CESifo WORKING PAPERS

9214
2021

July 2021

# How Does Reducing the Intensity of Tracking Affect Student Achievement? Evidence from German State Reforms

*Marc Piopiunik*

CESifo

## Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 9214*

# How Does Reducing the Intensity of Tracking Affect Student Achievement? Evidence from German State Reforms

## Abstract

To investigate the effects of reducing the intensity of tracking, this study exploits reforms across German states which combined the two lower secondary school tracks, sometimes additionally offering the possibility to acquire a university entrance qualification. Using a difference-in-differences approach, we find that reducing the tracking intensity significantly improves students' reading achievement. Lower-performing student groups – boys, students born abroad, and students from lower socio-economic status families – benefited in particular. In contrast, we find no effects on acquiring a middle school degree, attending the most academic track, or repeating a grade.

JEL-Codes: I210, I240, I280.

Keywords: school tracking, student performance, NEPS.

*Marc Piopiunik*
*ifo Institute – Leibniz Institute for Economic Research at the*
*University of Munich / Germany*
*piopiunik@ifo.de*

I thank Ludger Woessmann and my colleagues at the ifo Center for the Economics of Education for valuable comments and suggestions. Furthermore, I thank Natalie Obergruber, Lisa Simon, and Moritz Wagner for their contributions and excellent research assistance. Funding by the DFG Priority Programme 1646 "Education as a Lifelong Process. Analyzing Data of the National Educational Panel Study (NEPS)" is gratefully acknowledged.

# 1. Introduction

Countries differ greatly in both the intensity and timing of separating students into different secondary school tracks. Some countries separate students into school tracks with different academic levels as early as age 10, whereas other countries do not track students until the end of compulsory schooling at age 16 (OECD 2004, p. 262).[1] The intensity of tracking – i.e., how early and strictly students with differing ability are separated into different school types – is a potentially important feature of the school system since it might affect both the level and the distribution of students' achievement.[2] After World War II, several European countries have moved from a selective to a more comprehensive school system (Leschinsky and Mayer 1990). In many cases, however, it is rather difficult to isolate the impact of the tracking intensity because these reforms (e.g., in Sweden and Norway) simultaneously changed tracking practices as well as other crucial features of the school system, such as the duration of compulsory schooling.

This study investigates the effects of a reduction in the intensity of tracking on various student outcomes in Germany. To do so, we exploit school reforms that have been implemented in eight out of the 16 German states between 2009 and 2012. These reforms affected students in the lower school tracks, basic school (*Hauptschule*) and middle school (*Realschule*). One common key element of the reforms is that the basic school track has been combined with the middle school track into a new secondary school type.[3] This change made it easier for students at the basic school track to obtain a higher secondary school degree (which provides access to more advanced upper secondary schooling) simply by attending school for one more year. In addition, several reform states established a new school track that offers the possibility to acquire a university entrance qualification, that is, the school-leaving certificate that had traditionally been offered only by the most academic track (*Gymnasium*). The effects of these reforms are estimated in a difference-in-differences approach, comparing student outcomes before the reform to outcomes after the reform as well as between reform states and non-reform states.

[1] While the practice to separate students into different school types that we study is common practice in Europe, students in the United States and Canada are often tracked within schools to different classes based on past achievement (Betts 2011). Card and Giuliano (2016), for example, study the effect of selectively tracking high-achieving students within schools into separate classrooms for gifted students in the US.

[2] Student achievement, in turn, has been shown to affect individual earnings and economic growth (Hanushek and Woessmann 2008).

[3] The terms *school track* and *school type* are used interchangeably throughout the paper.

We use two cohorts of secondary school students from the National Educational Panel Study (NEPS). During school year 2010/2011, students from starting cohort 3 (SC3) attended grade 5, while students from starting cohort 4 (SC4) attended grade 9. Given that the reforms were introduced between school years 2009/2010 and 2012/2013, they affected students in cohort SC3 (*post-reform cohort*), but not students in cohort SC4 (*pre-reform cohort*). Besides measures on student achievement in reading, math, and science, NEPS also contains information on students' and their parents' background. We observe student outcomes in 9th, 10th, and 11th grade. These are relevant grade levels since 9th grade is the last year of compulsory schooling. Therefore, grade 9 is the last grade level when the performance of all students in the general education system can be assessed. Since the middle school degree is acquired after grade 10, the achievement of this degree can be observed in grade 11. Furthermore, the last important school track transition in the general school system occurs after grade 10, when students with a middle school degree may attend a Gymnasium, the most academic track, from grade 11 onward (or alternatively start an apprenticeship).

We find that reducing the intensity of tracking significantly improves students' reading achievement. Effects on students' math and science achievement are also positive, but weaker and statistically insignificant. We do not find robust evidence that the reforms affected the distribution of student achievement, but we find effect heterogeneities along several dimensions: the reforms particularly improved the performance of boys, of students born abroad, and of students with lower socio-economic status (SES), as measured by parental education. Thus, reducing the intensity of tracking seems to have benefited particularly those student groups with lower average performance. In particular, the reading gaps narrowed between boys and girls, between foreign-born and native-born students, and between students with lower and higher SES background. This also applies to the math achievement gaps, with the exception that the math gap between boys and girls widened somewhat.[4] In contrast to the effects on student achievement, we do not find any evidence that the reforms affected the likelihood of acquiring a middle school degree, attending the most academic track (in grade 10 or 11), or repeating a grade, neither overall nor for any student subgroup.

Since students of the post-reform cohort (SC3) joined the NEPS in grade 5, and because NEPS is a voluntary survey, some students left the NEPS between grade 5 and grade 9, when

[4] While boys perform better than girls in math, they are nevertheless often considered a disadvantaged student group since they perform worse in reading, are more likely to repeat a grade, and are more likely to leave school without a degree.

we observe the first outcomes. For this reason, we investigate whether sample attrition differs across reform and non-reform states. This is important since differential attrition could bias our estimates. For example, if attrition rates were higher among better-performing students in the control states, we would overestimate the reform effect on student achievement. However, using various information on students observed in their first NEPS survey (e.g., reading and math achievement), we find no evidence that attrition differs in a systematic way across students from reform states and students from control states. Therefore, attrition is unlikely to bias our results.

Theoretically, the impact of tracking on student achievement is ambiguous (Betts 2011). On the one hand, tracking might increase student achievement because teachers face more homogeneous classrooms in terms of ability, educational achievement, and aspiration (Pekkarinen 2014), which allows them to adjust their teaching style to students' ability level. Furthermore, schools can adjust the curriculum to the students' achievement level or adjust their resources, for example, by hiring teachers with certain qualifications. On the other hand, tracking might lower equality of opportunities since track placement might be affected by a student's socioeconomic status (see Dustmann (2004) for Germany). Tracking might also be detrimental when ability is measured with noise because then some students are likely to be allocated to the wrong track (Brunello et al. 2007). Opponents of tracking also argue that both low-performing and high-performing students benefit from interacting with each other: weak students benefit from the help of strong students, while strong students benefit through explaining the subject material to weak students since this consolidates their knowledge (Lazear 2001).

Given these arguments, tracking might cause an efficiency-equity trade-off in a school system. Empirically, however, there is no evidence for such a trade-off. Most studies find that tracking decreases equity while efficiency is not increased. In a cross-country study, Hanushek and Woessmann (2006) compare the achievement of fourth graders, i.e., prior to tracking, with the achievement of eighth or ninth graders, i.e., after some countries have already tracked their students into different school types. They find that achievement inequality increases in early-tracking countries, with no significant effects on the achievement level.[5] Using cross-country variation in the timing of tracking, Ammermueller (2013), Brunello and Checchi (2007), Schuetz et al. (2008), and Woessmann et al. (2009) similarly find that student achievement depends more strongly on family background in

[5] Exploiting a similar international setting, Ruhose and Schwerdt (2016) find overall no significant effect of tracking on the migrant-native achievement gap.

countries with early tracking.[6] Van Elk et al. (2011) find that early tracking on average reduces the probability to attend university.

Cross-sectional studies, however, are likely plagued by omitted variable bias arising from unobserved differences between countries or between regions within a country, such as different attitudes toward education. Furthermore, the classification of tracking vs. non-tracking countries has been criticized as comparing mainly different kinds of tracking, for example, within- versus between-school tracking (Betts 2011, Waldinger 2007).

These issues are circumvented in within-country studies, which typically evaluate tracking effects in quasi-experimental settings by exploiting variation in the timing and spatial spread of tracking reforms. In the 1950s, Sweden simultaneously replaced the academic and non-academic track with comprehensive schools, increased compulsory years of schooling, and introduced a nationally unified curriculum. Exploiting the successive implementation of the reform across municipalities, Meghir and Palme (2005) find that the reform increased schooling and earnings for students with low socioeconomic background. Aakvik et al. (2010) study a similar school reform in Norway in the 1960s and find that the reform lowered the impact of family background on educational attainment. In the 1970s, Finland replaced a two-track school system with a nine-year comprehensive school, thus postponing tracking into vocational and academic tracks from age 10 to age 15. In line with our results, Pekkala Kerr et al. (2013) find that the reform significantly improved the achievement of students with low SES parents on verbal, arithmetic, and logical reasoning tests. At the same time, the reform had no impact on the achievement of students with high SES parents. Pekkarinen et al. (2009) investigate the long-run effects of the Finnish reform and find that it also increased intergenerational income mobility. Galindo-Rueda and Vignoles (2007) investigate a reform in the United Kingdom that replaced early tracking with comprehensive schools, finding some evidence of positive effects on student achievement in the selective system. However, Pischke and Manning (2006) demonstrate that it is unlikely to eliminate selection bias in case of the U.K. reform. Hall (2012) investigates the effects of introducing a more comprehensive secondary school system in Sweden in the 1990s which prolonged and increased the academic content of the vocational track. While the reform increased the amount of upper secondary schooling of vocational students, it likely had no effect on the probability to enroll in a university or on subsequent earnings. Guyon et al. (2012) investigate an educational reform in Northern Ireland that led to a large increase in

[6] Bauer and Riphahn (2006), using a cross-section of Swiss cantons, find that late tracking reduces the positive effect of highly educated parents on their children's education.

the share of students admitted to the elite track at age 11. They find a strong positive overall effect of this detracking reform on the number of students passing national examinations at later stages and a negative effect on student achievement in non-elite schools who lost their most able students. Malamud and Pop-Eleches (2010, 2011) evaluate a reform in Romania that postponed tracking of students into vocational and academic schools by two years, finding overall no effect on university completion, labor-market participation, or earnings. For disadvantaged students, they find an increased probability to finish the academic track, which does, however, not translate into a higher probability to complete university. In sum, studies based on educational reforms tend to find that later tracking reduces the impact of family background on student achievement.

