational inequality; equality of opportunity; intergenerational mobility.

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

*The authors would like to thank Rafael Aigner, Mikkel Barslund, Ronny Freier, Stephan Klasen, Steffen Lohmann, Malte Reimers, Ramona Rischke, Viktor Steiner, Sebastian Vollmer, as well as 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

Can education policies help level the playing field between students from different social backgrounds? And if so, how? One major design feature of education systems that has frequently been related to equality of opportunity is tracking, the practice of grouping students by ability (Betts, 2010). Countries differ widely in the way they track students[1] but almost all education systems in practice feature some kind of tracking. Proponents 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 this paper we exploit a policy reform implemented during the 1970s in one of Germany's federal states, Lower Saxony. While most German states continued tracking students after fourth grade, the reform shifted the timing of tracking from grade four to grade six (roughly age ten and twelve, respectively). This was achieved through the introduction between 1972 and 1982 of a new intermediate school, the *orientation stage*[2] (henceforth, OS). We investigate the effects of this reform on the gradient between parental education and own education[3] based on a difference-in-differences (DD)-framework, comparing changes in this gradient across cohorts and states.

On average, the reform neither increased years of education nor the likelihood of being eligible to apply for university nor university graduation. We find, however, that the reform had a significant negative effect on the gradient in terms of years of education (which we measure as the time usually required to obtain the highest degree earned): years of education among individuals with uneducated parents increased by about one-third of a year and decreased by about the same time for individuals with educated parents. There is no evidence for a corresponding effect for females. Hence, the gap in years of education decreased by 1.4–1.6 years for males and by about 0.7–0.8 years overall. We observe a similar pattern when we investigate the impact on the likelihood of being eligible for university and obtaining a university degree. These results are robust to a wide range of changes in the empirical specification and in sample definitions.

In an extension to our analysis, we find no evidence for a positive effect of the reform on parental involvement, one of its stated aims. We argue that our results are consistent with the presence of peer effects and improvements in the allocation of students to tracks associated with de-tracking. We discuss recent evidence from the literature on gender differences in the development of non-cognitive skills and conclude that such differences potentially explain the gendered pattern in our findings.

The German reform we study offers a particularly well-suited setting to study the effects of

[1] 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).

[2] In German: *Orientierungsstufe*.

[3] We define this gradient as the gap in educational outcomes between individuals with educated parents and those with uneducated parents.

tracking: education policies in Germany are to a large extent 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 unlikely to be subject to confounding trends in unobserved socio-economic and institutional factors.[4] Because of Germany's system of equalization payments between states and sharing of tax revenues, there is 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 intergenerational mobility play a negligible role in Germany's education system. Finally, Germany stands out among OECD countries for early tracking at which large effects on educational careers may be expected.

Only a limited number of studies investigate the long-term effects of changes in the tracking regime.[5] Identification strategies that rely on cross-sectional variation (e.g. Bauer and Riphahn, 2006) potentially suffer from bias due to selection-on-unobservables or omitted variables (Betts and Shkolnik, 2000; Pischke and Manning, 2006; Waldinger, 2007). Hanushek and Woessmann (2006) exploit differences in the degree of tracking across countries based on an identification strategy that replaces changes within countries with changes across grades. They find a positive effect of tracking on performance inequality and no effect on performance levels. Waldinger (2007), however, presents evidence hinting at the presence of an omitted variable that affects both the intergenerational transmission of education and the likelihood of a country to implement early tracking.

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 impact 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.

There are also several recent studies that investigate policy reforms within countries. Malamud and Pop-Eleches (2011) exploit a policy reform implemented in Romania in 1973 to study the effect of postponing tracking on educational outcomes. The reform increased significantly the proportion of students in general as opposed to vocational secondary schools. While they find a sharp increase in the proportion of students eligible to apply for university overall and among disadvantaged students, they do not find an increase in the likelihood of disadvantaged students to attend university. Their results are thus in stark contrast to ours.[6]

Studies similar to our own rely on variation across both cohorts and space to analyze de-tracking reforms in Scandinavia. Meghir and Palme (2005), Pekkarinen et al. (2009), and Kerr et al. (2013) all report findings that are broadly in line with ours. Hall (2012), however, finds no effect of such a reform on university enrollment in Sweden. An important difference is that

[4] Waldinger (2007) argues that this may be an issue in cross-country comparisons such as Hanushek and Woessmann (2006). See, however, Woessmann (2010).

[5] See Betts (2010) for a review.

[6] Note, however, that the setting also differs in many important ways.

country-wide reforms in Scandinavia involved other significant changes to educational institutions and it is not possible to disentangle the effects of de-tracking from other components of these reforms.[7] Lower Saxony's reform, in contrast, entailed only adjustments to curricula in addition to delayed tracking. Nevertheless, our results below suggest that the pro-equality character of Germany's reform is broadly comparable to that of Scandinavian reforms.

The remainder of this paper is organized as follows. The next section discusses channels through which tracking may affect educational outcomes and educational inequality. Section 3 describes the German education system and the process that finally led to the introduction of OS schools in Lower Saxony. Section 4 describes the data and our identification strategy. Section 5 presents our main results and section 6 a number of robustness checks. Section 7 investigates one potential transmission channel, parental involvement with their children's academic performance. Section 8 offers an interpretation of our results. Section 9 concludes.

2 Conceptual framework

2.1 Peer effects

Tracking may affect the efficiency of educational production as well as the variation in outcomes through altering the relationship between individual and peer quality.[8] 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 peers (see, *inter alia,* Sacerdote, 2001; Ding and Lehrer, 2007; Lavy et al., 2011). Hence, if tracking succeeds in matching peers of similar quality, one would expect an increase in the overall variation in achievement. Moreover, if own achievement is linked to parental education, one would expect to find a positve relationship between tracking and the gradient between parental education and own education.

At the same time, it is often argued that tracking may allow for efficiency gains in educational production. 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 math scores benefits students.

[7]Sweden's reform entailed an increase in compulsory years of schooling and cash transfers to compensate families for foregone earnings. De-tracking in Finland coincided with the abolition of a vast network of private schools which were placed under municipal authority.

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

2.2 Misallocation of students to tracks

A second point relates to the precision with which students' are tracked. While track choice should arguably be based on academic potential, it is often maintained that this is difficult to observe initially and 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; DiPrete and Jennings, 2012).

There is indeed evidence that early tracking fails in separating students effectively by academic potential, particularly in Germany. First, there is considerable overlap in test scores between different school tracks (Baumert et al., 2003). Lehmann and Peek (1997) find that students with uneducated parents have to score significantly 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 education: Puhani and Weber (2007), Mühlenweg and Puhani (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.[9] 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.[10] 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.

