Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

ifo Working Papers

The Effects of Early Tracking on Student Performance: Evidence from a School Reform in Bavaria

Marc Piopiunik

Ifo Working Paper No. 153

January 2013

An electronic version of the paper may be downloaded from the Ifo website
www.cesifo-group.de.

Ifo Working Paper No.153

The Effects of Early Tracking on Student Performance: Evidence from a School Reform in Bavaria*

Abstract

This paper evaluates a school reform in Bavaria that moved the timing of tracking in low- and middle-track schools from grade 6 to grade 4; students in high-track schools were not affected. To eliminate state-specific and school-type-specific shocks, I estimate a triple-differences model using three PISA waves. The results indicate that the reform reduced the performance of 15-year-old students both in low- and middle-track schools. Further evidence suggests that the share of very low-performing students increased in low-track schools.

JEL Code: I20, I21, I24.
Keywords: Tracking, student performance, PISA.

Marc Piopiunik
Ifo Institute – Leibniz Institute
for Economic Research
at the University of Munich
Poschingerstr. 5
81679 Munich, Germany
Phone: +49(0)89/9224-1312
piopiunik@ifo.de

* I thank Stefan Bauernschuster, Francesco Cinnirella, Alexander Danzer, Oliver Falck, Elke Luedemann, Simon Wiederhold, and Ludger Woessmann, as well as conference participants 2012 at the RES in Cambridge, ESPE in Bern, EEA/ESEM in Malaga, and at the EALE in Bonn. and seminar participants at the Ifo Institute and at the University of Munich for their valuable comments and suggestions.

1 Introduction

Countries differ greatly in both the extent and the timing of separating students into different school tracks. Some countries separate students into more vocationally-oriented and more academic school tracks 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). Timing of tracking is a potentially important feature of the school system since it might affect both the level and the distribution of students' cognitive skills which, in turn, affect individual earnings and economic growth (Hanushek and Woessmann, 2008). During the last decades, 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 timing of tracking because the reforms simultaneously changed both the timing of tracking and other features of the school system, for example, the amount of compulsory years of schooling (e.g. in Sweden and Norway).

This paper studies the effect of separating students two years earlier into different school tracks on student achievement. The variation in the timing of tracking comes from a school reform in the German state of Bavaria in 2000, where students in the basic track (*Hauptschule*) and middle track (*Realschule*) were separated at the end of grade 6 prior to the reform and at the end of grade 4 after the reform. Importantly, the reform did not change the amount of schooling. Students in the most academic track (*Gymnasium*) were not affected by the reform; both before and after the reform, (future) Gymnasium students attend the four-year primary school together with the (future) basic and middle track students before entering Gymnasium (see Figure 1 for a graphical representation of the reform). The effect of the reform is estimated in a difference-in-differences-in-differences approach. Student performance before the reform is compared with student performance after the reform. To eliminate Germany-wide trends, performance is furthermore compared between students in Bavaria and students in other German states, where the timing of tracking did not change. Because the reform effect might still be confounded by state-specific or school-type-specific trends, performance is additionally compared between students in the school tracks affected by the reform (non-Gymnasium tracks) and students in the unaffected school track (Gymnasium).[1] Performance of 15-year-old students in math, reading, and science comes from the German extension studies (PISA-E) of the Programme for International Student Assessment (PISA) in 2000, 2003, and 2006.[2]

The results indicate that the reform lowered students' performance both in the basic and middle track. The performance decreased by about 13 PISA points in both tracks which

[1]The terms *school track* and *school type* are used interchangeably in this paper.

[2]Since an official data request to use the PISA-E student-level micro data was refused, the analyses in this paper are based on aggregated performance data published by the German PISA consortium. Most importantly, the reform effect on student performance is identified with the aggregated data, since the published data are representative for each school track within each state, and therefore vary at the same level as the Bavarian school reform.

equals approximately the performance gain in about half a school year. While the decline in the middle track might be due to the hiring of additional, inexperienced, teachers and to (unobservable) implementation problems related to the reform, lower student performance in the basic track is most likely due to peer effects only. Importantly, the results cannot be explained by a change in student composition since the share of students attending the different tracks remained stable in this period. Overall, the reform also increased the performance dispersion, suggesting greater inequality of opportunity since achievement is strongly correlated with family background (Hanushek and Woessmann, 2011). With early tracking, the share of very low-performing students increased in the basic track. An alternative measure of student performance—whether a student is below grade given her age—also indicates detrimental effects of the reform, especially for basic track students. Finally, using additional performance data from another student assessment in 2009, the negative impact on student performance seems to persist for several years after the reform went into effect.

The impact of tracking on student performance is theoretically ambiguous (see Betts (2011) and Meier and Schütz (2008)). On the one hand, tracking might increase student performance because teachers face more homogeneous classrooms, allowing them to adjust their teaching style to the students' ability level and to use different pedagogical methods. 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, e.g. 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, Giannini, and Ariga, 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.

Several educational reforms have been exploited to investigate the effect of timing of tracking on student performance and labor-market outcomes. 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, Salvanes, and Vaage (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 my results, Kerr, Pekkarinen, and Uusitalo (2012) find that the reform significantly improved the performance of students with low SES parents on verbal, arithmetic, and logical reasoning tests. At the same time, the reform had no negative impact on the performance of students with high SES parents. In line with these findings, Pekkarinen, Uusitalo, and Pekkala (2009) have shown that the Finnish reform also increased the intergenerational income mobility. Galindo-Rueda and Vignoles (2007) investigate a reform in the United Kingdom which replaced early tracking with comprehensive schools, finding some evidence of positive effects on student performance 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 enrol in a university or on subsequent earnings. Guyon, Maurin, and McNally (2012) investigate an educational reform in Northern Ireland that led to a large increase in the share of students admitted to the elite track at age 11. They find a strong positive overall effect of this de-tracking reform on the number of students passing national examinations at later stages and a negative effect on student performance in non-elite schools who lost their most able students. Malamud and Pop-Eleches (2010, 2011) evaluate a reform in Romania which 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 performance.