By international standards, Germany's 16 decentralized school systems are characterized by early tracking and high inequality. Almost all of these school systems track students at age 10 and many offer at least three different secondary school tracks.[7] Exploiting cut-off dates for primary school admission as an instrument, Dustmann et al. (2017) find that attending a higher track has no effect on years of schooling and adult earnings of the marginal student. However, Dustmann (2004) documents strong positive correlations between socioeconomic background and track choice, which turns into earnings differences in adulthood. Piopiunik (2014) evaluates a reform in the state of Bavaria, which shifted tracking from age 12 to age 10 for students in the basic and middle school track. He finds that the reform reduced average achievement and increased achievement inequality in these two tracks and increased the share of very low-performing students in the basic track. Overall, these results suggest that earlier tracking decreased equity of the Bavarian school system. Concerning the permeability of the German three-tier school system, previous studies find that only students from high socio-economic background take the opportunity of educational upgrading (Biewen and Tapalaga 2017; Buchholz and Schier 2015). Exploiting differences in tracking systems across states, Matthewes (2021) finds substantial achievement gains from comprehensive versus tracked schooling at ages 10-12, which are almost entirely driven by low-performing students.

We contribute to the literature of reform evaluation studies that investigate the effects of tracking on the level and equity of student outcomes, exploiting detracking reforms that occurred in eight of the 16 German states. To do so, we use comparable achievement tests (and other outcomes) administered to two large and representative student cohorts that

[7] Overall, cross-sectional studies find no significant performance differences between early-tracked and late-tracked students in Germany (Muehlenweg 2008; Woessmann 2010).

attended school either before or after the reforms. Importantly, student achievement is tested several years after students have been tracked, such that the tracking of students had enough time to unfold its effects on students' achievement. Because students are still in school at the time of testing (and when other outcomes are observed), factors other than the school system are unlikely to explain the estimated effects. In contrast to existing within-country studies, the reforms in the German states have the advantage that other key characteristics of the school systems, such as years of compulsory schooling and financial support for students from low socio-economic background, have not been changed. Additionally, although school systems vary across states, most other policy areas are managed nationally, affecting all states alike.

The paper proceeds as follows. Section 2 describes the school system in Germany and the school reforms in the eight German states. Section 3 describes the NEPS data and provides summary statistics for the pre-reform and post-reform student cohorts. Section 4 presents the empirical strategy, and Section 5 reports the main results, heterogeneity analysis, and attrition results. Section 6 concludes.

## 2. German School System and Detracking Reforms

In Germany, children start school in the year after they turn six years old and typically attend four grades in primary school (*Grundschule*).[8] At about age 10, students are separated into different secondary school types, which differ by both duration and curriculum. The secondary school track decision after primary school is based on teacher recommendations and/or on parents' wishes. At the end of primary school, neither ability tests nor centralized examinations exist that could provide information as to the students' academic potential. Instead, primary school teachers recommend a secondary school track for each student, which mostly depends on the student's grades in the two major subjects German and math (sometimes also science). This school track recommendation is binding in some, but not all states (Grewenig 2021). In states with a binding recommendation, school authorities define a cutoff for the average grade in German and math (and science) that is required to receive a recommendation for a certain school track.

Traditionally, West German states had three different school types: (i) basic schools (*Hauptschule*) provide basic general education and typically lead to a certificate after grade 9 (in few states after grade 10); (ii) middle schools (*Realschule*) provide a more

[8] In two of the 16 states, Berlin and Brandenburg, primary school lasts six years. Because authority and control over education policy lies with each state (*Bundesland*), the school structure differs somewhat across states. See Kultusministerkonferenz (2019) for a detailed description of the German school system.

extensive general education and last six years; and (iii) the most academic track, *Gymnasium*, covers eight or nine grades and is the only school type that exists in all German states. In East German states, integrated schools (called *Mittelschule* or *Regelschule*) exist instead of basic and middle schools, offering the school-leaving certificates typically obtained in these two school types. In some states, comprehensive schools (*Gesamtschule*) exist in addition to the other school types. This school type includes both lower and upper secondary education level and is typically attended only by a small fraction of students.

Importantly, even though the number (and names) of secondary school types differs across states, each state offers the same (three) school-leaving certificates, namely those traditionally acquired when graduating from basic school, middle school, and Gymnasium, respectively. We therefore refer to these three certificates as basic school degree, middle school degree, and university entrance qualification (*Hochschulzugangsberechtigung*, acquired at the Gymnasium) – independent of the name of the school type where a student acquires her certificate and independent of the state-specific name for that certificate (which may differ across states).

Students with a basic school degree typically enter an apprenticeship that combines part-time vocational school and firm-based training. Students with a middle school degree might do the same types of apprenticeship, but are also entitled to attend (full-time) vocational schools that lead to a higher education entrance qualification. Specifically, students can acquire an advanced qualification (*Fachhochschulreife*) at a vocational school that qualifies for attending a university of applied sciences (*Fachhochschule*). When graduating from Gymnasium, students obtain the university entrance qualification, which is a prerequisite for attending a university.

*Detracking Reforms in German States*

Between 2009 and 2012, eight out of the 16 German states have implemented educational reforms that decreased the intensity of secondary school tracking (referred to as "detracking reforms" in this paper). The reforms aimed at increasing equality of opportunities and individual advancement in the respective school systems. In all reform states, students with different academic abilities and different family backgrounds tend to attend secondary school together longer after the reforms. In some states, secondary school systems got more flexible as specific school tracks offered more than one school-leaving certificate (depending on the grade level successfully completed), especially at school types attended by low-performing students. In some states, this has been achieved by combining two existing school types into a single, new school type, which offers more

than one school-leaving certificate. In other states, the education reforms established a new, additional school type that offers two or even all three school-leaving certificates. In these new school types, students with different family backgrounds learn together in the same courses until they graduate. In the following, we describe the key elements of the eight detracking reforms (states ordered alphabetically). Note that the Gymnasium track exists in all states, both before and after the reforms.

*Baden-Wuerttemberg*. At the start of school year 2010/2011, a new school type has been established (called *Werkrealschule*), which complements the existing three tracks basic school, middle school, and Gymnasium. This new school type covers the grade levels 5 to 10. Besides offering the basic school degree after grade 9, it also enables students who obtained only a basic track recommendation from primary school to obtain a middle school degree after grade 10. In 2012/2013, comprehensive schools (*Gemeinschaftsschule*) have been established, with compulsory all-day schooling in grades 5 to 10. Courses are offered at three different academic levels, reflecting the curriculum standards of basic school, middle school, and Gymnasium. The basic level leads to the basic school degree after grade 9 (or 10) and the intermediate level to the middle school degree after grade 10. In comprehensive schools that include the upper secondary education level (grades 11 to 13), the university entrance qualification can be obtained after grade 13.

*Bavaria.* In 2011/2012, basic schools have been converted into "middle schools" (*Mittelschule*).[9] Like in basic schools, students can obtain the basic school degree after grade 9. In addition, the reform enabled students in this track also to acquire the middle school degree after grade 10. To do so, students have to attend the middle-school-degree path (*Mittlerer Reife-Zug*) from grade 7 onward.

*Berlin*. In 2010/2011, three school types – basic school, middle school, and comprehensive school – have been merged into a new type, the so-called integrated secondary school (*integrierte Sekundarschule*). The new school type follows the concept of "one school for all", that is, it aims at supporting both low- and high-performing students. Grades 7 to 10 are all-day schools. Integrated secondary schools offer all types of school-leaving certificates. In particular, they provide the opportunity to obtain the university entrance qualification. This is achieved by offering either their own upper secondary level (grades 11 to 13) or by cooperating with a vocational Gymnasium (*berufliches Gymnasium*) or by forming a cluster with another integrated secondary school. While the university entrance

[9] Although the English name is similar, this school type differs from the intermediate school track, called *Realschule* in Bavaria.

qualification is obtained after grade 13 at integrated secondary schools, it is acquired after grade 12 at the Gymnasium.[10]

*Bremen*. Starting in 2009/2010, all secondary school tracks containing the lower secondary education level (grades 5-10) except independent Gymnasium schools – that is, comprehensive schools and school centers (*Schulzentrum*) containing secondary schools and Gymnasium – have been gradually converted into so-called upper schools (*Oberschule)*. The restructuring had been completed by 2011/2012. All school-leaving certificates can be obtained at this new school type: the basic school degree after grade 9; the extended basic school degree (*erweiterte Berufsbildungsreife*) and middle school degree after grade 10; and the university entrance qualification after grade 13. While the Gymnasium offers the same degrees, the university entrance qualification can be acquired already after grade 12.