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

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

3 Background: Germany's education system

3.1 Tracking in Germany's education system

Most states in Germany track students into three different types of secondary schools at the end of fourth grade when students are about ten years old:[11] low-achieving primary students usually attend the lower vocational track. The leaving certificate, awarded after five additional years of schooling, qualifies graduates to enroll in upper secondary vocational training courses (i.e. apprenticeships in the dual-system). Students with about average marks from primary education usually attend 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. Enrollment in the academic track is recommended 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 that permits students to apply to a university.

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 (Dustmann et al., 2014): first, teachers in the academic track usually receive higher wages. Second, their university degree differs in terms of requirements, is more subject-oriented, and also typically takes one additional year to complete. Third, 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. Finally, Brunello and Checchi (2007) report that the ratio of students to teachers in German schools varied substantially across tracks in 2004 with 11.89 students per teacher in the general track and 21.25 students per teacher in the vocational track.

Tracking across schools may be inconsequential if the education system exhibits 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.

[11] 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.

On the other hand, upgrading upon completion of one of the two vocational tracks is quite common and is often cited as evidence 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 if their academic record meets requirements. 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.

3.2 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 and efforts to reform the system were evident in the early 20th century. However, they were not realized until the post-war decades when some states started experimenting with less stringent forms of tracking in a few selected schools. The goal at the time was to improve the selection of students into academic careers. However, efforts in this direction were terminated in Lower Saxony in 1964 (Schuchart, 2006). At the end of the 1950s, a federal advisory board formally recommended a prolongation of comprehensive schooling until sixth grade (Deutscher Ausschuss, 1959), a recommendation that 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, the *German Education Council*, presented a blueprint for structural reforms that first mentioned the introduction of an *orientation stage* 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. While all states would agree in principle to changes to the school system, the compromise would leave the decision over whether or not to delay tracking to the states (Ziegenspeck, 2000, p. 81). While initial trials with OS-type schools were evident in several states, ultimately, only Lower Saxony and the city state of Bremen would subsequently track students only at the age of twelve.[12]

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

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

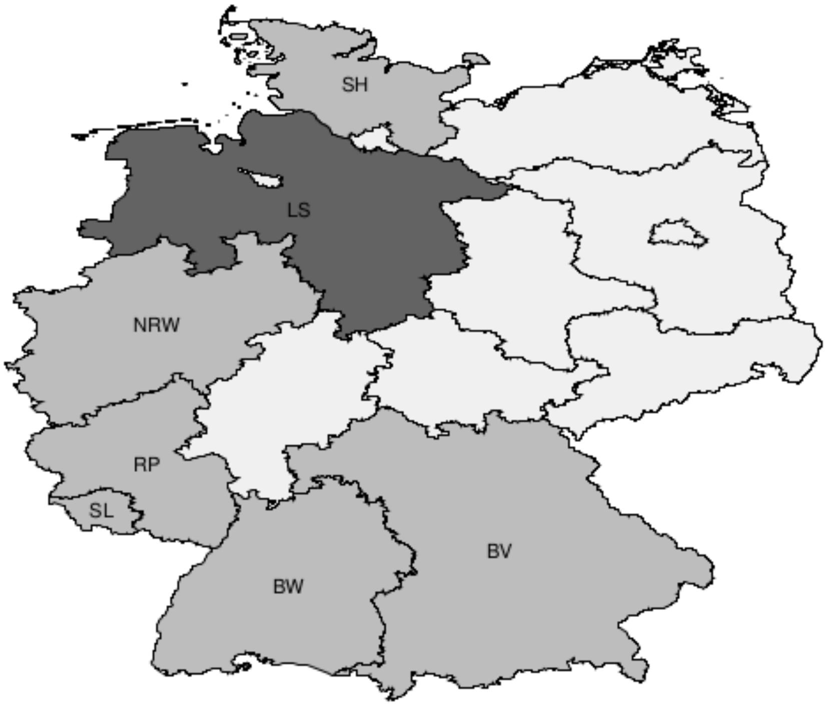

Figure 1: Treatment (dark gray) and control states (light gray). From North to South: SH (Schleswig-Holstein), LS (Lower Saxony), NRW (Northrhine-Westphalia), RP (Rhineland-Palatinate), SL (Saarland), BW (Baden-Württemberg), and BV (Bavaria).

differ in many important ways. We also exclude the state of Hesse which introduced an OS-style school but retained schools in 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 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 yet they eventually failed to abolish tracking after fourth grade.[13] 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. In Lower Saxony, OS schools were the norm until they were abolished in 2004.

There are several aspects of the design of the OS in Lower Saxony over and above delayed

[13]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).

tracking that are important for the present study: 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 OS schools were affiliated with secondary schools in the old three-tiered system in what amounts to an inconsequential re-labeling of fifth and sixth grade in the traditional 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). Students were also 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 the comparisons that we rely on in our analysis below are between tracking across schools and 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 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 also 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 associated with creating homogeneous groups with an effort to dilute the influence of parental background.

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 carried out on an annual basis between 1984 and 2012. See Wagner et al. (2007) for a description of the dataset.

Table 1: Means of key variables by state.

	LS	BV	BW	NRW	RP	SH
Panel A. Outcomes						
Years of education	12.62	12.49	12.69	12.89	12.33	12.55
Eligible for university	0.39	0.37	0.41	0.45	0.36	0.39
University degree	0.23	0.24	0.25	0.26	0.22	0.21
Panel B. Individual background characteristics						
Panel B1. Basic characteristics						
Age	45.20	45.39	44.59	45.44	46.45	44.81
Male	0.48	0.48	0.49	0.47	0.47	0.48
Educated parent	0.37	0.35	0.35	0.36	0.30	0.45
Panel B2. Childhood place of residence was mostly...						
...city.	0.12	0.16	0.12	0.30	0.10	0.15
...large town.	0.16	0.12	0.19	0.23	0.15	0.21
...small town.	0.22	0.22	0.21	0.21	0.22	0.22
...rural area.	0.46	0.46	0.42	0.20	0.49	0.39
Panel B3. Migratory background is...						
...direct	0.02	0.03	0.06	0.04	0.03	0.01
...indirect	0.05	0.08	0.12	0.07	0.06	0.05
Panel B4. Sample information						
Last observed in 2012	0.56	0.57	0.53	0.56	0.60	0.49
Observations	1,593	2,430	1,977	3,484	981	488

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

Outcome variables in our analysis are total years of education and two binary variables indicating whether an individual has attained the highest school-leaving certificate (i.e. is eligible to apply to a university without restrictions) and whether an individual has graduated from a university. Means are reported in panel A of table 1. 'Years of education' in this dataset refers 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.[14] Grade repetition will thus not be reflected in this measure of educational attainment.