The effect of tracking on student performance has also been investigated in field experiments. Duflo, Dupas, and Kremer (2011) conduct an experiment with primary school children in Kenya, in which half of the schools tracked students based on prior achievement, while the other schools allocated children randomly to classrooms. The authors find that within-school tracking improves math and literacy performance. However, it is doubtful whether the within-school tracking experiment in Kenyan primary schools is informative for whether (or at what age) students in developed countries should be separated into different secondary school types. Slavin (1990) reviews 15 tracking experiments in U.S. and British secondary schools. Overall, effects of tracking on student performance are mixed. However, almost all experiments lasted for only one year and students in the different ability groups faced the same curriculum. Thus, effects might be very different when tracking is instituted on a permanent basis and the curriculum can be adopted to the differing ability levels.

In contrast to reform evaluation studies that exploit changes in the timing of tracking within countries over time, cross-sectional studies 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. Hanushek and Woessmann (2006) compare the performance of fourth graders, i.e. prior to tracking, with the performance of eighth or ninth graders, i.e. after some countries have already tracked their students into different school types. They find that performance inequality increases in early-tracking countries, with no significant effects on the performance level. Waldinger (2007), however, shows that these results are sensitive to model specification and to the sample of countries included. Using cross-country variation in the timing of tracking, Ammermüller (2005), Brunello and Checchi (2007), Schütz, Ursprung, and Woessmann (2008), and Woessmann, Lüdemann, Schütz, and West (2009) similarly find that student performance depends more strongly on family background in countries with early tracking. Van Elk, van der Steeg, and Webbink (2011) investigate the effect of timing of tracking in the Netherlands, exploiting the fact that students can enrol in tracked schools or in comprehensive schools at age 12. Using regional variation in the supply of schools as instruments, they find that early tracking has negative effects on completing higher education. Several cross-sectional studies have investigated the effect of tracking on student performance in Germany, which overall find no significant performance differences between early-tracked and late-tracked students (Mühlenweg (2008), Baumert, Becker, Neumann, and Nikolova (2009), and Woessmann (2010)). In sum, also cross-sectional studies find that student performance depends more strongly on family background in countries with early tracking.

This paper contributes to the small literature that investigates the effect of timing of tracking on student performance by exploiting school reforms. In particular, it studies the impact of a reform in the German state of Bavaria that moved the timing of tracking in low- and middle-track schools from grade 6 to grade 4. To this end, comparable achievement tests—administered to representative and large samples of students—are used for several student cohorts who either attended the old or the new school system. Furthermore, student achievement is measured several years after students have been tracked, such that the tracking of students had enough time to unfold its effects on students' skills. Because students are still in school at the time of testing, factors other than the school system are unlikely to explain the estimated effects. Since the reform affected only two out of three school tracks in Bavaria, we can use the unaffected track (Gymnasium) as a control group to control for Bavaria-specific trends. Using students in other German states as an additional control group, we can also take school-type-specific performance trends into account. Another advantage of the Bavarian school reform is that it did not change compulsory years of schooling, in contrast to several previous school reforms.

The paper proceeds as follows. Section 2 briefly describes the school system in Germany and Section 3 the school reform in Bavaria as well as potential reform channels. Section 4 describes the German PISA data and provides descriptive statistics. Section 5 presents the estimation strategy, and Section 6 reports the main results and robustness checks. Section 7 concludes.

2 The German School System

In Germany, children start school in the year after they turn six years old and typically attend four grades in primary school (*Grundschule*).[3] At about age 10, students are separated into different secondary school types, which differ both by duration and by curriculum. Basic schools (*Hauptschule*) provide basic general education and typically lead to a certificate after grade 9 (in few states after grade 10). Middle schools (*Realschule*) provide a more extensive general education and last six years. Instead of basic and middle schools, Saarland and the East German states have integrated schools (often called *Mittelschule* or *Regelschule*), which offer the school-leaving certificates typically obtained in basic and middle schools. The most academic track *Gymnasium* typically covers nine grades and is the only school type that exists in all German states.[4] In most states, comprehensive schools (*Gesamtschule*) exist in addition to the other school types. This school type encompasses all lower and upper secondary education levels and is typically attended only by a small fraction of students. While West German states traditionally have three different school types (basic schools, middle schools, and Gymnasium) and East German states two different school types (integrated schools and Gymnasium), each state offers the same three school-leaving certificates: those acquired at the end of basic school, middle school, and Gymnasium.

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 vocational schools that lead to a higher education entrance qualification. Specifically, students can acquire a technical school degree (*Fachhochschulreife*), which qualifies for a polytechnic (*Fachhochschule*). The Gymnasium-leaving certificate (*Abitur*) is a prerequisite for attending a university or other institution of higher education. Thus, the Gymnasium is the only secondary school track that provides direct entry into tertiary education.

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). The school track recommendation is binding in some (e.g., Bavaria), but not all states. In states with a binding recommendation, school

[3] Because authority and control over education policy lies with each state (*Bundesland*), the school structure differs somewhat across states. In two of the 16 states, Berlin and Brandenburg, primary school lasts six years. See Lohmar and Eckhardt (2010) for a detailed description of the German school system.

[4] In almost all East German states, as well as in Hamburg and Saarland, Gymnasium lasts only eight years. All other German states are currently also shortening Gymnasium duration from nine to eight years, with Schleswig-Holstein being the last state to complete this reform in 2016.