*Hamburg*. In 2010/2011, basic schools and middle schools were replaced by so-called district schools (*Stadtteilschulen*). At both district schools and Gymnasium, the basic school degree can be obtained after grade 9 and the middle school degree after grade 10. While the university entrance qualification can be obtained after grade 13 at district schools, it can be obtained at Gymnasium already after grade 12. Importantly, all students in district schools learn together, with academically weaker students being supported and challenged in the same way as academically stronger students.[11]

*Rhineland-Palatinate.* Starting in 2009/2010, all basic schools and middle schools have been combined into "middle schools plus" (*Realschule plus*). However, the conversion process was finished only in 2013/2014, i.e., two (three) years after the pre-reform cohort (SC4) graduated from middle (basic) school.[12] Middle school plus offers the two school-leaving certificates that students previously acquired at basic and middle schools: the basic school degree (called *Berufsreife*) after grade 9 and the middle school degree (*Sekundarabschluss I*)

[10] A few years earlier, in 2008/2009, community schools (*Gemeinschaftsschulen*) were set up as a pilot project. All community schools offer all school-leaving certificates, either by providing their own upper secondary level or by cooperating with other schools with an upper secondary level. The introduction of the detracking reform in Berlin has been academically supervised by Neumann et al. (2017). They find that the reform increased the share of students in upper secondary education, but without improving subject knowledge. Inequalities with respect to socio-economic background did not change either.

[11] To cope with the greater diversity of the student body, classes in district schools are smaller than classes in Gymnasium. To achieve smaller classrooms, additional teachers have been hired. Furthermore, teacher training has been strengthened and children with disabilities have been integrated in the new school type.

[12] In contrast to the reforms in the other seven states, it is therefore possible that some students of the pre-reform cohort have been affected by the reform. As a robustness check, we have excluded the small state Rhineland-Palatinate from the sample; all results are very similar (results available upon request).

after grade 10.[13] Furthermore, integrated comprehensive schools (*Integrierte Gesamtschule*), which cover grades 5-13 and offer all three school-leaving certificates, existed already since the 1970s. However, since 2009 many new integrated comprehensive schools have been constructed.

*Saarland.* In 2012/2013, community schools (*Gemeinschaftsschulen*) have replaced basic schools and middle schools. In addition to offering the basic school degree after grade 9 and middle school degree after grade 10, community schools also contain the upper secondary education level, which leads to the university entrance qualification after grade 13. Importantly, at community schools educational career paths remain open as long as possible, thus ensuring permeability between educational pathways.

*Schleswig-Holstein.* In 2011, the three-tier school system became a two-tier system when basic schools and middle schools have been converted into regional schools (*Regionalschulen*). Regional schools offer the basic school degree after grade 9 and the middle school degree after grade 10. In 2014, these new regional schools have been converted into community schools (*Gemeinschaftsschulen*). While regional schools only offered basic and middle school degrees, the conversion into community schools enabled students to acquire also the university entrance qualification after grade 13, at least in those community schools that also have an upper secondary education level (about 23% of schools). In contrast, most Gymnasium schools offer this qualification after grade 12 (the remaining Gymnasium schools after grade 13).

These educational reforms vary with respect to whether the newly established school type offers a university entrance qualification (*Hochschulzugangsberechtigung*) or not. In Baden-Wuerttemberg, Bavaria, and Rhineland-Palatinate, the reforms did not affect the university entrance qualification, which can be obtained only at a Gymnasium in these three states.[14] In contrast, the new school types established in Berlin, Bremen, Hamburg, Schleswig-Holstein, and Saarland provide students the opportunity to obtain a university entrance qualification, rendering these new schools potential competitors for the Gymnasium.

[13] *Realschule plus* exists in two versions. In *cooperative* schools, students are taught separately in "degree-oriented" classes after the common orientation stage (grades 5 and 6). Each school year teachers check whether students can transfer to the course track that leads to the higher degree, i.e., the middle school degree. In *integrative* schools, students continue to learn together in the same classroom after the orientation stage, regardless of the aspired education degree.

[14] However, in Baden-Wuerttemberg the university entrance qualification can also be acquired at the newly established comprehensive schools, but only at those schools that include the upper secondary education level (grades 11 to 13).

Finally, the non-reform states either have a two-tier school system (states in East Germany) or have a school system with three or more tracks (in West Germany).

# 3. Data

This section describes the National Educational Panel Study (NEPS), in particular the two student cohorts used in this study, and presents summary statistics separately by cohort.

## 3.1 National Educational Panel Study

We use data from the National Educational Panel Study (NEPS), which analyzes educational processes in Germany from early childhood to late adulthood. To do so, NEPS follows various age cohorts – called starting cohorts (SC) – over time. To investigate the effects of the detracking reforms, we use two cohorts: starting cohort 3 (SC3) and starting cohort 4 (SC4). SC3 includes students enrolled in 5th grade during school year 2010/2011, and SC4 includes students enrolled in 9th grade during the same school year.[15] The first surveys were conducted in the fall of 2010, that is, at the beginning of the school year 2010/2011. Besides assessing students' achievement in reading, math, and science, NEPS also elicits information via student and parent background questionnaires (Blossfeld et al. 2011).[16]

NEPS samples students in a two-step procedure. First, a representative set of schools is drawn. Importantly, schools are sampled after stratifying schools by type and region, which ensures that both SC3 and SC4 samples include students from all school types and all 16 states. Second, within selected schools, entire classrooms are randomly drawn. If there is more than one 5th (9th)-grade classroom, then two classrooms are (randomly) sampled. In principle, students can choose not to participate in NEPS. However, since participating schools support the survey, most students do participate.[17]

---

[15] This study uses data from the National Educational Panel Study (NEPS) Starting Cohort Grade 5 (SC3) (doi:10.5157/NEPS:SC3:3.1.0) and Starting Cohort Grade 9 (SC4) (doi:10.5157/NEPS:SC4:9.0.0) (Blossfeld et al. 2011). From 2008 to 2013, NEPS data were collected as part of the Framework Program for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research (BMBF). As of 2014, NEPS is carried out by the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg in cooperation with a nationwide network. Data are available from https://www.neps-data.de/en-us/home.aspx.

[16] NEPS also surveyed school principals and teachers. Unfortunately, relevant information is not consistently available for the same grade level for our two cohorts (SC3 and SC4). For example, class size is reported in 6th grade for SC3, but only in 9th grade for SC4. Similarly, while information on teachers (e.g., gender, year of birth, and school grades) has been elicited in teacher questionnaires, this information is available for teachers in different grade levels. Hence, teacher information is missing for many students in 9th/10th grade.

[17] On average, in each sampled classroom about 70 percent of students participate. School-aged children need consent from their parents to participate in the survey.

Students from starting cohort 4 (SC4) attended grades 5 to 9 during the school years 2006/2007 to 2010/2011, but are observed for the first time in 9th grade. Students from starting cohort 3 (SC3) attended 5th grade in 2010/2011, which is when they are observed for the first time. The first detracking reforms started in 2009/2010 and the last reforms took place in 2012/2013. Hence, the reforms affected students from cohort SC3, but did not affect students from cohort SC4. For our analysis, we combine student outcomes in 9th/10th/11th grade from cohort SC3 with exactly the same outcomes in the same grade levels for students from cohort SC4. Thus, we observe the outcomes of SC4 students during the years 2010-2013 and the same outcomes of SC3 students (who are four years younger) during the years 2014-2017.

From the sample, we exclude students in general education schools who did not participate in the relevant NEPS surveys when our outcomes of interest are observed, that is, in the surveys in grade 9, 10, and/or 11.[18] Furthermore, we exclude students in special education schools and a few students with missing information on age (13 students) and gender (107 students). The final samples consist of (at most) 6,606 ninth-graders from SC3 and (at most) 15,169 ninth-graders from SC4. As some outcomes are missing for these students, sample sizes differ somewhat across outcomes. Appendix Table A1 reports the (maximal) sample sizes by state and starting cohort.

## 3.2 Summary Statistics

Table 1 reports summary statistics separately for the pre-reform (SC4) and post-reform (SC3) cohort.[19] Student achievement in reading, math, and science has been assessed in grade 9, both for SC3 and SC4 students. We have standardized test scores to have mean 0 and standard deviation 1 separately by cohort and test domain. While 79% of SC3 students obtain a *middle school degree* (which includes also all students attending 11th grade at Gymnasium), this is true for only 68% of SC4 students. This difference arises mainly because 15.7% of SC3 students have left the NEPS survey between 5th and 9th grade, with lower-performing students being more likely to leave NEPS.[20] (We investigate attrition in Section 5.4 and find no evidence that attrition patterns differ systematically between reform and non-reform states.) Another outcome of interest is whether a student attends

[18] Of course, we use all available information. For example, if a student participates in grades 9 and 10, we use this information, even when the student did not participate in the survey in grade 11.

[19] Appendix Table A2 reports all summary statistics for the two cohorts combined.

[20] Because of the attrition of SC3 students, NEPS added new students to starting cohort 3 (SC3) at the start of grade 7. Out of the 2,205 students added, 250 (11.3%) left the NEPS between grade 7 and grade 9.

Gymnasium, the most academic secondary school track. We observe whether students attend a Gymnasium in grade 10 and in grade 11.[21] Consistent with the strong difference in obtaining a middle school degree and the underlying attrition pattern, SC3 students are more likely than SC4 students to attend a Gymnasium (55% versus 35% in 11th grade). Another outcome of interest is whether students have repeated a grade. Unfortunately, this information is missing for most SC3 and SC4 students. As a proxy for grade repetition, we use students' age at the end of grade 9 (in June). Students in cohort SC4 are about 3 months older than students in SC3. This is consistent with the fact that students who left the NEPS survey perform substantially worse than students who have not left the survey since lower-performing students also tend to be older, either because they repeated a grade or because they started primary school one year later.

In all regressions, we control for student and family background characteristics. In both cohorts, gender is balanced. 4% of SC3 students and 6% of SC4 students were born abroad.[22] Number of books at home is a frequently-used measure of the family's socioeconomic status, reported in six different categories (ranging from "0-10 books" to "more than 500 books"). Consistent with higher attrition rates among lower-performing SC3 students, SC3 students on average possess more books at home than students from cohort SC4. Similarly, highest parental education, which corresponds to mother's and father's highest education level, is on average also somewhat higher among SC3 students.[23] A similar difference across cohorts also applies to fathers' occupation, which is measured in 12 categories, ranging from 1 (agricultural professions) to 12 (managers).[24] Since occupation categories are not strictly ordinally ordered, we include a dummy for each category in the regressions.