Variables that serve as controls in our analysis capture socio-demographic characteristics such as year of birth, gender, and migrant status. We also include in our analysis a complete set of indicators for the size of the respondent's locality during childhood.[15] Means of these variables are reported in panel B of table 1.

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

[14]The maximum this variable takes is 18 years, 13 years until the highest school-leaving certificate plus five years in order to obtain a university degree. Obtaining a BA degree or a degree from a university of applied sciences 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.

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

certificate and the type of tertiary schooling or vocational training completed. We code a binary variable equal to unity if either the mother or father has (i) attained the highest school-leaving certificate, has (ii) completed the upper vocational track of secondary *and* a full course vocational education, or has (iii) completed training as a clerk, a public health worker, a civil servant, or an engineer, or holds a degree from a tertiary education institution.[16] 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).

Logit regressions reveal that our indicator of parents' educational attainment is highly predictive of parents' type of occupation[17] and the likelihood that respondents engaged in extracurricular activities at the age of 15 (such as having actively played a musical instrument or having done sports).[18] To further investigate differences between educated and uneducated parents, we also correlate this variable with an indicator of parents' preoccupation with respondents' academic achievement. This variable is coded on a five-point scale ranging from 'very much' to 'not at all.' We code the highest two levels as a binary variable. Results from logit regression in which we condition on own year of birth, parents' years of birth, parents' age, and the state of the last school visit indicate that having educated parents increases the odds that they were at least 'rather preoccupied' by a factor of two and 2.5 for males and females, respectively.

The GSOEP data do not provide direct information about the type of school individuals 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 an OS school in a given school year.[19] 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_c against birth years for individuals that attended schools in Lower Saxony.

There is usually no direct information in the GSOEP dataset about the sate in which an individual resided at the age of ten. A second challenge is thus to infer the state in which

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

[17] Conditional on fathers' (mothers') age and year of birth, we find that being educated increases their odds of having worked as a skilled, white-collar worker by a factor greater than four (three). These estimates are highly significant.

[18] Conditional on own year of birth, parents' years of birth, and parents' age, having educated parents increases the odds of having played an instrument and having done sports by a factor of 2.5 and 2.6, respectively.

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

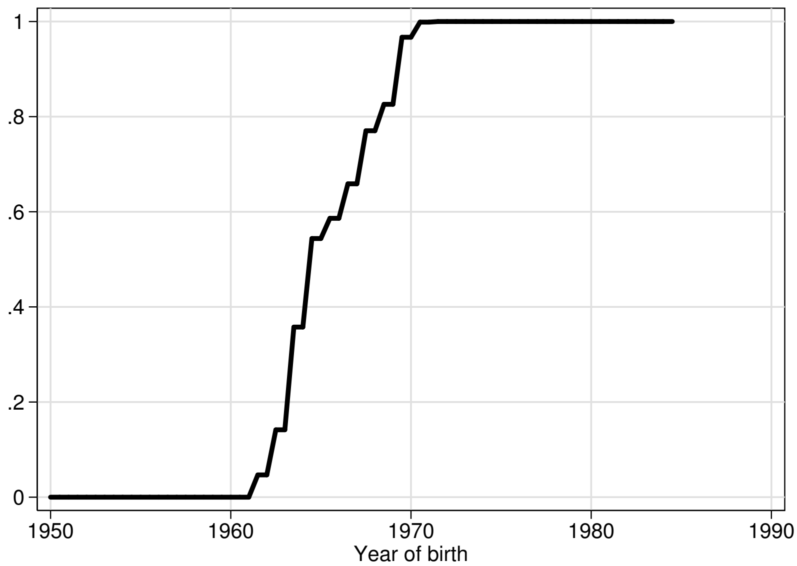

Figure 2: OS share against birth year in Lower Saxony.

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 and the state of the last school visit. We also drop observations on individuals that did not attend school in one of the West German states. The exact procedure is described in appendix A. While these steps are reasonable, there will still be misclassification error. Note, however, that confounding treatment and control will usually bias coefficients towards zero. Our estimates may 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 B4 of table 1).

As was discussed in section 3, we retain only observations on individuals that are likely to have received their schooling in one of seven states. Due to low case counts and associated privacy issues, two states, Rhineland-Palatinate and Saarland, were until recently treated as one entity in the dataset. We further retain only observations on individuals above the age of 28 as these are likely to have completed their education.[20] In consequence, all individuals in our sample were born before 1985. We also exclude individuals born before 1950.

[20] 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.

4.2 Empirical specifications

We estimate the causal effect of the reform on educational outcomes and inequality in educational outcomes and conduct this analysis separately for males and females. Our analysis is retrospective: individuals are observed only after they have acquired their education. We employ a DD estimator with a first difference across cohorts and a second difference across individuals in different states. Since the school year starts after the summer, a cohort c consists of all individuals born during the second half of a given year or during the first half of the subsequent year.

Define the treatment variable as the product of the percentage of a cohort in OS schools in Lower Saxony and an indicator variable for Lower Saxony, i.e. $D_{sc} = OS_c \times LS_s$. Note that D_{sc} is not a binary variable but takes on a range of values between zero and unity for cohorts born during the 1960s. We are primarily interested in the effect of the reform on the gradient between parental education and own education. Therefore, we include an interaction between the treatment variable and an indicator variable that is unity whenever the individual has educated parents. Our main model can be written as

$$y_{isc} = \beta_1(D_{sc} \times I[b = high]) + \beta_2 D_{sc} + \mathbf{x}'_{isc}\gamma + \lambda_s^b + \tau_c + \epsilon_{isc}, \tag{4.1}$$

where y_{isc} is the educational outcome for individual i with parental background $b \in \{low, high\}$ who went to school in state s and is a member of cohort c. As outcomes we consider years of education (as defined above) and binary indicators of university eligibility and graduation. For the latter two, the model is a linear probability model (LPM).[21]

$\mathbf{x}_{isc}$ is a matrix of time-invariant demographic and socio-economic variables. We include indicators for the respondent's gender, migrant status, and locality size during childhood. Preliminary data analysis suggests that there was considerable variation in the gradient across states prior to the reform. We therefore allow the gradient to differ across states by including state-background-fixed effect denoted λ_s^b. τ_c denote cohort-fixed effects. When the sample is pooled across males and females, λ_s^b, 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.

The parameters of interest are β_1 and β_2. The former measures the effect of the reform on the gradient and the latter the effect on individuals with uneducated parents. The effect on individuals with educated parents is calculated as the sum over both coefficients.

The reform was implemented during a time of a nation-wide educational expansion, mostly a response to rapid population growth during the 1960s and 1970s. We therefore estimate an alternative model that allows for differences in trends in educational attainment across states by including state-cohort-fixed effects, denoted ϕ_{sc}:

$$y_{isc} = \beta_1(D_{sc} \times I[b = high]) + \mathbf{x}'_{isc}\gamma + \lambda_s^b + \phi_{sc} + \epsilon_{isc}. \tag{4.2}$$

[21]Estimating probit and logit models does not alter our results.