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.[5]

3 The School Reform in Bavaria and Potential Reform Channels

Before the reform, the more able students attended the Gymnasium track after the four-year primary school, while all other students attended the basic school track (see left-hand panel in Figure 1). After two years in basic school, i.e. after grade 6, the better-performing students switched to a middle school, which covered four grades, i.e. grades 7 to 10. Because student performance differed strongly at the beginning of middle school, the Bavarian parliament decided in April 2000 to institute six-year middle schools (*sechsstufige Realschule*). The hope was that tracking students two years earlier would help to reduce the performance dispersion and to raise the performance level. After the reform, middle schools thus contain six grades (grades 5 to 10) and start immediately after the four-year primary school (see right-hand panel in Figure 1). After the reform, basic and middle school students are therefore separated two years earlier than before the reform. Importantly, before the reform, basic and (future) middle school students actually studied together until the end of sixth grade since classrooms were not formed on the basis of students' abilities.[6]

Most importantly, the reform did not affect students in the Gymnasium track. First, the Gymnasium track has always started immediately after the four-year primary school, both before and after the reform. Second, the reform did not change the grade point average required at the end of primary school to attend a Gymnasium. For these reasons, the reform is unlikely to have changed the student composition in the Gymnasium track. This is supported by official statistics which show that the distribution of 15-year-old students (the age cohort tested in PISA) across the different school tracks did not change after the reform, i.e. between school year 1999/2000 (PISA 2000) and 2005/2006 (PISA 2006) (see Table 1). The rather constant student shares furthermore indicate that there was no substantial reaction to the school reform which might have led more (or less) students to choose the new six-year middle school track. This is a reassuring finding as it implies that the results are likely not confounded by a change in the student composition in either the basic or middle track.

[5] School grades in Germany range from 1 (very good) to 6 (fail). Students in Bavaria, for example, need an average grade of 2.33 in the main subjects German, math, and science (Heimat- und Sachkunde) to receive a recommendation for the Gymnasium track and an average grade of 2.66 to receive a recommendation for the middle school track.

[6] Implementation details and changes related to the reform reported in this section come from discussions with several employees of the Bavarian State Ministry of Education working in the departments responsible for basic schools, middle schools, and the pilot project (see below), respectively. For information on the political debates that accompanied the pilot project and the introduction of the six-year middle schools, see Baik (2011).

The state-wide implementation of the six-year middle school did, however, not take place within one or two school years, but rather extended over a longer period. Even before the reform in 2000, some middle schools offered six grade levels. Very few private six-year middle schools already existed even in the 1970s. Several middle schools added a fifth and sixth grade level until 1992 when the Bavarian State Ministry of Education started a pilot project to test the functioning of six-year middle schools.[7] The new school law of April 2000 did not require an immediate implementation of the reform, but rather allowed middle schools to add a fifth and sixth grade level within several years. This provided schools time to hire additional teachers and to build additional buildings, if necessary. The state-wide implementation of six-year middle schools was complete by school year 2003/2004. Because basic school students in grade 5 and 6 were guaranteed to switch to a middle school in seventh grade in the first years after the reform, 2004/2005 was the last school year in which students started attending a four-year middle school (cf. Bavarian State Ministry of Education, 2008, p. 80).

The reform may have affected student performance through several channels. As the reform changed the peer groups of both basic and (future) middle school students in grades 5 and 6 by separating them two years earlier, peer effects—i.e. any impact that classmates' background, behavior, or outcomes have on own outcome—is probably the most important channel (see Epple and Romano (2011) and Sacerdote (2011) for reviews of the peer effects literature). The most convincing studies of peer effects exploit random assignment of students into classrooms and tend to find large peer effects (Whitmore, 2005; Kang, 2007; Graham, 2008; Duflo et al., 2011). Effect sizes differ across studies, with a one standard deviation increase in peer mean achievement raising own performance by between 25 and 60 percent of a standard deviation. These findings suggest that the lower-performing basic school students should be negatively affected by the reform, whereas the better-performing middle school students should benefit. Besides the average performance, the heterogeneity of classmates' performance might also be important. In this respect, the literature is rather ambiguous: some studies find positive effects of performance dispersion, some studies negative effects, and still others find no effect (cf. Epple and Romano, 2011). In sum, the peer effects literature suggests that the reform likely hurt students in basic schools because they were earlier separated from more able peers, whereas the impact on the performance of middle school students is less clear. On the one hand, their performance may increase because they lost lower-ability peers. On the other hand, their performance may decrease if a more heterogeneous mix of classmates is beneficial.

[7] In principle, all middle schools could apply for the pilot project. However, the entity paying the material costs of the middle school (*Sachaufwandsträger*), the municipality for state-run schools and (usually) the church for private schools, had to approve the application. The involved costs could be substantial. In some cases, the need for additional classrooms meant to construct a new building. Furthermore, middle schools could only participate in the project if they were able to hire additional teachers. At the end, the school supervisory board decided which middle schools could participate in the pilot project.

Besides peer effects, the reform might have affected student performance also through other channels. Because middle school students were subject to different curricula in grades 5 and 6 after the reform, the curricula of grades 7-10 had to be revised. While the curriculum is more demanding in middle school than in basic school, it is unclear whether a more demanding curriculum automatically leads to better student performance. Because the curricula did not change in basic schools, there is no curriculum effect for basic school students. A further potential channel, again only affecting middle school students, is a change in teachers. After the reform, middle school students are taught by middle school teachers instead of basic school teachers in grades 5 and 6. Furthermore, middle schools had to hire additional new teachers because two new grades were added. This means that after the reform there were more inexperienced teachers. This might lower student performance since existing research indicates that beginning teachers perform significantly worse than more experienced teachers (see, e.g., Murnane and Phillips (1981) and Rivkin, Hanushek, and Kain (2005)). Since the number of basic school students decreased and the number of middle school students increased after the reform, school resources, such as spending per student, student-teacher ratio, and hours of instruction per teacher/class, might have changed in both tracks. However, official statistics (see Figure 4) indicate that policymakers adapted resources such that these indicators developed rather constant in both tracks in the reform years. Therefore, resource effect should play only a minor role.