[21] Attending a Gymnasium also contains a small group (ca. 1%) of students who attend the track in a comprehensive school (*Gymnasialzweig*) that leads to the university entrance qualification.

[22] We use students' rather than parents' country of birth since the latter information is missing substantially more often (for more than 50% of students).

[23] Highest parental education includes nine ISCED-97 categories, ranging from "secondary school certificate" to "doctorate, habilitation". Note that one intermediate category (ISCED-97 level 4B) is not present in our sample.

[24] Some background variables have missing values. Since we consider multiple explanatory variables and since some of these variables are missing for a substantial fraction of students, dropping all student observations with any missing value would result in a substantial sample reduction. We therefore imputed missing values for the control variables as follows: a separate category for father's occupation (since categories have no strict ordinal ordering); separately imputing mother's and father' education with a logit regression, using number of books and students' migration background as explanatory variables; and median value imputation of number of books at home and foreign born. To ensure that imputed data are not driving our results, all regressions include an indicator for each imputed variable that equals 1 for imputed values and 0 otherwise.

# 4. Empirical Strategy

To estimate the effects of the detracking reforms, we apply a difference-in-differences approach. In particular, we exploit the fact that these reforms reduced the intensity of tracking in eight out of 16 of states between 2009/2010 and 2012/2013. As the reforms affected the SC3 cohort, these students constitute the *post-reform* cohort. In contrast, the reforms did not affect students from SC4 who thus constitute the *pre-reform* cohort. The reform effect is identified by comparing the outcomes between the pre-reform and post-reform cohorts across reform and non-reform states. As noted above, outcomes of students in the pre-reform and post-reform cohort are observed in the same grade levels (grade 9, 10, or 11). We estimate the reform effects as follows:

$$y_{ics} = \beta_0 + \beta_1 reform\ state_{cs} + \beta_2 post_c + \beta_3 reform\ state_{cs} * post_c + X_{ics}\beta_4 + \varepsilon_{ics}, \quad (1)$$

where $y_{ics}$ is an outcome of student *i* in cohort *c* (pre-reform or post-reform) who attends school in state *s.* Outcomes include students' achievement in reading, math, and science (assessed in 9th grade), a dummy for whether the student obtained a middle school degree (observed in 11th grade), two dummies for whether the student attended a Gymnasium in 10th and 11th grade, respectively, and age at end of grade 9 (in months). The binary indicator $reform\ state_{cs}$ equals 1 for the eight states that implemented a detracking reform and 0 for all non-reform states. The binary indicator $post_c$ equals 1 for all SC3 students (affected by the reforms) and 0 for all SC4 students (not affected). We are interested in $\beta_3$, the coefficient on the interaction between $reform\ state_{cs}$ and $post_c$, which yields the average effect of the detracking reforms in the eight German states on outcome $y_{ics}$. $X_{ics}$ is a vector of student and family background characteristics of student *i*, such as gender, migration background, number of books at home, highest parental education, and father's occupation. $\varepsilon_{ics}$ is the idiosyncratic error term.

Standard errors are clustered at the state level to reflect both the level of the school reforms as well as possible serial correlation of outcomes within states over time (Bertrand et al. 2004). To account for the small number of 16 independent clusters (states), we compute bootstrapped standard errors (clustered at the state level) that provide an asymptotic refinement with a limited number of clusters.

To investigate effect heterogeneities, we split the sample separately by four background characteristics: gender, migration background, highest parental educational, and number of books at home. To assess whether reform effects differ statistically significantly, we conduct t-tests on the $\beta_3$ coefficients from the two subsamples (e.g., females and males).

The outcomes in Equation (1) are measured at the individual student level. However, existing studies suggest that tracking effects might be greater on educational inequality (Hanushek and Woessmann 2006). To investigate effects on educational inequality, we compute various inequality measures of students' reading, math, and science achievement at the state-cohort level, such as the standard deviation and percentile gaps (e.g., 95th-5th percentile gap). Thus, we run the same type of regression at the state-cohort level:

$$y_{cs} = \beta_0 + \beta_1 reform\ \text{state}_{cs} + \beta_2 \text{post}_c + \beta_3 reform\ \text{state}_{cs} * \text{post}_c + X_{cs}\beta_4 + \varepsilon_{cs}, \quad (2)$$

where $y_{cs}$ is an inequality measure of student achievement (e.g., standard deviation) in reading, math, and science, respectively, of cohort *c* (pre-reform or post-reform) in state *s.* $X_{cs}$ is a vector of student and family background characteristics aggregated at the state-cohort level, such as the share of male students, share of students born abroad, and four different shares of parents' highest education level. $\varepsilon_{cs}$ is the idiosyncratic error term at the state-cohort level. Again, standard errors are clustered at the state level.

# 5. Results

This section first presents the results of the impact of the detracking reforms on the level and inequality of student achievement (Section 5.1) and then on middle school degree acquisition, Gymnasium attendance, and grade repetition (Section 5.2). Afterwards, Section 5.3 explores the heterogeneity of these effects by gender and socio-economic background. Finally, Section 5.4 explores whether attrition patterns of students in the post-reform cohort (SC3) differ systematically across reform and non-reform states.

## 5.1 Reform Effects on Student Achievement

We first investigate the effects of the detracking reforms on the level of students' reading, math, and science achievement.[25] One advantage of the NEPS data is that student achievement has been assessed in grade 9, that is, the last year of compulsory schooling. Therefore, we observe student achievement when all students are still attending the general school system.[26]

The detracking reforms improved students' reading achievement by about 0.18 standard deviations (SD) when not accounting for student and family background characteristics

[25] Berlin is excluded from all student achievement analyses because reading, math, and science scores are not available for the post-reform cohort (SC3).

[26] Students who acquire a basic school degree (typically at the end of grade 9) tend to begin an apprenticeship afterwards, thus leaving the general school system after grade 9.

(Column 1 of Table 2). When including these characteristics, the effect decreases slightly to 0.15 SD (Column 2). On average, students in reform states perform as well as students from non-reform states. Although test scores are standardized by cohort, students from the post-reform cohort (SC3) perform on average worse (-0.17 SD) than students from the pre-reform cohort (SC4) when student-level covariates are accounted for. This difference arises due to the survey attrition of SC3 students between 5th grade (when they enter NEPS) and 9th grade (when student achievement is assessed).[27] In particular, students with higher SES background had lower attrition rates and are therefore somewhat overrepresented in the SC3 sample compared to the SC4 sample (see Table 1). Hence, when we control for these background characteristics, the coefficient on the *post* dummy turns significantly negative. As expected, boys perform substantially worse than girls in reading (-0.21 SD). Similarly, students born abroad have lower German reading skills than students born in Germany (-0.23 SD). As expected, we also find that students with more books at home and students with higher-educated parents have higher reading skills (results not shown).

We also find that the detracking reforms increased students' math achievement by 0.14 SD (Column 3). However, when adding the student-level covariates, the effect decreases to 0.09 SD and becomes statistically significant only at the 14% level (Column 4). Similar to the reading model, the coefficient on the post-cohort dummy turns significantly negative (-0.15 SD) once we add the student-level covariates. Male students perform better than girls in math (0.31 SD), and students born abroad perform lower than native students (-0.18 SD).

The results pattern for students' science achievement is very similar to the math results, except that the reform effects seem to be somewhat weaker (0.12 SD without and 0.07 SD with covariates; Columns 5 and 6). Compared to math, the gap in science achievement between boys and girls is only about half as large (0.18 SD), while it is even larger between foreign-born and German-born students (0.26 SD).

The detracking reforms might not only affect students' achievement level, but may also affect the inequality of student achievement (e.g., Hanushek and Woessmann 2006). To investigate this, we use various inequality measures of student achievement, such as the standard deviation and several percentile gaps, which measure the dispersion across the overall achievement distribution (p95–p5, p90–p10, p75–p25), at the top of the distribution (p90–p50), and at the bottom (p50–p10). All inequality measures are computed at the state-

[27] We investigate survey attrition in detail in Section 5.4, showing that attrition does not systematically differ across reform and non-reform-states. This finding suggests that the estimated reform effects are not biased by (systematic) attrition of SC3 students.

cohort level, separately for reading, math, and science achievement.[28] All regressions control for the share of male students, the share of foreign-born students, and the share of parents with a specific education level (grouped into four categories), computed also at the state-cohort level. Without these state-cohort-level covariates, results are very similar.

We find only weak evidence that the reforms reduced the inequality of students' reading achievement (Panel A of Table 3). While the coefficients on all inequality measures are negative, none of them is statistically significant at the 10% level. The coefficients on math achievement inequality tend to be closer to zero and all of them are statistically insignificant (Panel B). Regarding effects in science, it seems that achievement inequality has decreased at the top, while it has widened at the bottom of the achievement distribution (Columns 5 and 6 in Panel C).

Overall, we find that the detracking reforms substantially improved students' reading achievement. The effects on math and science achievement are also positive, but weaker and statistically insignificant. Overall, we do not find strong evidence that the reforms changed the inequality of student achievement in reading, math, or science.

## 5.2 Reform Effects on Middle School Degree Acquisition, Gymnasium Attendance, and Grade Repetition

After investigating reform effects on students' performance in three important domains, we now turn to potential effects on the probability of acquiring a specific school degree. Since the reforms mainly targeted lower-performing students who traditionally attended basic schools and acquired a basic school degree (the least academic ("lowest") school-leaving certificate), we first investigate whether the reforms increased the likelihood to acquire a middle school degree, that is, a higher degree. Such an effect is possible since the reforms either enabled basic schools to additionally offer a middle school degree (to better-performing students) or because the basic school track has been merged with the middle school track, thus offering both school-leaving certificates. Note that students who successfully complete grade 10 at a Gymnasium (or comprehensive school) automatically obtain a middle school degree as well.