Note that the main effect of the reform is no longer identified and thus excluded here. β_1, however, is still identified from within-cohort, within-state variation.

The identifying assumption for the causal effect of the reform on educational outcomes in the framework above is that trends between states would not have systematically differed in the absence of the reform. While this may be reasonable, the assumption required to obtain a causal effect of the reform on the gradient was so far more demanding: we required that the gradient remains constant over time except for the effect of the reform. We relax the latter assumption by (1) allowing for common changes in the gradient across cohorts and (2) allowing for linear changes in the gradient that vary across states.

A specification that allows for changes in the gradient across cohorts that are common to all states is

$$y_{isc} = \beta_1(D_{sc} \times I[b = high]) + \mathbf{x}'_{isc}\gamma + \lambda_s^b + \psi_c^b + \phi_{sc} + \epsilon_{isc}, \tag{4.3}$$

where ψ_c^b denotes a set of cohort-background-fixed effects. Here, the assumption required to obtain a causal estimate of the effect of the reform on the gradient is that there would not have been any systematic differences in trends *in the gradient* absent the reform.

Since we observe several cohorts that turned ten before the introduction of OS schools in our dataset, we may also relax the assumption of common trends in the gradient in the absence of the reform by replacing cohort-background-fixed effects with linear state- and background-specific trends. This allows us to pick up state-specific changes in the gradient that are discernible prior to the introduction of the reform. The specification is

$$y_{isc} = \beta_1(D_{sc} \times I[b = high]) + \mathbf{x}'_{isc}\gamma + \lambda_s^b + \eta_s^b c + \phi_{sc} + \epsilon_{isc}, \tag{4.4}$$

where η_s^b denotes background- and state-specific parameters that multiply the cohort variable. Note that we again include state-cohort-fixed effects. (4.4) mostly serves a robustness check as linear trends are likely more restrictive than cohort-background-fixed effects.

A general concern one may have in estimating the above specifications is that the reform may have caused selective migration: if families decided to move from or to Lower Saxony in response to the reform and if this decision was correlated with parental education, our estimates 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 seems implausible that a reform that only affected two grades would have induced this behavior.

If there had been an important migratory response, however, the reform would be associated with a change in the state-specific probability of having educated parents. To check that this is not the case, we regress the binary indicator of having educated parents on the treatment variable 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

Table 2: Impact of the reform on years of education and inequality in years of education by gender: OLS estimates.

	(1)	(2)	(3)	(4)	(5)
Panel A. Pooled sample ($N = 10,953$)					
Reform × ed. parents		-0.613*	-0.731**	-0.798**	-1.542**
		(0.315)	(0.339)	(0.358)	(0.712)
Reform	0.076	0.305			
	(0.137)	(0.193)			
R-squared	0.164	0.164	0.190	0.195	0.192
Panel B. Females ($N = 5,698$)					
Reform × ed. parents		-0.066	-0.100	-0.048	-0.867
		(0.350)	(0.358)	(0.406)	(0.768)
Reform	-0.011	0.014			
	(0.180)	(0.226)			
R-squared	0.166	0.166	0.191	0.195	0.192
Panel C. Males ($N = 5,255$)					
Reform × ed. parents		-1.235**	-1.435***	-1.629***	-2.380*
		(0.500)	(0.507)	(0.543)	(1.217)
Reform	0.173	0.617**			
	(0.207)	(0.274)			
R-squared	0.158	0.159	0.186	0.192	0.187
Fixed effects:					
Cohort	✓	✓			
State-cohort			✓	✓	✓
State-background	✓	✓	✓	✓	✓
Cohort-background				✓	
Linear cohort trend:					
State-background-specific					✓

Standard errors clustered at the state-cohort-level in parentheses. *, **, and *** denote significance at the ten-, five-, and one-percent level, respectively. All regressions include dummies for the size of respondents' childhood place of residence and 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.

an effect on the probability of having educated parents: all coefficients are close to zero and statistically insignificant. Selection due to strategic migration is not a confounding element in our analysis.

5 Results

We first consider years of education as the outcome variable. Results from estimating all the above equations are reported in table 2. Standard errors clustered at the state-cohort-level are reported in parentheses in all tables that follow.[22]

We exclude the interaction effect between parental background and the reform from (4.1) in column (1) in order to obtain an estimate of the average effect of the reform. Column (2) reports results from estimating (4.1) and column (3) from estimating (4.2), the model including state-cohort-fixed effects. Finally, columns (4) and (5) reports results from estimating (4.3) and (4.4), respectively.

Results reported in panel A are for the pooled sample. There is no indication here of an effect of the reform on years of education on average (column (1)). The coefficient is positive yet small and insignificantly different from zero. If we include the interaction term, however, we find that the reform had a significant negative effect on the gradient between parental and own education (column (2)). The point estimate on the interaction term is -0.61 years and significant at the ten-percent level. We may calculate from the two coefficients reported in column (2) the effect on individuals with uneducated and educated parents, respectively, as the estimate of the main effect, 0.31, and the sum of the two coefficients, -0.31. Both effects fall just short of being significant at the ten-percent level (p-values of 0.12 and 0.16, respectively).

Our conclusions regarding the reform's effect on the gradient does not change when we include state-cohort-fixed effects (column (3)). Our estimate of the effect on the gradient is now somewhat larger in absolute terms at -0.73 years and is significant at the five-percent level. Allowing for a common trend in the gradient over time results in a similar estimate (column (4)). Including linear, state- and background-specific cohort trends results in an increase in the estimate in absolute terms by a factor of close to two. However, the standard error is also considerably larger. The coefficient remains significant at the five-percent level. This suggests that state-specific trends in the gradient that were present prior to the introduction of the reform cannot explain our finding.

Turning to panels B and C of the same table it is apparent that the effect on the gradient is entirely due to an effect on males. Both the estimate for the overall effect of the reform (column (1)) and the estimate on the interaction term for females are close to zero and insignificant at conventional levels. The coefficient estimate reported in column (5) is large and negative but also insignificant. Estimates in panel C, on the other hand, suggest that the gap in years of education narrowed considerably for males. The coefficient estimate is -1.24 in column (2) and -1.44 when we include state-cohort-fixed effects. Both estimates are statistically significant. A similar estimate is obtained when we allow for cohort-background-fixed effects (column (4)). Again, we

[22] We also experimented with conventional, robust ('sandwich'-type) standard errors, and standard errors clustered at the level of cohorts and states. Clustering at the state-cohort-level turned out to be the most conservative option and not very different from conventional standard errors, robust standard errors, and standard errors clustered at the level of cohort. Clustering at the state-level resulted in much smaller standard errors. We address potential problems with standard errors in a robustness check in section 6.1 below.