4 German PISA Data

Data on student performance come from the German extension studies (PISA-E) of the Programme for International Student Assessment (PISA), conducted by the Organisation for Economic Co-operation and Development (OECD). PISA tested representative samples of 15-year-old students in math, reading, and science literacy in 2000, 2003, and 2006. The tests emphasize understanding as well as flexible and context-specific application of knowledge, containing both multiple-choice and open-answer questions. The German extension studies used the same tests as the international PISA, but increased sample sizes (in 2003 and 2006; see below) to ensure representative sampling within each school track in each of the 16 German states.[8]

As mentioned above, the aggregated performance data used in this study—published by the German PISA consortium (see Baumert, Artelt, Klieme, Neubrand, Prenzel, Schiefele, Schneider, Tillmann, and Weiß, 2002; Prenzel, Baumert, Blum, Lehmann, Leutner, Neubrand, Pekrun, Rost, and Schiefele, 2005; Prenzel, Artelt, Baumert, Blum, Hammann, Klieme, and Pekrun, 2008)—identify the reform effect since the data vary at the same level as the Bavar-

[8]PISA-E 2003, for example, tested a total of 44,580 students in 1,487 schools, with state sample sizes ranging from 1,618 students in Saxony-Anhalt to 4,904 students in Hamburg. Sample sizes of PISA-E 2000 (ca. 36,000 students) and PISA-E 2006 (ca. 40,000 students) are somewhat smaller.

ian school reform, i.e., at the school-track*state level. Note that in specifications without control variables at the student level, regressions with aggregated outcomes produce results identical to those obtained using the underlying student-level data, given that observations are weighted accordingly (see Angrist and Pischke, 2009, Chapter 3.1.2).

PISA-E 2003 and 2006 are the principal performance tests used for the analysis because the performance outcomes published for these surveys are representative for each school track in each of the 16 German states (see, e.g. Prenzel et al., 2005, pp. 14 and 386 for PISA-E 2003).[9] In contrast, performance in PISA-E 2000 is only reported for (ninth grade) students in the Gymnasium track but is not separately reported for the non-Gymnasium tracks. PISA-E 2000 also reports performance measures for all students in a state, both for 15-year-olds and ninth graders. Performance of PISA-E 2000 is primarily used to provide evidence that pre-reform trends were similar in Bavaria and control states.

To estimate the effects of the reform on student performance requires performance data that are comparable over time. Reading performance is directly comparable across all PISA surveys 2000, 2003, and 2006. Math performance, in contrast, is in principle comparable only between 2003 and 2006 (see Prenzel et al., 2005, p. 73), and science performance only between 2000 and 2003 due to changes in the test (see Prenzel et al., 2008, p. 54f,149,383).[10] However, if non-comparability of test scores is uncorrelated with the school reform in Bavaria, then including non-comparable subjects yields unbiased estimates of the reform effects. Robustness checks (not shown) show that reform effects are quite similar if non-comparable subjects are excluded. To be consistent across models, all specifications therefore pool all three PISA subjects.

The official PISA-E publications do not contain information on the number of basic and middle school students attending the new school system in Bavaria. (Note that even with the individual student-level PISA-E micro data, it is not possible to identify whether a student attended the old or the new system.) However, since the PISA-E studies sampled schools randomly, we can compute the expected share of basic and middle school students attending the new system. First, the PISA-E 2003 and 2006 publications report the grade-level distributions of tested students for each school track in each state.[11] Second, official Bavarian statistics report for each school year the number of students transiting from basic schools to four-year middle schools (old system) and from primary schools to six-year middle schools (new system) (Bavarian State Ministry of Education, 2008, p. 114). Combining this information allows to compute the approximate share of basic and middle

[9]To be precise, performance is reported for each general education school type which is attended by at least 5 percent of 15-year-old students in a state.

[10]PISA had a special focus on reading literacy in 2000, with about half the testing time devoted to this subject. Math was the focus in PISA 2003 and science in PISA 2006. The international PISA scale was standardized to have a mean of 500 and a standard deviation of 100 for all subjects in PISA 2000 and whenever a subject was the focus (except for a standard deviation of 95 for science in PISA 2006).

[11]Because grade level distributions are not reported for PISA-E 2000, the PISA-E 2003 figures are used as a proxy.

school students attending the new tracking regime.[12] According to these calculations, about 9.2 percent of basic school students and 8.5 percent of middle school students who were tested in PISA-E 2000 attended the "new system." These students were tracked early not because of the reform, but because of the few early-adopting middle schools and the middle schools participating in the pilot project. In PISA-E 2003, these shares increased to 26.2 percent for basic school and 25.4 percent for middle school students. Since only a minority of students was affected by the new tracking regime, PISA-E surveys 2000 and 2003 are considered pre-reform periods. In contrast, a large majority of students tested in PISA-E 2006 went to school under the new tracking regime (77.4 percent in the basic track and 74.7 percent in the middle track), rendering the PISA-E 2006 survey the post-reform period. Because the share of students affected by the reform did not increase from 0 to 100 percent between PISA-E 2003 and PISA-E 2006, the estimated effects provide lower bounds of the true reform effects.

Because PISA randomly sampled schools, and not individual students, one can alternatively compute the expected share of students attending the new system for each PISA-E survey based on the share of middle schools offering six grade levels relative to all middle schools. (Note that in all PISA tests, the average basic track student and the average middle track student attended grade 9.) During the school year 1995/1996, when the majority of basic and middle school students tested in PISA 2000 attended grade 5 (the first grade level of the new six-year schools), only 41 out of 326 middle schools (12.6%) contained six grade levels. This share increased slightly to 65 out of 326 schools (19.9%) until the school year 1998/1999, when the majority of PISA 2003 students attended grade 5. In 2001/2002, the second school year after the reform, when most of the PISA 2006 students were in fifth grade, already 224 out of 334 middle schools (67.0%) offered six grade levels (cf. Bavarian Statistical Office (1996) and following volumes). In this case, the increase of the expected share of students attending the new system between PISA 2000 and PISA 2003 is rather small (7.3 percentage points). This suggests that we should not expect a discernible effect of the introduction of six-year middle schools on student performance between the two pre-reform periods PISA-E 2000 and 2003.