We do not find that the reforms affected the probability of acquiring a middle school degree (Columns 1 and 2 of Table 4). Consistent with the attrition pattern (see Section 5.4), we find

[28] Again, Berlin is excluded because reading, math, and science achievement is not available for the post-reform cohort (SC3).

that students from the post-reform cohort (SC3) are about 11 percentage points more likely to obtain a middle school degree than students from the pre-reform cohort (SC4). Compared to an overall likelihood of acquiring a middle school degree of 71% in our sample (Appendix Table A2), this is a substantial difference. While this indicates that attrition rates (in the SC3 cohort) are higher among lower-performing students, it does not bias our reform effects (see Section 5.4).

To avoid the reform, children who would have attended a middle school before the reform, might attend a Gymnasium, the most academic school track, after the reform. Such a change in track choice might be a potential channel of the reform effects on student achievement.[29] A change in track choice might inhibit potentially positive peer group effects for students in the lower tracks, but might have negative peer effects for students in the highest track. We investigate effects on the probability of attending a Gymnasium in 10th grade, that is, the last grade level of middle schools and integrated schools. Furthermore, we also look at Gymnasium attendance in 11th grade, that is, the grade level after students have acquired the middle school degree.[30] We find only very small positive effects on attending Gymnasium in grade 10 and 11 (Columns 3 to 6). The coefficients indicate an increase by one or two percentage points, but are statistically insignificant. Boys and foreign-born students are less likely to attend a Gymnasium (by 4 to 7 percentage points).

Finally, we investigate whether the detracking reforms had any impact on grade repetition. At noted above, information on grade repetition is missing for many students in our sample. Therefore, we use students' age in months at the end of 9th grade (in June) as a proxy. Students' age is a good proxy for grade repetition since the detracking reforms did not change the age cutoffs for entering primary school. We find a small negative effect, suggesting that students affected by the reforms are about 0.3 months younger at the end of grade 9; however, these coefficients are imprecisely estimated (Columns 7 and 8). Thus, we find only little evidence that the reforms decreased the incidence of grade repetition.

Overall, we find no impact of the detracking reforms on the probability of acquiring a middle school degree, attending the most academic track (Gymnasium), or repeating a grade.

[29] Some evidence that such a change in track choice might have taken place is provided for the small state of Berlin (Neumann et al. 2017).

[30] We also investigated the effect on attending a Gymnasium in 9th grade. The result is very similar to that of attending Gymnasium in 10th grade.

## 5.3 Heterogeneity by Gender and Socio-Economic Background

In this section, we investigate whether reform effects differ by gender, migration background, and family background, as measured by parents' highest education level and number of books at home. To do so, we repeat the previous analyses for each subsample separately. All regressions control for gender, migration background, highest parental education, number of books at home, and father's occupational status, except for the covariate that is used for splitting the sample.

We have estimated that the detracking reforms increased students' reading achievement by about 0.15 SD on average (Table 2). Splitting the sample by gender, we find that this effect is more pronounced for boys (0.21 SD) than for girls (0.10 SD) (Column 1 in Panel A of Table 5). While the effect for boys is statistically significant at the 1% level, it is not statistically significant at conventional levels for girls. Due to the large standard errors, the difference between the effect on boys and that on girls is statistically significant only at the 13% level.[31] Furthermore, the effect on reading achievement is stronger for foreign-born students than students born in Germany (0.23 SD vs. 0.15 SD; Panel B). Concerning family background, we find that reform effects are stronger for students whose parents have lower education levels (0.22 SD vs. 0.10 SD; Panel C) and students with fewer books at home (0.21 SD vs. 0.14 SD; Panel D).[32] The difference between the parental education subsamples is statistically significant at the 5% level. In sum, the detracking reforms seem to have improved particularly the reading achievement of the lower-achievement groups: boys, students born abroad, and student with lower SES background.

The pattern of reform effects on students' math achievement is very similar (Column 2): boys (who perform better in math than girls), students born abroad, and students with less-educated parents benefit from substantial math achievement gains. In contrast to reading, the effects seem to be very similar for students with few books and students with many books at home. Since the overall effects are substantially weaker on math than on reading achievement (0.09 vs. 0.15 SD), the effects by subgroups are also weaker. For example, the positive effect for boys is 0.14 SD (vs. 0.21 SD in reading) and for students born abroad 0.20 SD (vs. 0.23 SD in reading). The difference across the parental education subsamples is statistically significant at the 10% level. Results for the gender and migration background

[31] All subsample differences in Table 5 are statistically insignificant at the 10% level except those differences explicitly mentioned in the text.

[32] *High parental education* contains all students whose mother or father have at least a post-secondary, non-tertiary degree (ISCED-97 code 4A and higher).

subsamples for science achievement are similar to those of math, but there are no differences with respect to parental education and effects are stronger (rather than weaker) for students with many books at home (significant at the 10% level) (Column 3).

Finally, we investigate effect heterogeneities with respect to the probability of acquiring a middle school degree, attending the most academic track (Gymnasium), or repeating a grade. Recall that we did not detect any significant effects on these outcomes in the full student sample. Similarly, we find no evidence that the detracking reforms affected these outcomes for any student subgroup we investigate (Table 6). None of the reform*post-reform coefficients is statistically significant at conventional levels. Furthermore, we do not find any significant effect differences across student subsamples.

Overall, we find that the detracking reforms have substantially improved the reading and math achievement of boys, students born abroad, and students with lower SES background. This implies that the reading gap has decreased between foreign-born and native-born students as well as between students with lower SES background and students with higher SES background. Concerning the achievement gaps between boys and girls, the results point in different directions: while the reading gap has narrowed, the math gap has widened. In contrast to substantial heterogeneous effects on student achievement, we do not find evidence that the reform effects on acquiring a middle school degree, attending the most academic track (Gymnasium), or repeating a grade differ across subgroups. This suggests that the detracking reforms had no effects on these outcomes.

## 5.4 Attrition Analysis

The post-reform cohort (SC3) joined NEPS at the start of 5th grade, that is, four years before we observe the first outcomes for these students.[33] However, because NEPS is a voluntary survey, students are allowed to leave the survey at any time. Since some students indeed left the survey, the NEPS team added so-called "top-up" students to starting cohort 3 (SC3) at the start of grade 7 to make up for the students who left. This implies that four years passed between the first NEPS survey of the *initial students* and the time we observe the first outcomes (in grade 9); similarly, two years passed for the *top-up students* (between grade 7 and grade 9). We therefore investigate the attrition patterns for both initial students and top-up students. In particular, we assess whether attrition differs systematically across reform and non-reform states since differential attrition could bias our estimates. For

[33] In contrast, the pre-reform cohort (SC4) joined NEPS in 9th grade. Hence, we observe the first outcomes for these students immediately after joining the survey.

example, if attrition rates were higher among high-performing students in non-reform states, we would overestimate the reform effects on student achievement.

Table 7 presents results from the attrition analysis separately for initial students (Panel A) and top-up students (Panel B). Here, attrition means that a student left the NEPS survey before 9th grade, that is, before we observe the first outcomes. We investigate whether attrition occurred differentially across reform and non-reform states with respect to four student outcomes/characteristics. The information on these four variables – reading and math achievement, number of books at home, and father's education – stem from the first school year when students joined NEPS, that is, from grade 5 for initial students and from grade 7 for top-up students.

We find that students who left the NEPS before grade 9 (coefficient on a*ttritioned*) have substantially lower reading and math skills than students who do not leave the NEPS. This difference amounts to 35%-45% of a standard deviation and is highly statistically significant (Columns 1 and 2). Furthermore, students who left the NEPS are more likely to come from families with lower socioeconomic background, as measured by the number of books at home and father's education (Columns 3 and 4). These findings hold for both initial students and top-up students. Furthermore, the results suggest that students in reform states perform somewhat better and tend to come from higher SES families than students in non-reform states; however, only one of the eight coefficients are statistically significant at the 10% level.

Most importantly, we find no evidence that attrition patterns with respect to these student characteristics differ significantly across reform and non-reform states. The respective coefficients for are very small, both for initial students and top-up students. Therefore, differential attrition is unlikely to bias the estimated reform effects.

# 6. Conclusion

We investigate the effects of a reduction in the intensity of tracking on student outcomes by exploiting education reforms in eight of the 16 German states. These reforms affected students in the lower school tracks, typically extending the periods of learning together and providing differentiated learning opportunities to promote the individual abilities of all students. We estimate the effects in a difference-in-differences approach, using a pre-reform and a post-reform student cohort from the National Educational Panel Study from reform and non-reform states, which contain student outcomes in grades 9 to 11.

The detracking reforms significantly improved students' average reading achievement. The reading improvement is stronger for boys, students born abroad, and students with lower socio-economic background. In addition, the reforms also increased the math achievement for these groups. Therefore, the detracking reforms seem to have benefited particularly lower-performing students. Despite the positive effects on student achievement (particularly in reading), the reforms had no significant impact on the probability of acquiring a middle school degree or on attending the most academic track (*Gymnasium*), neither overall nor for any student subgroup investigated. Hence, the reforms seem to have failed one goal – to make it easier for students at the basic school track level to upgrade their education by an additional year of schooling and obtaining a higher secondary school degree. Finally, we do not find any evidence that the reforms affected the likelihood of repeating a grade.

When trying to isolate the impact of the intensity of tracking on student outcomes, one advantage of the German detracking reforms – in contrast to existing reform evaluation studies – is that other key characteristics of the school systems, such as the timing of tracking, the years of compulsory schooling, and financial support for students from low socio-economic background, have not been changed. This makes the detracking reforms in the German states a particularly valuable setting to investigate the effects of intensity of tracking on student outcomes.

# References


Aakvik, A., Salvanes, K. G., & Vaage, K. (2010). Measuring heterogeneity in the returns to education using an education reform. *European Economic Review*, 54, 483–500.