Table 3: Impact of the reform on the probability of being eligible for university: OLS estimates.

	(1)	(2)	(3)	(4)	(5)
Panel A. Pooled sample ($N = 10,953$)					
Reform × ed. parents		-0.050	-0.081	-0.101	-0.159
		(0.063)	(0.066)	(0.065)	(0.137)
Reform	-0.008	0.010			
	(0.024)	(0.037)			
R-squared	0.123	0.123	0.149	0.153	0.150
Panel B. Females ($N = 5,698$)					
Reform × ed. parents		0.021	0.002	0.013	0.028
		(0.075)	(0.076)	(0.078)	(0.148)
Reform	-0.023	-0.032			
	(0.032)	(0.045)			
R-squared	0.126	0.126	0.151	0.155	0.153
Panel C. Males ($N = 5,255$)					
Reform × ed. parents		-0.130	-0.173**	-0.227***	-0.391**
		(0.082)	(0.082)	(0.083)	(0.172)
Reform	0.009	0.056			
	(0.032)	(0.044)			
R-squared	0.100	0.101	0.127	0.133	0.129
Fixed effects:					
Cohort	✓	✓			
State-cohort			✓	✓	✓
State-background	✓	✓	✓	✓	✓
Cohort-background				✓	
Linear cohort trend:					
State-background-specific					✓

Standard errors clustered at the state-cohort-level in parentheses. *, **, and *** denote significance at the ten-, five-, and one-percent level, respectively. All regressions include dummies for the size of respondents' childhood place of residence and 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.

observe that both the estimate and the associated standard error increase in absolute terms when we replace cohort-background effects with linear cohort trends. However, the estimate remains significant at the ten-percent level.

We conclude that the reform had no effect on average but narrowed the gap in years of education between sons of educated and sons of uneducated parents. Coefficient estimates suggest a decrease in years of education of about three-fifths of a year for the latter and this effects is significant at the five-percent level (column (2)). The sum over both coefficients in column (2), panel C, the effect of the reform on years of education for males with educated parents, is -0.62 and close to being significant at conventional levels (p-value of 0.11).

Next, table 3 presents estimates of the effect of the reform on the probability of obtaining the highest school-leaving certificate (and being thus eligible for university). The general pattern we observe is similar to that for years of education: there is no effect of the reform on this outcome on average. The coefficient on the interaction term is negative for the pooled sample. But, in contrast to our results for years of education, it is not statistically significant. The point estimates in this case suggest that the reform lowered the gap in the proportion of individuals that have attained a degree by five to 16 percentage points.

There is no evidence that the reform in any way altered the probability of university eligibility among females. All coefficients are close to zero and insignificant (panel B, table 3). On the other hand, the reform seems to have increased equality of opportunity in this respect among males: the coefficient estimates in columns (3) and (4), for instance, suggest that the gap in the probability of attaining eligibility decreased by 17 and 23 percent and these estimates are significant at the five- and one-percent level, respectively. As above, the effect seems more pronounced when we include linear trends but is less precisely estimated (column (5)).

As one would expect given the previous results, there is no indication that the reform increased the probability of obtaining a university degree on average (column (1) of table 4). The effect on the gradient, however, is again seizable even for the pooled sample: once we control for state-cohort-fixed effects (columns (3)–(5)), the coefficient estimates on the interaction term indicate that the gap decreased by about 15 percentage points. These estimates are significant at the one-percent level except when we include state-background-specific cohort trends. However, this is likely to be a power issue as the point estimate hardly changes while the standard errors increase by a factor of about two. Again, the effect is driven by males for which the gap decreases by more than 20 percentage points according to most estimates (panel C).

6 Robustness

In this section we conduct different robustness checks that address potential problems with our analysis. We first investigate the robustness of our results to aggregating data at the level of cohort-states. Next, we investigate whether our results are robust to alternative sample restrictions.

6.1 Accounting for grouped data

The previous section shows that the result of a decrease in the gradient for males is robust to several alternative estimation methods. One concern may be that the treatment variable varies only at the state-cohort-level. Since the number of clusters is limited, standard errors may be unreliable. Following Angrist and Pischke (2009, p. 313), we collapse our dataset at the state-cohort-level. For each dependent variable of interest, we compute the mean difference between

Table 4: Impact of the reform on the probability of graduating from university: OLS estimates.

	(1)	(2)	(3)	(4)	(5)
Panel A. Pooled sample ($N = 10,953$)					
Reform × ed. parents		-0.123**	-0.144**	-0.155***	-0.166
		(0.053)	(0.056)	(0.057)	(0.121)
Reform	-0.000	0.046			
	(0.023)	(0.029)			
R-squared	0.112	0.113	0.143	0.150	0.145
Panel B. Females ($N = 5,698$)					
Reform × ed. parents		-0.042	-0.060	-0.065	-0.132
		(0.064)	(0.065)	(0.072)	(0.159)
Reform	-0.031	-0.014			
	(0.027)	(0.034)			
R-squared	0.104	0.104	0.135	0.142	0.136
Panel C. Males ($N = 5,255$)					
Reform × ed. parents		-0.214**	-0.238***	-0.255***	-0.210
		(0.085)	(0.085)	(0.089)	(0.209)
Reform	0.034	0.111***			
	(0.035)	(0.042)			
R-squared	0.106	0.107	0.137	0.144	0.140
Fixed effects:					
Cohort	✓	✓			
State-cohort			✓	✓	✓
State-background	✓	✓	✓	✓	✓
Cohort-background				✓	
Linear cohort trend:					
State-background-specific					✓

Standard errors clustered at the state-cohort-level in parentheses. *, **, and *** denote significance at the ten-, five-, and one-percent level, respectively. All regressions include dummies for the size of respondents' childhood place of residence and 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.

individuals with educated and those with uneducated parents. We regress this gap for males and females separately on the reform variable and include both cohort- and state-fixed effects. We also include variables that capture the share of individuals with direct and indirect migratory background and four variables indicating the level of urbanization of the childhood place of residence. We estimate these models by weighted least squares using the group size as weights.

There are two advantages of this approach: first, it makes our identifying assumption with respect to the reform's effect on the gradient explicit: the gradient is the outcome variable

Table 5: DD estimates of the reform's effect on educational gradients based on grouped data and by gender.