In principle, students in all other German states can be used as a control group for students in Bavaria. However, a control state should not have introduced major educational reforms that may have affected student performance. Therefore, I exclude all states which introduced central exit exams in the basic, middle, or Gymnasium track between 2003 and 2006, i.e. between the pre-reform and post-reform period, because central exams are

[12] Direct information on whether *basic* school students attended the old or new system is obviously not available because basic school students do not switch school type after entering basic school. Therefore, I assume that the share of basic school students in the new system equals the share of middle school students in the new system (which I compute with the transition figures). The small differences between basic and middle school students attending the new system is due to the fact that students in basic school attend on average slightly lower grade levels than middle school students as 15-year-olds.

frequently found to increase student performance (see, e.g., Jürges, Schneider, and Büchel, 2005).[13] Furthermore, the sample includes all secondary school tracks except comprehensive schools (*Gesamtschulen*) because these schools offer all types of school degrees. Therefore, it is not clear whether comprehensive schools should be considered as treated (non-Gymnasium) or untreated (Gymnasium) track. However, results (not shown) are quite similar if comprehensive schools are included in the sample and either assigned to treated or untreated track.

Table 1 reports summary statistics of outcomes at the state level for Bavaria and control states for all PISA-E surveys. As almost all regressions pool the PISA subjects math, reading, and science, all performance measures are simple averages across the three subjects. Since the PISA scaling reflects the difficulty of test items, there are also competency levels available (see OECD, 2002, Chapter 16). Students achieving only competency level 1 will likely have difficulties of finding an apprenticeship position; therefore, these students are considered "students at risk." Table 2 reports outcomes for the non-Gymnasium tracks and Gymnasium for Bavaria and control states separately for pre-reform (PISA-E 2003) and post-reform (2006) period. The control variables, such as average class size and share of migrants (at least one parent born in a foreign country), vary also at the school-track*state level. To ensure that the non-Gymnasium tracks (either basic and middle school, such as in Bavaria, or integrated school) receive the same total weight as the control school track (Gymnasium) in the regressions, each non-Gymnasium track is weighted by the share of 15-year-old students attending the respective school track as a fraction of all non-Gymnasium students in the state. Results (not shown) are similar if each school track is weighted by the share of students attending that track or if each school track is weighted by the inverse of the number of different school tracks in a state.

5 Identification Strategy

To estimate the effects of the Bavarian school reform, I first use a difference-in-differences approach which compares the performance change of the Bavarian basic and middle track students between PISA-E 2003 and 2006 with the corresponding performance change in the non-Gymnasium tracks in the control states:

$$\Delta^2_{nG} = (Y^{2006}_{Bavaria,nG} - Y^{2003}_{Bavaria,nG}) - (Y^{2006}_{control,nG} - Y^{2003}_{control,nG}) \quad (1)$$

where the subscript nG denotes non-Gymnasium tracks and $Y^{t}_{s,nG}$ represents the performance in the non-Gymnasium tracks in state s (Bavaria or control state) at time t (2003 or

[13] Central exit exams were introduced in Brandenburg in 2003 and 2005, Hessen in 2004, Hamburg and Mecklenburg-Western Pomerania in 2005, and Berlin, Bremen, and Lower Saxony in 2006. Among the seven excluded states are the two German states, Berlin and Brandenburg, that track students after six years of primary school.

2006). Four different performance measures are used (see Table 2): average performance, standard deviation of performance, share of students achieving at most competency level 1 (low performers), and share of students achieving competency level 4 or higher (top performers).

Performance is pooled across all non-Gymnasium tracks and all PISA subjects for PISA-E 2003 and 2006 to estimate the double-differences estimator in Equ. (1):

$$Y_{st} = \alpha_1 + \alpha_2 Bavaria_s + \alpha_3 PISA2006_t + \beta(Bavaria * PISA2006)_{st} + \epsilon_{st} \qquad (2)$$

where Y_{st} is the outcome in state s at time t and ϵ is a random error term. $Bavaria$ identifies school tracks in Bavaria and captures average performance differences between students in Bavaria and control states. $PISA2006$ captures Germany-wide performance trends. The coefficient of interest, β, identifies the effect of the Bavarian school reform, i.e. β measures the difference-in-differences defined in Equation (1).

The identifying assumption that underlies the difference-in-differences estimate is that the performance changes of students in the non-Gymnasium tracks between PISA-E 2003 and 2006 would have been similar in Bavaria and control states without the school reform in Bavaria. The drawback of the difference-in-differences approach is that it does not control for state-specific shocks, which might similarly affect all students in a state, for example, due to changes in primary school. Because the Bavarian reform did not affect students in Gymnasium, I use the performance change of Gymnasium students to control for Bavaria-specific shocks. Adding Gymnasium students as a further control group yields a difference-in-differences-in-differences, or triple-differences, approach (see Hamermesh and Trejo, 2000):

$$\Delta^3 = \Delta^2_{nG} - \Delta^2_G \qquad (3)$$

where Δ^2_{nG} is the double-difference estimator for the non-Gymnasium tracks from Equ. (1) and Δ^2_G is the analogous estimator for the Gymnasium track. The second term (Δ^2_G), which reflects performance changes of Gymnasium students in Bavaria relative to Gymnasium students in other states, is supposed to capture Bavaria-specific time trends. The interpretation of the triple-differences coefficient as a causal effect of the reform therefore relies on a weaker assumption: without the reform, the performance difference between students in the non-Gymnasium tracks and students in the Gymnasium track would have developed similarly in Bavaria and control states. The performance of Gymnasium students is added to the double-differences model in Equ. (2) to compute the triple-differences estimator:

$$
\begin{aligned}
Y_{ist} =& \alpha_1 + \alpha_2 Bavaria_s + \alpha_3 PISA2006_t + \alpha_4 nonGym_i + \alpha_5 (Bavaria * PISA2006)_{st} \\
&+ \alpha_6 (Bavaria * nonGym)_{is} + \alpha_7 (PISA2006 * nonGym)_{it} \\
&+ \beta (Bavaria * PISA2006 * nonGym)_{ist} + \epsilon_{ist}
\end{aligned} \tag{4}
$$

where Y_{ist} is the performance in school track i in state s at time t, and *nonGym* identifies non-Gymnasium school tracks. The new interaction term *PISA2006*nonGym* controls for non-Gymnasium-specific trends in test scores and *Bavaria*PISA2006* for Bavaria-specific trends. The coefficient β identifies the triple differences in Equ. (3). All standard errors are clustered at the state level, allowing for correlation of error terms within states over time.[14]