Ammermueller, A. (2013). Institutional Features of Schooling Systems and Educational Inequality: Cross-Country Evidence From PIRLS and PISA. *German Economic Review* 14 (2), 190-213.

Argys, Laura M., Daniel I. Rees, Dominic J. Brewer (1996). Detracking America's Schools: Equity at Zero Cost? *Journal of Policy Analysis and Management* 15 (4): 623-645.

Bauer, Philipp, Regina T. Riphahn (2006). Timing of School Tracking as a Determinant of Intergenerational Transmission of Education. *Economics Letters* 91 (1): 90-97.

Bauer, Philipp C., Regina T. Riphahn (2013). Institutional Determinants of Intergenerational Education Transmission: Comparing Alternative Mechanisms for Natives and Immigrants. *Labour Economics* 25: 110-122.

Bertrand, Marianne, Esther Duflo, Sendhil Mullainathan (2004). How Much Should We Trust Differences-in-Differences Estimates? *Quarterly Journal of Economics* 119 (1): 249-275.

Betts, Julian R. (2011). The Economics of Tracking in Education. In *Handbook of the Economics of Education*, Vol. 3, edited by Eric A. Hanushek, Stephen Machin, Ludger Woessmann. Amsterdam: North Holland: 341-381.

Betts, Julian R., Jamie L. Shkolnik (2000). The Effects of Ability Grouping on Student Achievement and Resource Allocation in Secondary Schools. *Economics of Education Review* 19 (1): 1-15.

Biewen, Martin, Madalina Tapalaga (2017). Life-cycle Educational Choices in a System with Early Tracking and 'Second Chance' Options. *Economics of Education Review* 56: 80-94.

Blossfeld H-P, Roßbach H-G, Maurice Jv, (2011): editors. *Education as a lifelong process. The German National Educational Panel Study (NEPS)*. Wiesbaden: Springer; 2011.

Brunello, Giorgio, Daniele Checchi (2007). Does School Tracking Affect Equality of Opportunity? New International Evidence. *Economic Policy* 22 (52): 781-861.

Brunello, G., Giannini, M., & Ariga, K. (2007). The optimal timing of school tracking: A general model with calibration for Germany. In L. Woessmann & P. Peterson (Eds.), *Schools and the equal opportunity problem* (pp. 129–156). Cambridge, Mass: The MIT Press

Buchholz, S., A. Schier (2015). New Game, New Chance? Social Inequalities and Upgrading Secondary School Qualifications in West Germany. *European Sociological Review* 31, 603-615.

Card, David, Laura Giuliano (2016). Can Tracking Raise the Test Scores of High-Ability Minority Students? *American Economic Review* 106 (10): 2783-2816.

Dustmann, Christian (2004). Parental background, secondary school track choice, and wages. *Oxford Economic Papers* 56, 209-230.

Dustmann, Christian, Patrick A. Puhani, Uta Schönberg (2017). The Long-term Effects of Early Track Choice. *Economic Journal* 127 (603): 1348-1380.

Epple, Dennis, Elizabeth Newlon, Richard Romano (2002). Ability Tracking, School Competition, and the Distribution of Educational Benefits. *Journal of Public Economics* 83 (1): 1-48.

Figlio, David N., Marianne E. Page (2002). School Choice and the Distributional Effects of Ability Tracking: Does Separation Increase Inequality? *Journal of Urban Economics* 51 (3): 497-514.

Galindo-Rueda, F., A. Vignoles (2007). The heterogeneous effect of selection in U.K. secondary schools., in: Woessmann, L., Peterson, P.E. (Eds.), *Schools and the Equal Opportunity Problem*. (MIT Press), 103-128.

Grewenig, Elisabeth (2021). School Track Decisions and Teacher Recommendations: Evidence from German State Reforms. ifo Working Paper No. 353

Guyon, N., Maurin, E., & McNally, S. (2012). The effect of tracking students by ability into different schools: A natural experiment. *Journal of Human Resources*, 47, 684–721.

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

Hanushek, Eric A., Ludger Woessmann (2006). Does Early Tracking Affect Educational Inequality and Performance? Differences-in-Differences Evidence across Countries. *Economic Journal* 116 (510): C63-C76.

Hanushek, Eric A., Ludger Woessmann (2008). The Role of Cognitive Skills in Economic Development. *Journal of Economic Literature* 46 (3): 607-668.

Hanushek, Eric A., Ludger Woessmann (2011). The Economics of International Differences in Educational Achievement. In *Handbook of the Economics of Education*, Vol. 3, edited by Eric A. Hanushek, Stephen Machin, Ludger Woessmann. Amsterdam: North Holland: 89-200.

Kultusministerkonferenz (2019). The education system in the Federal Republic of Germany 2016/2017. In *Eurydice National Dossier*. Secretariat of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany.

Lazear, Edward.P. (2001). Educational Production. *Quarterly Journal of Economics* 116, 777-803.

Leschinsky, A., & Mayer, K. U. (Eds.). (1990). *The Comprehensive School Experiment Revisited: Evidence from Western Europe*. Frankfurt am Main: Peter Lang.

Luedemann, Elke, Guido Schwerdt (2013). Migration Background and Educational Tracking. *Journal of Population Economics* 26 (2): 455-481.

Malamud, Ofer, Cristian Pop-Eleches (2010). General education versus vocational training: Evidence from an economy in transition. *Review of Economics and Statistics*, 92, 43–60.

Malamud, Ofer, Cristian Pop-Eleches (2011). School tracking and access to higher education among disadvantaged groups. *Journal of Public Economics*, 95, 1538–1549

Matthewes, Sönke H. (2021). Better Together? Heterogeneous Effects of Tracking on Student Achievement. *Economic Journal* 131 (635): 1269-1307.

Meghir, Costas, Mårten Palme (2005). Educational Reform, Ability and Family Background. *American Economic Review* 95 (1): 414-424.

Muehlenweg, A. (2008). Educational effects of alternative secondary school tracking regimes in Germany. *Schmollers Jahrbuch* (*Journal of Applied Social Science Studies*), 128, 351–379

Neumann, M., M. Becker, J. Baumert, K. Maaz, O. Koeller (2017). Zentrale Ergebnisse der wissenschaftlichen Begleitstudie zur Berliner Schulstrukturreform - Zusammenfassung fuer die Presse. (DIPF, Max-Plack-Institut fuer Bildungsforschung, IPN).

OECD (2004). *Learning for tomorrow's world: First results from PISA 2003*. Paris: OECD

Pekkala Kerr, Sari, Tuomas Pekkarinen, Roope Uusitalo (2013). School Tracking and Development of Cognitive Skills. *Journal of Labor Economics* 31 (3): 577-602.

Pekkarinen, Tuomas (2014). School Tracking and Intergenerational Social Mobility. *IZA World of Labor* 214: 56.

Pekkarinen, Tuomas, Roope Uusitalo, Sari Kerr (2009). School Tracking and Intergenerational Income Mobility: Evidence from the Finnish Comprehensive School Reform. *Journal of Public Economics* 93 (7-8): 965-973.

Piopiunik, Marc (2014). The Effects of Early Tracking on Student Performance: Evidence from a School Reform in Bavaria. *Economics of Education Review* 42: 12-33.

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

Ruhose, Jens, Guido Schwerdt (2016). Does Early Educational Tracking Increase Migrant-Native Achievement Gaps? Differences-in-Differences Evidence across Countries. *Economics of Education Review* 52: 134-154.

Schuetz, Gabriela, Heinrich W. Ursprung, Ludger Woessmann (2008). Education Policy and Equality of Opportunity. *Kyklos* 61 (2): 279-308.

van Elk, Roel, Marc van der Steeg, Dinand Webbink (2011). Does the Timing of Tracking Affect Higher Education Completion? *Economics of Education Review* 30 (5): 1009-1021.

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

Woessmann, Ludger (2009). International Evidence on School Tracking: A Review. *CESifo DICE Report - Journal for Institutional Comparisons* 7 (1): 26-34.

Woessmann, Ludger (2010). Institutional Determinants of School Efficiency and Equity: German States as a Microcosm for OECD Countries. *Jahrbücher für Nationalökonomie und Statistik / Journal of Economics and Statistics* 230 (2): 234-270.

Woessmann, Ludger, Elke Luedemann, Gabriela Schuetz, Martin R. West (2009). *School Accountability, Autonomy and Choice around the World*. Cheltenham: Edward Elgar.

**Table 1: Summary Statistics by Cohort**

| **Pre-reform cohort (SC4)** | | | | | |
|---|---|---|---|---|---|
| | Count | Mean | Min | Max | SD |
| **Outcomes** | | | | | |
| Math score (grade 9) | 14,505 | 0.00 | -3.61 | 3.79 | 1.00 |
| Reading score (grade 9) | 13,873 | -0.00 | -3.74 | 2.63 | 1.00 |
| Science score (grade 9) | 14,453 | -0.00 | -3.56 | 5.30 | 1.00 |
| Middle school degree | 14,319 | 0.68 | 0.00 | 1.00 | 0.47 |
| Gymnasium (grade 10) | 14,973 | 0.34 | 0.00 | 1.00 | 0.47 |
| Gymnasium (grade 11) | 14,760 | 0.35 | 0.00 | 1.00 | 0.48 |
| Age in months (grade 9) | 15,169 | 188.36 | 140.09 | 230.14 | 7.63 |
| **Explanatory Variables** | | | | | |
| Male | 15,169 | 0.50 | 0.00 | 1.00 | 0.50 |
| Born abroad | 15,169 | 0.06 | 0.00 | 1.00 | 0.24 |
| Number of books at home | 15,169 | 3.79 | 1.00 | 6.00 | 1.39 |
| Highest parental education | 15,169 | 5.67 | 1.00 | 10.00 | 2.58 |
| Father occupation | 15,169 | 7.95 | 1.00 | 13.00 | 4.29 |
| **Post-reform cohort (SC3)** | | | | | |
| | Count | Mean | Min | Max | SD |
| **Outcomes** | | | | | |
| Math score (grade 9) | 4,789 | -0.00 | -3.49 | 4.40 | 1.00 |
| Reading score (grade 9) | 4,507 | 0.00 | -2.95 | 4.35 | 1.00 |
| Science score (grade 9) | 4,783 | -0.00 | -3.68 | 4.09 | 1.00 |
| Middle school degree | 6,283 | 0.79 | 0.00 | 1.00 | 0.41 |
| Gymnasium (grade 10) | 6,235 | 0.38 | 0.00 | 1.00 | 0.49 |
| Gymnasium (grade 11) | 5,311 | 0.55 | 0.00 | 1.00 | 0.50 |
| Age in months (grade 9) | 6,606 | 185.50 | 145.12 | 218.14 | 6.15 |
| **Explanatory Variables** | | | | | |
| Male | 6,606 | 0.51 | 0.00 | 1.00 | 0.50 |
| Born abroad | 6,606 | 0.04 | 0.00 | 1.00 | 0.20 |
| Number of books at home | 6,606 | 4.17 | 1.00 | 6.00 | 1.35 |
| Highest parental education | 6,606 | 6.19 | 1.00 | 10.00 | 2.59 |
| Father occupation | 6,606 | 9.08 | 1.00 | 13.00 | 4.15 |