	Females			Males		
	Years of education (1)	Eligibility (2)	Graduation (3)	Years of education (4)	Eligibility (5)	Graduation (6)
Reform	-0.035	0.011	-0.061	-1.736***	-0.259***	-0.256***
	(0.496)	(0.086)	(0.076)	(0.561)	(0.098)	(0.092)
R-squared	0.196	0.183	0.224	0.267	0.262	0.235
Observations	209	209	209	211	211	211

Asymptotic standard errors in parentheses. *, **, and *** denote significance at the ten-, five-, and one-percent level, respectively. Observations weighted by the number of individuals in each cohort-state. All regressions include sets of cohort- and state-fixed effects. We also include the share of individuals with direct and indirect migratory background and the share of individuals that grew up in rural areas, small towns, large towns, and cities in each cohort-state-cell as additional controls. Based on GSOEP v29-data.

in a standard difference-in-differences equation. Second, it is also more evident now that the asymptotics of our analysis are based on the number of cohort-state-groups. Because the group means are close to being normally distributed, however, the finite-sample properties should closely resemble those of a regression with normally distributed errors. Usual asymptotic standard errors are therefore more likely to be reliable than clustered standard errors.

Results reported in table 5 confirm our previous findings: for males, the introduction of OS schools decreased the gap in years of education by about 1.7 years, and the gaps in university eligibility and graduation by 26 and 25 percentage points, respectively. Standard errors are very similar to those reported in columns (4) of tables 2–4. There is no evidence for an effect of the reform on the gradient for females.

6.2 Alternative samples

We now analyze the robustness of our result concerning the effect of the reform on the gradient to changes in the underlying sample. We consider (1) only fully treated or untreated cohorts, (2) only non-migrants, (3) only individuals that grew up in non-rural areas, (4) only individuals that grew up in states with levels and trends in enrollment in pre-primary similar to those of Lower Saxony, and (5) only individuals that grew up with both parents present. In all cases, we focus on the effect on the gradient. We therefore estimate (4.3), the specification that allows for state-cohort-fixed effects and common trends in the gradient. We only consider males here as we did not find any effects for females. Estimating models on the respective samples for females does not allow any other conclusion (results not reported).

A first concern is self-selection into OS schools. The introduction of OS schools in Lower Saxony took almost one decade and during this time period selection may have played a role.

Given our findings above, it seems plausible that educated parents would have decided to send their children to schools in the old, three-tiered system as long as this was possible. In any case, the co-existence of two tracking regimes potentially produces outcomes that differ from those under complete de-tracking.[23] We therefore also estimate (4.3) including only cohorts that were either fully treated or fully untreated, i.e. cohorts born before 1962 or after 1971.

A second subsample considers only non-migrants. The effect of such a reform on migrants may differ significantly and there may also be problems in classifying the educational attainment of migrants' parents. Also, Lower Saxony has one of the smallest migrant populations in our sample.

A third subsample excludes individuals that grew up in rural areas. This is motivated by the fact that Lower Saxony is thinly populated in comparison to other states. It is the second-largest state of Germany in terms of area yet only the fourth-largest in terms of inhabitants. Kramer (2002) documents an increase in the access to upper vocational and academic track schools especially in rural areas of Lower Saxony following the construction of new schools during the 1960s and early 1970s. It may thus be that a differential change in access to different types of schools during the 1970s confounds our results.

An additional concern relates to the percentage of children enrolled in pre-primary, an educational institution that has been linked to inequality in educational outcomes (Schütz et al., 2008). Pre-school enrollment has repeatedly been shown to affect both cognitive and behavioral outcomes, particularly among disadvantaged children. See Barnett (1992), Currie (2001), Cunha et al. (2010), and Blau and Currie (2006) for reviews and Garces et al. (2002), Belfield et al. (2006), Magnuson et al. (2007), Berlinski et al. (2009), Schlotter and Wößmann (2010), and Schlotter (2011) for recent within-country evidence. Therefore, systematic differences in trends and levels in pre-school enrollment between states may affect outcomes and thus confound our results.

Differences in pre-primary enrollment are unlikely to explain our results for two reasons. First, much of the evidence on the effect of pre-school education comes from the US yet pre-school institutions in Germany are very different in that they offer mainly supervision and are not directly targeted towards disadvantaged segments of the population (Schlotter and Wößmann, 2010). While Schlotter (2011) finds some evidence that pre-school education positively affects assertiveness and the ability to make friends, Schlotter and Wößmann (2010) find no evidence for an effect on reading test scores.

Second, while there are differences in the levels of pre-school education, there are no major differences in trends. Our data do not contain any information about whether respondents attended any pre-school institution and official records by state are not readily available. The number of places available per one hundred children between the ages of three and six are tabulated, however, in Erning et al. (1987, p. 37) and we reproduce this information in figure 3. We observe that pre-school capacity was lower in Lower Saxony than in all other states except

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

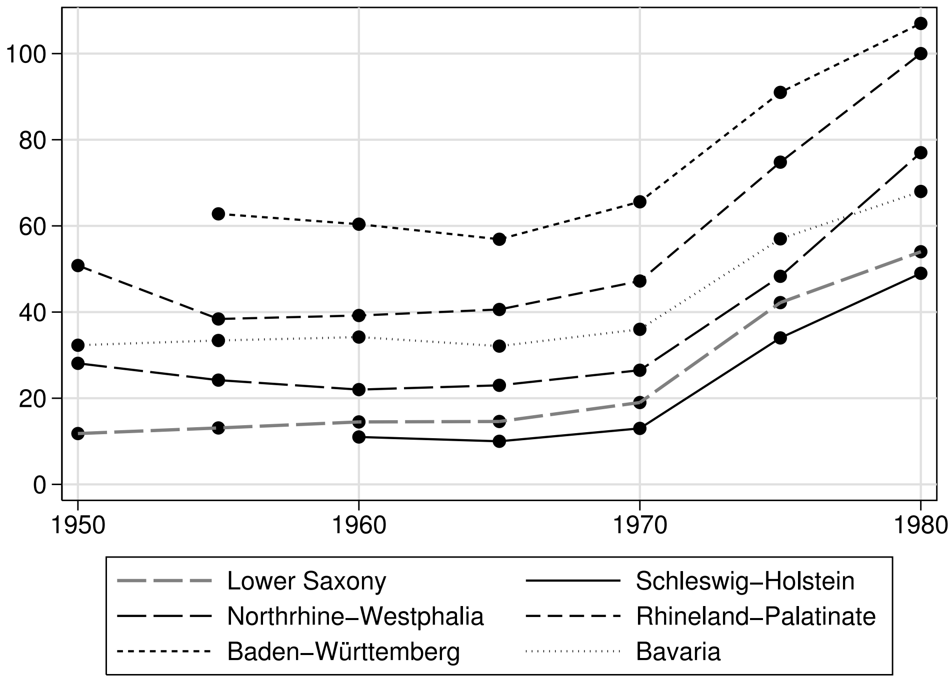

Figure 3: Pre-school capacity per 100 children between the age of three and six. *Source:* Erning et al. (1987, p. 37).