Identification of reform effects would be hampered if other major educational reforms were implemented in Bavaria or in the control states between the pre-reform (2003) and the post-reform period (2006). Especially other reforms in the basic or middle school track in Bavaria would confound the effects of the tracking reform. Indeed, a new type of class (*M-Klassen*) was introduced in the basic track in Bavaria in the school year 1999/2000. These classes provide basic track students the opportunity to acquire a middle school degree after grade 10. To prepare students for the higher degree, the curriculum of these classes is more demanding.[15] Therefore, this new type of class might have increased the performance of basic track students. However, the new class type existed already during the pre-reform period and the share of ninth-grade basic school students attending this class type increased only marginally between 2002/2003 (16.3 percent) and 2005/2006 (18.1 percent). Given that this increase is rather small and given that the effect of this class type on student performance is likely limited, it very likely does not confound the effects of the tracking reform. Another important reform in Bavaria reduced the length of Gymnasium from nine to eight years. But since the first cohort graduated from the eight-year Gymnasium only in 2010/2011, this reform did not affect the Gymnasium students who were tested in PISA-E 2003 or 2006. A Germany-wide education reform was implemented by the Secretariat of the Standing Conference of the Ministers of Education (*Kultusministerkonferenz*) which decided to introduce new educational standards that define general educational goals and specify competencies that students of a certain grade level should have acquired. Because these new standards were implemented in all German states at the same time, this reform does not confound the Bavarian reform effect since Germany-wide shocks are eliminated in the triple-differences model.[16] Some major educational reforms were implemented in some of

[14]Clustering standard errors at the school-track*state level yields quite similar results.

[15]A voluntary tenth grade in basic schools has existed since the school year 1994/1995. However, special classes that prepared students for the middle school degree did not exist.

[16]Educational standards were implemented in the subjects German, math, and first foreign language for the basic school degree (grade level 9) in the school year 2005/2006 and for the middle school degree (grade level 10) one year earlier (see also Section 6.5).

the control states between 2003 and 2006, such as the shortening of Gymnasium duration from nine to eight years. These reforms will be addressed in the results section.[17]

6 Results

Figure 2 depicts the development of students' reading performance in the non-Gymnasium tracks and Gymnasium in Bavaria and in the control states over the three PISA surveys. While the average reading performance of Gymnasium students developed similarly in Bavaria and control states both before and after the reform, the reading performance of non-Gymnasium students developed similarly only before the reform, but decreased in Bavaria after the reform. Figure 3 presents a very similar pattern for the state-specific shares of very low-performing and high-performing students across all PISA subjects math, reading, and science. In line with Figure 2, both shares—adjusted for Germany-wide trends[18]—developed more or less similarly in Bavaria and control states before the reform, while the share of low-performing students increased considerably in Bavaria, by about 3 percentage points, relative to the control states after the reform.[19] Apart from a small number of students in special education schools and some students in vocational schools, low-performing students attend either a basic or middle school, while most top-performing students attend a Gymnasium. Figure 3 thus suggests that the reform increased the share of low-performing students in Bavaria, without affecting the share of top-performing students.

Before presenting the regression results, I briefly report simple diff-in-diff estimates based on the descriptive statistics for Bavaria only (cf. Table 2). These estimates compare the performance changes between the non-Gymnasium tracks and the Gymnasium in Bavaria between 2003 and 2006. Using (only) the Gymnasium students in Bavaria as a control group, the results indicate that the reform lowered average student performance in the non-Gymnasium tracks (basic and middle schools) by 9 PISA points between 2003 and 2006. While the effect on the standard deviation is very small (0.7 PISA points), the share of low-performing students increased considerably in the non-Gymnasium tracks (1.3 percentage points; or 8 percent compared to the baseline share in PISA 2003). At the same time, the share of top-performing students strongly decreased in basic and middle schools compared to Gymnasium students (-4.4 percentage points). While these simple calculations do not take into account school-track-specific performance trends, they provide first hints that the Bavarian reform lowered student achievement in the affected tracks.

[17]Three states (Mecklenburg-West Pomerania, Saarland, and Saxony-Anhalt) have replaced the basic and middle track by integrated schools around the year 2000. Because these reforms have largely been finished until 2003 (see Table 4.1 in Lüdemann, 2011), this does not confound my estimates.

[18]The shares are regressed on survey year dummies because math and science performance is not directly comparable across all PISA waves.

[19]Plotting the performance development for each control state separately (not shown) reveals that the increase in the share of low-performing students was higher in Bavaria than in any control state.

6.1 Main Results

First, I present the results of the double-differences approach which compares performance changes between PISA 2003 and 2006 of students in basic and middle schools in Bavaria with the corresponding changes of students in the non-Gymnasium tracks in the control states (Columns (1)-(4) in Table 3). The estimates suggest that the reform lowered the average performance across all PISA subjects by 13 PISA points (Column 1) and increased the performance dispersion by 4 points (Column 2). The increase in performance dispersion might be explained by an increase in the correlation between test scores and family background, indicating lower equality of educational opportunities, which is in line with earlier studies Brunello and Checchi (2007). Finally, the lower performance level seems to be due both to a larger share of low-performing students (Column 3) and to a smaller share of top-performing students (Column 4).

The results of the triple-differences model, which controls for state-specific shocks by adding Gymnasium students as an additional control group, are very similar (Columns 5 to 8): the effect on the performance level and on the share of top performers is somewhat weaker, while it is somewhat stronger on the performance dispersion, and very similar on the share of low performers. The similarity of the double-differences and triple-differences results indicates that performance trends in Bavaria are similar to those in the control states. Finally, the coefficients on the other variables are in line with well-known findings: Bavarian students outperform students in other German states and German students improved their performance between PISA 2003 and PISA 2006 (cf. Prenzel et al., 2008).