*Notes:* Summary statistics reported separately for the pre-reform cohort (NEPS starting cohort 4, SC4) and the post-reform cohort (SC3). *Age in months* is measured in June of 9th grade. *Middle school degree* equals 1 if a student obtained the middle school degree (includes, among others, all students who attend grade 11 at Gymnasium); 0 otherwise. *Gymnasium (grade 10/11)* equals 1 if a student attends a Gymnasium or the Gymnasium track (*Gymnasialzweig*) at a comprehensive school in 10th/11th grade. Math, reading, and science achievement were assessed in 9th grade; test scores are z-standardized (with mean 0 and standard deviation 1) separately by cohort. *Number of books at home* contains six categories, ranging from "0–10 books" (=1) to "more than 500 books" (=6). *Highest parental education* corresponds to highest level of mother's and father's education according to the ISCED-97 classification, ranging from "secondary school certificate" (=1) to "doctorate, habilitation" (=10). *Father occupation* contains twelve categories, with missing values as a separate category.

**Table 2: Reform Effects on Student Achievement**

| | Reading | | Math | | Science | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Post x Reform state | 0.180** | 0.152*** | 0.145** | 0.087 | 0.121** | 0.067 |
| | (0.076) | (0.056) | (0.064) | (0.059) | (0.053) | (0.049) |
| Reform state | 0.001 | –0.025 | 0.106 | 0.081 | 0.037 | 0.015 |
| | (0.069) | (0.055) | (0.086) | (0.077) | (0.066) | (0.048) |
| Post | –0.088 | –0.170*** | –0.076 | –0.148*** | –0.063 | –0.152*** |
| | (0.065) | (0.050) | (0.050) | (0.049) | (0.044) | (0.042) |
| Male | | –0.215*** | | 0.313*** | | 0.182*** |
| | | (0.011) | | (0.014) | | (0.015) |
| Born abroad | | –0.228*** | | –0.185*** | | –0.265*** |
| | | (0.021) | | (0.022) | | (0.025) |
| Further covariates | No | Yes | No | Yes | No | Yes |
| N | 18,110 | 18,110 | 18,974 | 18,974 | 18,921 | 18,921 |

*Notes:* Least squares regressions. Dependent variables: reading (Columns 1 + 2), math (Columns 3 + 4) and science (Columns 5 + 6) achievement, assessed in 9th grade; test scores are z-standardized (mean 0, standard deviation 1) separately by cohort. *Reform state* equals 1 for students attending school in a reform state; 0 otherwise. *Post* equals 1 for students in the post-reform cohort (SC3); 0 for students in the pre-reform cohort (SC4). *Further covariates* include number of books at home, highest parental education, and father's occupation. Berlin is excluded since student achievement is not available for the post-reform cohort (SC3). Standard errors are clustered at the state level and bootstrapped with 1,000 replications. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

**Table 3: Reform Effects on Inequality of Student Achievement**

| | SD | p95–p5 | p90–p10 | p75–p25 | p90–p50 | p50–p10 |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| **Panel A: Reading** | | | | | | |
| Post x Reform state | -0.076 | -0.295 | -0.125 | -0.082 | -0.124 | -0.001 |
| | (0.052) | (0.191) | (0.158) | (0.107) | (0.134) | (0.070) |
| Reform state | 0.031 | 0.150 | 0.001 | 0.027 | 0.046 | -0.045 |
| | (0.042) | (0.170) | (0.143) | (0.094) | (0.104) | (0.077) |
| Post | 0.034 | 0.238 | 0.132 | 0.099 | 0.170 | -0.038 |
| | (0.050) | (0.187) | (0.164) | (0.087) | (0.116) | (0.087) |
| **Panel B: Math** | | | | | | |
| Post x Reform state | -0.018 | -0.088 | -0.062 | 0.002 | -0.133 | 0.071 |
| | (0.060) | (0.203) | (0.179) | (0.125) | (0.145) | (0.078) |
| Reform state | 0.032 | 0.114 | 0.025 | 0.048 | 0.010 | 0.015 |
| | (0.058) | (0.197) | (0.157) | (0.107) | (0.119) | (0.057) |
| Post | -0.034 | -0.116 | 0.085 | -0.009 | -0.015 | 0.100 |
| | (0.078) | (0.242) | (0.227) | (0.168) | (0.179) | (0.075) |
| **Panel C: Science** | | | | | | |
| Post x Reform state | -0.014 | -0.046 | -0.051 | -0.042 | -0.181* | 0.130* |
| | (0.051) | (0.263) | (0.120) | (0.121) | (0.095) | (0.068) |
| Reform state | 0.012 | 0.104 | 0.128 | 0.041 | 0.092 | 0.036 |
| | (0.049) | (0.203) | (0.137) | (0.101) | (0.091) | (0.061) |
| Post | 0.016 | 0.098 | 0.116 | -0.002 | 0.150** | -0.034 |
| | (0.049) | (0.212) | (0.129) | (0.083) | (0.070) | (0.090) |
| N (each panel) | 30 | 30 | 30 | 30 | 30 | 30 |

*Notes:* Least squares regression. Dependent variables: dispersion measures of reading, math and science achievement, assessed in 9th grade; student-level test scores have been z-standardized (mean 0, standard deviation 1) separately by cohort. Dispersion measures are computed at the state-cohort level. *SD* (Column 1) is the standard deviation; Columns 2–6 contain various percentile gaps; e.g., p95–p5 (Column 2) is the difference between the 95th and 5th percentile of the state-cohort-specific achievement distribution. *Reform state* equals 1 for students attending school in a reform state; 0 otherwise. *Post* equals 1 for students in the post-reform cohort (SC3); 0 for students in the pre-reform cohort (SC4). All regressions additionally control for the state-cohort-specific share of male students, share of foreign-born students, and shares of parents with a specific education level (grouped into four categories). Berlin is excluded since student achievement is not available for the post-reform cohort (SC3). Standard errors are clustered at the state level and bootstrapped with 1,000 replications. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

**Table 4: Reform Effects on Middle School Degree, Gymnasium Attendance, and Age in 9th Grade**

| | Middle school degree | | Gymnasium (grade 10) | | Gymnasium (grade 11) | | Age (grade 9) | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Post x Reform state | –0.023 | –0.031 | 0.045 | 0.022 | 0.032 | 0.013 | –0.442 | –0.310 |
| | (0.041) | (0.038) | (0.062) | (0.054) | (0.035) | (0.030) | (0.609) | (0.572) |
| Reform state | 0.019 | 0.013 | –0.010 | –0.016 | 0.002 | –0.005 | –0.263 | –0.211 |
| | (0.029) | (0.028) | (0.029) | (0.027) | (0.042) | (0.033) | (0.597) | (0.549) |
| Post | 0.126*** | 0.110*** | 0.022 | –0.008 | 0.185*** | 0.137*** | –2.650*** | –2.243*** |
| | (0.022) | (0.022) | (0.037) | (0.034) | (0.029) | (0.027) | (0.377) | (0.360) |
| Male | | –0.029*** | | –0.048*** | | –0.071*** | | 1.210*** |
| | | (0.007) | | (0.007) | | (0.007) | | (0.118) |
| Born abroad | | –0.059*** | | –0.054*** | | –0.041*** | | 3.960*** |
| | | (0.015) | | (0.017) | | (0.010) | | (0.191) |
| Further covariates | No | Yes | No | Yes | No | Yes | No | Yes |
| N | 20,602 | 20,602 | 21,208 | 21,208 | 20,071 | 20,071 | 21,775 | 21,775 |

*Notes:* Least squares regressions. Dependent variables: *middle school degree* equals 1 if the student obtained a middle school degree (includes all students who attend grade 11 at Gymnasium); 0 otherwise (Columns 1 and 2). *Gymnasium (grade 10/11)* equals 1 if the student attends a Gymnasium or the Gymnasium track (*Gymnasialzweig*) at a comprehensive school in 10th/11th grade (Columns 3–6); *age (grade 9)* is the age in months in June of 9th grade (Columns 7 and 8). *Reform state* equals 1 if the student attends school in a reform state; 0 otherwise. *Post* equals 1 for students in the post-reform cohort (SC3); equals 0 for all students in SC4. *Further covariates* include number of books at home, highest parental education, and father's occupation. Standard errors are clustered at the state level and bootstrapped with 1,000 replications. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