Schleswig-Holstein and that capacity increased substantially in all states from 1970 onwards. If the effect of enrollment would be linear, at least, we would not expect trends to be systematically correlated with the roll-out of OS schools in Lower Saxony.

It may be that the effect of pre-school education depends on the level of enrollment (Schütz et al., 2008). This will be the case, for instance, if there are positive effects of enrollment on educational attainment and children of educated parents are more likely to enroll initially. In this case, the effect of enrollment will follow an inverted u-pattern with equality of opportunity decreasing at low levels of enrollment and increasing at high levels. Schütz et al. find some evidence for such a relationship and report a turning point at an enrollment rate at about 60 percent. Note that this would result in *decreasing* inequality of opportunity in states with levels of enrollment near the turning point around 1970 (such as Baden-Württemberg) and *increasing* inequality in states with low levels of enrollment such as Lower Saxony.

Nevertheless, as an additional robustness check, we investigate the sensitivity of our results to excluding both Baden-Württemberg and Rhineland-Palatinate, two states that differed the most from Lower Saxony in that they had the highest levels of pre-school enrollment (see figure 3).

A final concern may be that systematic differences in family's demographic make-up confound

Table 6: Impact of the reform on inequality in educational outcomes for males. Robustness to alternative sample definitions.

	excl. partially treated (1)	excl. migrants (2)	excl. rural (3)	excl. BW and RP (4)	excl. incomplete families (5)
Panel A. Years of education.					
Reform × ed. parents	-1.872***	-1.744***	-1.822**	-1.485***	-1.691**
	(0.634)	(0.544)	(0.756)	(0.520)	(0.665)
R-squared	0.197	0.197	0.212	0.192	0.204
Panel B. University eligibility.					
Reform × ed. parents	-0.237**	-0.250***	-0.235**	-0.205**	-0.223**
	(0.101)	(0.085)	(0.108)	(0.079)	(0.095)
R-squared	0.142	0.139	0.158	0.138	0.141
Panel C. University degree.					
Reform × ed. parents	-0.333***	-0.265***	-0.277**	-0.232***	-0.285***
	(0.092)	(0.091)	(0.124)	(0.085)	(0.104)
R-squared	0.161	0.151	0.166	0.143	0.151
Observations	3,327	4,651	3,295	3,815	4,636

Standard errors clustered at the state-cohort-level in parentheses. *, **, and *** denote significance at the ten-, five-, and one-percent level, respectively. All regressions include sets of dummies capturing the size of respondents' childhood place of residence and migrant status. Based on GSOEP v29-data.

our results. In particular, the absence of a father or a mother may both affect our indicator of parental education and may also be related to educational outcomes. While it is not clear why there should be systematic differences between states, we nevertheless investigate the robustness of our results to the exclusion of all individuals that did not grow up in complete families, i.e. with a mother and father present. The information is obtained from two questions in the dataset: parents' death years and a question about whether the respondent had conflicts with the parent at the age of 15.

Results for years of education, university eligibility, and university graduation are reported in panels A, B, and C of table 6, respectively. Note that the sample size changes considerably in comparison to the main sample. The samples used in columns (1) through (5) retain 58, 82, 58, 67, and 81 percent of the original sample, respectively. Nevertheless, we find that our estimates for effect of the reform on the gradient do not change significantly and remain highly significant throughout. They range from -1.48 to -1.87 and are qualitatively similar when we use university eligibility or graduation as the outcome variable. Overall, none of the above concerns seems to affect our main result of an attenuating effect of the reform on the gradient for males.

Table 7: Impact of the reform on parental involvement: OLS estimates.

	Females			Males		
Parents' education	All	High	Low	All	High	Low
	(1)	(2)	(3)	(4)	(5)	(6)
Reform	-0.047	-0.080	-0.042	-0.029	0.015	-0.040
	(0.04)	(0.07)	(0.05)	(0.03)	(0.07)	(0.04)
Observations	5,698	2,045	3,653	5,255	1,875	3,380
R-squared	0.06	0.10	0.07	0.09	0.11	0.10

Standard errors clustered at the state-cohort-level in parentheses. *, **, and *** denote significance at the ten-, five-, and one-percent level, respectively. All regressions include state-of-school-visit-, cohort-, and age-fixed effects. Further controls are dummies capturing the size of respondents' childhood place of residence and migrant status. Based on GSOEP v29-data.

7 Effect of the reform on parental involvement

Previous sections show that the reform had no effect on educational outcomes on average. The effects on females that we find were small and insignificant. For males, on the other hand, we find that the effect varied by parental education with positive effects on individuals with uneducated parents and negative effects on individuals with educated parents. What accounts for this increase in equality of opportunity for males? In this section, we investigate one plausible channel through which the effects of the reform may have operated: parental preoccupation with their children's academic achievement.

The OS placed special emphasis on engaging parents with low educational attainment with the aim to increase the likelihood that they consider sending their children to the academic track granted teachers judged them suitable. Increased involvement of parents may thus be one channel through which educational attainment increased in some groups. Our data contain one item that captures the self-reported involvement of parents' with their children's academic performance. While we would prefer objective measures of parental involvement to self-reports and the question does not reference any particular grade, we are cautiously optimistic that this item serves us as a proxy. As discussed in section 4, the respective question asked respondents to rank how much parents cared about academic achievement on a five-point scale from 'very much' to 'not at all' and we code the top two responses as a binary variable. We regress this variable on the reform variable, our usual control variables, as well as fixed effects for schooling state and cohort. We also include age-fixed effects as the perception and judgment of past parental involvement may change with age. We run this regression separately for males and females and further disaggregate the analysis by parental education. Results are reported in table 7. While these coefficients are mostly negative and similar across subgroups, they are all insignificantly different from zero. There is thus no indication that the introduction of the OS increased parents' preoccupation with their children's academic achievement.

8 Discussion

Results presented in previous sections suggest that the reform achieved for male students what it was intended to achieve: it attenuated the importance of parental background for students' educational prospects. There is no evidence for a trade-off between efficiency and equity; the reform did not seem to have lowered overall educational attainment.

The effect on total years of education is comparatively large given that the reform delayed tracking by only two years: the effect we estimate on the gradient for males ranges from a reduction by 1.4 to 2.4 years of education. The preferred estimate is obtained from specifications that allow for common changes in the gradient across cohorts and state-cohort-fixed effects, i.e. the one reported in column (4), panel C of table 2, which indicates a decrease in the gap by 1.6 years. The unadjusted gradient in Lower Saxony prior to the reform was 3.2 years in Lower Saxony. It decreased to 1.9 years on average for all full treated cohorts in that state. Hence, it seems that the reform's negative effect on the gradient may have been to some extent abrogated by a modest increase in the gradient evident for other states.