How large is the effect of the reform on the performance level? To get an idea of the effect size, note that the competency increase in one school year is on average about 40 points in PISA 2003 (cf. Prenzel et al., 2005) and 25 to 30 points in PISA 2006 (see Prenzel et al., 2008, p. 59). Hence, the effect of 10.5 PISA points is nontrivial, reflecting the performance increase of between a quarter and a third of a school year. Relative to the performance standard deviation of Bavarian basic and middle school students in PISA-E 2003 (90.5 points; see Table 2), the effect on mean achievement is 11.6 percent. The effect size can also be compared to those of previous tracking studies. Kerr et al. (2012), for example, find for the Finnish comprehensive school reform an increase in test scores of men from low-education families by 3 percent of a standard deviation and in the case of verbal test scores of 6 percent of a standard deviation. Based on a field experiment, Duflo et al. (2011) find that within-school tracking in Kenyan primary schools increased test scores by 0.14 standard deviations (and 0.18 standard deviations after controlling for baseline scores and other control variables). These comparisons indicate that the effect size estimated for the Bavarian tracking reform seems to be in the range of effect sizes of previous tracking studies.

6.2 Robustness Checks

This section presents several robustness checks. First, alternative states are used as control group for Bavaria. Then, additional control variables at the school-track level are added to the model. Finally, evidence on similar pre-reform trends is presented as well as reform effects based on state-level performance measures which are similar to those based on school-track specific performance measures.

6.2.1 Alternative Control States

Causal interpretation of the reform effect in the triple-differences model relies on the assumption that the development of the performance difference between non-Gymnasium tracks and Gymnasium track would have been similar in Bavaria and the control states without the Bavarian reform. However, this assumption might not hold, for example, if the control states that performed quite poorly in PISA-E 2003 exerted more pressure on (or provided more support for) non-Gymnasium school types than on Gymnasium schools to improve student performance. This might have been the case, since the more able Gymnasium students on average considerably outperform non-Gymnasium students in all German states. Therefore, political pressure, accompanied maybe with minor educational reforms, might have led to a stronger increase in student performance in the non-Gymnasium tracks than in the Gymnasium track in the control states. In this case, the assumption would be violated and the estimated reform effect would be biased. To test whether the results are driven by such a "catching-up" process, I use only the three (respectively seven) best-performing states (after top-performing Bavaria) in the pre-reform period assessment, PISA-E 2003, as control states.[20] Panels C and D of Table 4 show that using only high-performing states as control group leads to very similar results than using the main control group (see Panel A).

A further alternative control group for Bavaria consists of all other German states. Including also those states that have introduced central exit exams between 2003 and 2006 changes results only slightly (Panel B). Finally, I use only those states as control group that have a school structure similar to Bavaria, i.e. states with large shares of students attending only the three tracks basic school, middle school, and Gymnasium (Baden-Württemberg, Lower Saxony, and Schleswig-Holstein). Using the states with a similar school structure indicates somewhat stronger effects in the triple-differences specifications (see Panel E). Note that some estimates lose statistical significance as samples get smaller.

Finally, I have also investigated whether the effects are driven by one of the states in the main control group. To do so, I have excluded each control state individually from the analysis (results not shown). The coefficients on the reform indicator are quite similar for any

[20]While Bavaria is the best-performing German state in the PISA assessments (with an average score of 521.6 points in the math and science tests in 2000 and 2003), this does not mean that there is no scope for improvement: the average score of the PISA winner Finland in the same tests was on average 541.5 points, which is substantially, about 20 percent of the international standard deviation, higher than in Bavaria.

one of the eight control states excluded from the sample. The robustness of the reform effects with respect to alternative control states strengthens the interpretation that the coefficients on the reform indicator indeed reflects the effect of the Bavarian reform. Especially since minor educational changes, which might have occurred in some school tracks in the control states, are hard to detect, it is reassuring that the estimated reform effects are robust to using various alternative sets of control states.

6.2.2 Controlling for School-Track-Specific Factors

While the distribution of 15-year-old students, the age cohort tested in PISA, across the Bavarian school tracks did not change after the reform (see Section 6.3), there might be other, school-track-specific factors which might have been affected by the reform. One factor is the share of migrants since migrants in Germany have lower test scores than natives (cf. Prenzel et al., 2008). Furthermore, differential changes in PISA-E participation rates might affect the reform estimates if the likelihood of taking the test is correlated with the student's performance, especially with lower-performing students being more likely not to participate (see Chapter 1.4 in Prenzel et al. (2008) and Chapter 2.4 in Prenzel et al. (2008) on non-participation rates in PISA-E 2003 and 2006). Finally, average class size might affect student performance, although this is unlikely as Woessmann (2010) finds that a state's average class size is not statistically significantly associated with student performance in Germany. All these school-track-specific factors are added in the specifications of Table 5. However, the reform coefficients barely change when these controls are included, indicating that the reform effects are not mediated through any of these channels.[21]

6.2.3 Pre-Reform Trends

Since PISA tested students also in 2000, it is possible to investigate whether test scores developed similarly in Bavaria and control states before the reform. Because school-track-specific outcomes in PISA-E 2000 are reported only for the Gymnasium track, the pre-reform trends between 2000 and 2003 rely almost only on performance measures at the state level, such as percentiles. Because some specifications also include important control variables at the state level, pre-reform trends are estimated with the difference-in-differences framework of Equ. (2), with the dependent variables being the difference between a low-performance and a high-performance measure. The measure of low performance (e.g. 10th percentile)

[21] Another robustness test (not shown) concerns the fact that the Gymnasium students of Sachsen-Anhalt and Saarland who participated in PISA 2006 were subject to the newly introduced G8 reform which shortened the Gymnasium duration from nine to eight years (see footnote 4). Because these reforms might have affected student performance, I have added a dummy variable for the affected tracks in all specifications, leaving results basically unaffected.