**Table 5: Reform Effects on Student Achievement (Heterogeneity Analysis)**

| | Reading | Math | Science |
|---|---|---|---|
| | (1) | (2) | (3) |
| **Panel A: Gender** | | | |
| Male | 0.213*** | 0.136* | 0.119** |
| | (0.068) | (0.074) | (0.058) |
| Female | 0.099 | 0.061 | 0.027 |
| | (0.068) | (0.065) | (0.062) |
| **Panel B: Migration background** | | | |
| Born abroad | 0.226* | 0.199** | 0.170 |
| | (0.118) | (0.079) | (0.106) |
| Born in Germany | 0.147*** | 0.082 | 0.057 |
| | (0.056) | (0.058) | (0.050) |
| **Panel C: Parental education** | | | |
| Low parental education | 0.216*** | 0.143** | 0.075 |
| | (0.072) | (0.060) | (0.060) |
| High parental education | 0.099* | 0.043 | 0.055 |
| | (0.053) | (0.066) | (0.058) |
| **Panel D: Number of books at home** | | | |
| Less than 100 books | 0.208*** | 0.086 | 0.005 |
| | (0.074) | (0.066) | (0.059) |
| More than 100 books | 0.139** | 0.105 | 0.120* |
| | (0.067) | (0.069) | (0.066) |

*Notes:* Least squares regressions. Each cell comes from a different regression, reporting the coefficient on *Post x Reform state*. Dependent variables: reading (Column 1), math (Column 2) and science (Column 3) achievement, assessed in 9th grade; test scores are z-standardized (mean 0, standard deviation 1) separately by cohort. Subsamples are indicated in the left column; *high parental education* includes all students whose mother or father have at least a post-secondary, non-tertiary degree. Each regression includes the following controls: *Reform state* equals 1 for students attending school in a reform state; 0 otherwise. *Post* equals 1 for students in the post-reform cohort (SC3); 0 for students in the pre-reform cohort (SC4); gender (except in Panel A), migration background (except in Panel B), highest parental education (except in Panel C), number of books at home (except in Panel D), and father's occupation. Berlin is excluded since student achievement is not available for the post-reform cohort (SC3). Standard errors are clustered at the state level and bootstrapped with 1,000 replications. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

**Table 6: Reform Effects on Middle School Degree, Gymnasium Attendance, and Age in 9th Grade (Heterogeneity Analysis)**

| | Middle school degree | Gymnasium (grade 10) | Gymnasium (grade 11) | Age (grade 9) |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| **Panel A: Gender** | | | | |
| Male | –0.030 | 0.001 | 0.003 | –0.456 |
| | (0.036) | (0.054) | (0.039) | (0.579) |
| Female | –0.032 | 0.045 | 0.023 | –0.146 |
| | (0.043) | (0.057) | (0.034) | (0.579) |
| **Panel B: Migration background** | | | | |
| Born abroad | –0.067 | 0.014 | –0.005 | 0.388 |
| | (0.076) | (0.058) | (0.064) | (1.146) |
| Born in Germany | –0.023 | 0.024 | 0.020 | –0.396 |
| | (0.037) | (0.055) | (0.030) | (0.547) |
| **Panel C: Parental education** | | | | |
| Low parental education | –0.030 | 0.015 | –0.006 | –0.339 |
| | (0.045) | (0.049) | (0.039) | (0.685) |
| High parental education | –0.039 | 0.027 | 0.020 | –0.298 |
| | (0.038) | (0.060) | (0.029) | (0.499) |
| **Panel D: Number of books at home** | | | | |
| Less than 100 books | –0.070 | –0.002 | –0.018 | –0.168 |
| | (0.046) | (0.041) | (0.041) | (0.750) |
| More than 100 books | –0.013 | 0.039 | 0.033 | –0.454 |
| | (0.036) | (0.063) | (0.037) | (0.523) |

*Notes:* Least squares regressions. Each cell comes from a different regression, reporting the coefficient on *Post x Reform state*. Dependent variables: *middle school degree* equals 1 if the student obtained a middle school degree (includes all students who attend grade 11 at Gymnasium); 0 otherwise (Column 1). *Gymnasium (grade 10/11)* equals 1 if the student attends a Gymnasium or the Gymnasium track (*Gymnasialzweig*) at a comprehensive school in 10th/11th grade (Columns 2+3); *age (grade 9)* is the age in months in June of 9th grade (Column 4). Subsamples are indicated in the left column; *high parental education* includes all students whose mother or father have at least a post-secondary, non-tertiary degree. Each regression includes the following controls: *Reform state* equals 1 if the student attends school in a reform state; 0 otherwise. *Post* equals 1 for students in the post-reform cohort (SC3); equals 0 for all students in SC4; gender (except in Panel A), migration background (except in Panel B), highest parental education (except in Panel C), number of books at home (except in Panel D), and father's occupation. Standard errors are clustered at the state level and bootstrapped with 1,000 replications. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

**Table 7: Attrition Analysis**

| | Reading | Math | # Books | Father education |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| **Panel A: Initial students** | | | | |
| Reform state x attritioned | 0.029 | 0.027 | –0.058 | –0.394 |
| | (0.102) | (0.080) | (0.144) | (0.387) |
| Attritioned | –0.381*** | –0.455*** | –0.495*** | –0.492 |
| | (0.046) | (0.033) | (0.109) | (0.303) |
| Reform state | 0.149 | 0.125 | 0.178 | 0.369 |
| | (0.099) | (0.124) | (0.114) | (0.268) |
| N | 5,193 | 5,193 | 5,256 | 3,065 |
| **Panel B: Top-up students** | | | | |
| Reform state x attritioned | 0.035 | –0.014 | –0.081 | –0.612 |
| | (0.137) | (0.204) | (0.180) | (0.654) |
| Attritioned | –0.388*** | –0.350* | –0.378** | –0.756*** |
| | (0.113) | (0.196) | (0.152) | (0.284) |
| Reform state | 0.127 | 0.186 | 0.211* | 0.265 |
| | (0.148) | (0.118) | (0.114) | (0.438) |
| N | 2,142 | 2,141 | 960 | 968 |

*Notes:* Least squares regressions. Attrition analysis for the post-reform cohort (SC3). *Initial students* (Panel A) joined the NEPS survey in 5th grade and *top-up students* (Panel B) in 7th grade. Dependent variables: The information on all dependent variables stem from the first school year when students joined the NEPS; reading achievement (Column 1), math achievement (Column 2), number of books at home (Column 3), and father's education (Column 4). *Attritioned* equals 1 if the student has left the NEPS survey until the end of 9th grade; 0 otherwise. *Reform state* equals 1 if the student attends school in a reform state; 0 otherwise. Standard errors are clustered at the state level and bootstrapped with 1,000 replications. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

# Appendix

**Table A1: Number of Student Observations by State and Cohort**

| | Post-reform cohort (SC3) | Pre-reform cohort (SC4) |
|---|---|---|
| **Reform states** | | |
| Baden-Wuerttemberg | 1,072 | 2,413 |
| Bavaria | 996 | 2,147 |
| Berlin | 96 | 341 |
| Bremen | 80 | 97 |
| Hamburg | 123 | 252 |
| Rhineland Palatine | 289 | 691 |
| Saarland | 68 | 224 |
| Schleswig-Holstein | 328 | 525 |
| **Total** | **3,052** | **6,690** |
| **Non-reform states** | | |
| Brandenburg | 589 | 325 |
| Hessen | 399 | 1,381 |
| Lower Saxony | 650 | 1,638 |
| Mecklenburg-West Pomerania | 141 | 237 |
| North Rhine Westphalia | 1,345 | 3,858 |
| Saxony | 153 | 410 |
| Saxony Anhalt | 74 | 287 |
| Thuringia | 203 | 343 |
| **Total** | **3,554** | **8,479** |

*Notes:* Number of student observations by state and cohort. The total sample size of starting cohort 4 (SC4) is 15,169 students and that of starting cohort 3 (SC3) is 6,606 students.

**Table A2: Summary Statistics (Pre-Reform and Post-Reform Cohorts Combined)**

| | Count | Mean | Min | Max | SD |
|---|---|---|---|---|---|
| **Outcomes** | | | | | |
| Math score (grade 9) | 19,294 | 0.00 | -3.61 | 4.40 | 1.00 |
| Reading score (grade 9) | 18,380 | 0.00 | -3.74 | 4.35 | 1.00 |
| Science score (grade 9) | 19,236 | -0.00 | -3.68 | 5.30 | 1.00 |
| Middle school degree | 20,602 | 0.71 | 0.00 | 1.00 | 0.45 |
| Gymnasium (grade 10) | 21,208 | 0.35 | 0.00 | 1.00 | 0.48 |
| Gymnasium (grade 11) | 20,071 | 0.40 | 0.00 | 1.00 | 0.49 |
| Age in months (grade 9) | 21,775 | 187.49 | 140.09 | 230.14 | 7.34 |
| **Explanatory Variables** | | | | | |
| Male | 21,775 | 0.51 | 0.00 | 1.00 | 0.50 |
| Born abroad | 21,775 | 0.05 | 0.00 | 1.00 | 0.23 |
| Number of books at home | 21,775 | 3.90 | 1.00 | 6.00 | 1.39 |
| Highest parental education | 21,775 | 5.83 | 1.00 | 10.00 | 2.59 |
| Father occupation | 21,775 | 8.29 | 1.00 | 13.00 | 4.28 |

*Notes:* Summary statistics for pre-reform cohort (SC4) and post-reform cohort (SC3) combined. *Age in months* is measured in June of 9th grade. *Middle school degree* equals 1 if a student obtained the middle school degree (includes, among others, all students who attend grade 11 at Gymnasium); 0 otherwise. *Gymnasium (grade 10/11)* equals 1 if a student attends a Gymnasium or the Gymnasium track (*Gymnasialzweig*) at a comprehensive school in 10th/11th grade. Math, reading, and science achievement were assessed in 9th grade; test scores are z-standardized (with mean 0 and standard deviation 1) separately by cohort. *Number of books at home* contains six categories, ranging from "0–10 books" (=1) to "more than 500 books" (=6). *Highest parental education* corresponds to highest level of mother's and father's education according to the ISCED-97 classification, ranging from "secondary school certificate" (=1) to "doctorate, habilitation" (=10). *Father occupation* contains twelve categories, with missing values as a separate category.