Returns to one additional year of education for males in Germany are usually estimated to be around seven percent and often somewhat higher for tertiary education (Lauer and Steiner, 2000; Ammermüller and Weber, 2005). The implied changes in hourly wages for males with educated and uneducated parents that we find, $(0.617-1.235)\times 0.07 \approx -4.3$ percent and $0.617\times 0.07 \approx 4.3$ percent, would thus be neither very large nor negligible. They are well in line with findings in Meghir and Palme (2005) who report estimates of -7.7 percent and 3.1 percent, respectively. The main difference is that we find no effects for females while Meghir and Palme find a positive effect for females with uneducated fathers (about 3.8 percent) and a negative effect for females with educated fathers (about -4.2 percent).

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.

A plausible explanation of our main findings is that tracking age interacts with gender differences in the development of non-cognitive skills. This would require that educated parents have an advantage in the production of these skills, that the resulting advantage dissipates over time (for instance, through in-class peer effects), and that at least one of the above is more relevant for boys than for girls. If these skills enhance learning or educators reward these skills, we would expect to find that delayed tracking is associated with a less prominent role of parental education for boys.

Differences in the levels and trajectories of cognitive and non-cognitive development between boys and girls are well-established in the literature. Matthews et al. (2009) 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. Lenroot et al. (2007) find important differences in the trajectories of brain development between boys and girls with boys generally trailing behind.

Recently, gender differences in non-cognitive skills have directly been linked to the gender gap in academic achievement. Based on data from the US, DiPrete and Jennings (2012) show that while there are no gender gaps in the returns to social and behavioral skills for children from kindergarten though fifth grade, girls lead boys by nearly 0.4 standard deviations at the start of kindergarten. They demonstrate that this gap grows over time and explains a considerable fraction of the gender gap in academic outcomes at that age. Moreover, these skills are significantly related to parents' socio-economic status and the presence of a father. While they reckon that this may reflect teachers rewarding social and behavioral skills, they also find evidence that these skills enhance learning. This is in line with Kerr et al. (2013) who find benefits from comprehensive schooling in terms of higher cognitive skills for students from disadvantaged backgrounds.

9 Conclusion

Recent research suggests that design features of education systems are an important determinant of the strength of the relationship between parental and own education. While tracking has often been found to increase variation in outcomes and the salience of parental background, it 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 one of Germany's federal states in the 1970s affected educational outcomes of individuals differentiated by gender and parental education. Based on a difference-in-differences estimation strategy, we find no evidence that the reform led to a decrease in educational attainment for cohorts affected, that is, there is no evidence that tracking increases efficiency. We present strong evidence, however, that the reform is associated with increased equality of opportunity. The effect is entirely driven by males: the reform benefited males with uneducated parents at the expense of males with educated parents. These findings are robust to a number of alternative specifications and restrictions imposed on the underlying sample. There is no indication that the reform had any effect on females.

We find no evidence that the reform affected parental involvement, one of its main aims. It seems plausible that the gender differences we observe are related to changes in the development of non-cognitive skills for males. However, more work on the precise transmission channel is clearly warranted. The interaction between gender differences in skill development and tracking seems an interesting alley for future research.

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.

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.

Barnett, S. (1992). Benefits of Compesatory Preschool Education. *Journal of Human Resources*, 27(2):279–312.

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

Bauer, P. and Riphahn, R. T. (2006). Timing of School Tracking as a Determinant of Intergenerational Transmission of Education. *Economics Letters*, 91:90–97.

Baumert, J., Trautwein, U., and Artelt, C. (2003). 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.

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

Belfield, C. R., Nores, M., Barnett, S., and Schweinhart, L. (2006). The High/Scope Perry Preschool Program. *Journal of Human Resources*, 41(1):162–190.

Berlinski, S., Galiani, S., and Gertler, P. (2009). The Effect of Pre-primary Education on Primary School Performance. *Journal of Public Economics*, 93:219–234.

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.

Blau, D. and Currie, J. (2006). Pre-School, Day Care, and After-School Care: Who's Minding the Kids? In Hanushek, E. A. and Welch, F., editors, *Handbook of the Economics of Education*, volume 2, pages 1163–1278. Elsevier, Amsterdam.

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.

Cunha, F., Heckman, J. J., Lochner, L., and Masterov, D. V. (2010). Interpreting the Evidence on Life Cycle Skill Formation. In Hanushek, E. A. and Welch, F., editors, *Handbook of the Economics of Education*, volume 1, pages 697–812. North Holland, Amsterdam.

Currie, J. (2001). Early Childhood Education Programs. *Journal of Economic Perspectives*, 15(2):213–238.

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.

DiPrete, T. A. and Jennings, J. (2012). Social/Behavioral Skills and the Gender Gap in Educational Achievement. *Social Science Review*, 41:1–15.

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., 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.

Erning, G., Neumann, K., and Reyer, J. (1987). *Geschichte des Kindergartens: Institutionelle Aspekte, Systematische Perspektiven, Entwicklungsverläufe*, volume 2. Lambertus, Freiburg, Germany.

Garces, E., Thomas, D., and Currie, J. (2002). Longer-Term Effects of Head Start. *American Economic Review*, 92(4):999–1012.

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.

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.

Kramer, W. (2002). *50 Jahre Schulenwicklung in Niedersachen*. Bibliotheks- und Informationssystem der Universität Oldenburg, Oldenburg, Germany.

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.

Magnuson, K. A., Ruhm, C., and Waldfogel, J. (2007). Does Prekindergarten Improve School Preparation and Performance. *Economics of Education Review*, 26:33–51.

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

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.

Schlotter, M. (2011). Age at Preschool Entrance and Noncognitive Skills before School—An Instrumental Variable Approach. *Ifo Institute for Economic Research Working Paper No. 112.*

Schlotter, M. and Wößmann, L. (2010). Frühkindliche Bildung und Spätere Nicht-kognitive Fähigkeiten: Deutsche und Internationale Evidenz. *Ifo Institute for Economic Research Working Paper No. 91.*

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.

A Inferring the state of schooling at age ten

We proceed as follows in order to narrow down the state in which individuals received schooling at the age of ten:

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 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.

Table 8 reports the frequencies with which individuals were classified according to the above steps by states. It shows that for the majority of individuals, we make the assumption that the state of last school visit is also the state they lived in at the age of ten. Only ten to 15 percent of individuals are classified based on the first state in which they are observed.

Table 8: Provenance of information on state of school visit at the age of ten.

Schooling state is...	LS	BV	BW	NRW	RP	SH
...state at age ten.	0.03	0.04	0.05	0.04	0.02	0.02
...state of last school visit.	0.71	0.70	0.67	0.68	0.68	0.73
...childhood state.	0.15	0.14	0.14	0.16	0.19	0.14
...first state in which observed.	0.10	0.12	0.13	0.12	0.10	0.11
Observations	1,593	2,430	1,977	3,484	981	488

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