thereby reflects the performance of students in the non-Gymnasium tracks, whereas the high-performance measure (e.g. 90th percentile) reflects performance of Gymnasium students.[22]

Panel A of Table 6 shows that the performance gap between low- and high-performing students developed quite similarly in Bavaria and control states before the reform. Only the coefficient on the performance gap between the 10th and 90th percentile is somewhat larger, though statistically insignificant. One issue that might bias the estimates is that the low-performance measures do not only reflect the performance of basic and middle school students, the treatment group, but also the performance of special education students. Because students typically attend special education schools either since the beginning or directly after primary school, special education students typically never attend a basic or middle school and are therefore not affected by the reform. The coefficient on *Bavaria*PISA 2003* might therefore be biased if the share of special education students developed differently in Bavaria than in the control group. Similarly, the coefficient might be affected by differential trends in the share of Gymnasium students. Controlling for the shares of special education and Gymnasium students does, however, not affect the coefficient of interest.[23]

6.2.4 Reform Effects with State-Level Outcomes

The state-level performance measures used to estimate the pre-reform trends can also be used to estimate the reform effects by adding the PISA 2006 performance. This has the advantage that pre-reform trends and the reform effects are based on the same outcomes. The coefficients on the interaction term *Bavaria*PISA 2006* in Columns 1 to 4 (Panel B) are negative and statistically significant, indicating that the performance of low-ability students declined in Bavaria relative to low-ability students in the control states. Consistent with Figure 3, the reform increased the share of very low-performing students (Columns 5 and 6). The reform effect on mean reading performance is about 15 PISA points (Columns 7 and 8). As before, controlling for state-specific shares of special education and Gymnasium students does not change the results. Furthermore, the reform effects are quite similar if either the first pre-reform period (PISA-E 2000) or the second pre-reform period (PISA-E 2003) is omitted (results not shown).

Given that the state-specific performance measures also include the test scores of students in other non-Gymnasium tracks, i.e. special education and vocational schools, it is reassuring that the reform effects estimated with performance outcomes at the state level are very similar to the effects based on outcomes at the school-track level. For example, the effect on the share of low-performing students is 4 to 5 percentage points with state-level outcomes (see Columns (5) and (6) in Panel B of Table 6) and is 5.2 percentage points in the

[22]Note that without the control variables, the double-differences approach with the differenced dependent variable yields results identical to those in triple-differences models with the low-performance measure as dependent variable and the high-performance measure as control variable.

[23]The share of vocational school students is not included because these students previously attended either a basic or middle school, thus rendering the share of vocational school students endogenous.

model which includes only the share of low-performers in the basic and middle track (see Column (7) in Table 3). Furthermore, the reform effect on mean reading performance is 14 to 15 PISA points in the model in which non-Gymnasium performance includes test scores of special education and vocational school students (Columns (7) and (8) in Panel B of Table 6) and is 15.6 points in the model which excludes these two groups (see Column (5) in Panel A of Table 7). The remarkably similar results across specifications with state-level and school-track-level outcomes indicates that models with outcomes at the school-track level that ignore the performance of vocational school students (who were also affected by the reform) yield unbiased reform effects.

6.3 Effect Heterogeneity

All specifications so far pooled the three PISA subjects math, reading, and science. The change in the timing of tracking might have had, however, differential effects on the performance in these subjects. Therefore, I estimate the main specifications (see Table 3) for each PISA testing domain separately. The results in Table 7 show that the effects in the double-differences models on the performance level is quite similar across subjects, with effects on the other performance measures varying a bit across domains. Controlling for state-specific performance trends in the triple-differences models alters the effects somewhat. While the effect on mean performance remains almost unchanged for reading and science, it decreases by half for math. The effect on the performance dispersion now becomes significantly positive for reading and math. The effects on the share of low and high performers change only little. Overall, the reform seems to have affected the performance in the three subjects in a similar way: it lowered the achievement level, increased the performance dispersion, increased the share of low performers, and reduced the share of top performers (except in science).

Interestingly, the evidence that timing of tracking has stronger effects on reading than on math performance has also been found by a previous study. For the Finnish school reform, which postponed tracking by five years, Kerr et al. (2012) find for men from low-education families that the reform had a much stronger effect on the performance in the verbal test than in the math and logical reasoning tests. As discussed in Section 6.3, the most important channel in the case of the Bavarian school reform is likely peer effects. Since reform effects seem to be strongest for reading, peer effects should therefore also be larger for reading than for math (or science). There is support for this assumption as previous work on Austria—a country with the same language, similar culture, and the same early tracking age than in Germany—has found significant positive peer effects on students' reading achievement, but less significant effects for math (Schneeweis and Winter-Ebmer, 2007). Importantly, these results are also based on PISA test scores. The results further indicate that peer effects are larger for students from low socioeconomic backgrounds (which corresponds to the socioeconomic background of students in German basic and middle schools). Consistent with their finding, other studies suggest that peer effects in math are more important in early school years, while peer effects in language become more important in later grades (Levin, 2001; Vigdor and Nechyba, 2007).

So far, we have investigated the average effect of the reform on basic and middle school students. However, reform effects might differ across the two tracks because basic school students lost better-performing peers in grades 5 and 6, while middle school students lost lower-performing peers and faced a new curriculum and new teachers. To investigate the effects on basic and middle school students separately, I use as control states only those states that have a school structure similar to Bavaria, that is, states with large shares of students in the three tracks basic school, middle school, and Gymnasium (see also Panel E of Table 4). Using only these states as control group has the advantage that basic track students in Bavaria are compared with basic track students in other states, nut not with, for example, students in integrated schools. To estimate the effect on basic school students, I exclude the middle school track of all states such that the sample includes only basic school and Gymnasium tracks (see Panel A in Table 8). Similarly, I exclude all basic school tracks to estimate the reform effects on middle school students (Panel B).

The coefficients on the reform indicator suggest that the reform lowered the performance level of basic and middle school students by a similar amount.[24] This seems odd since the reform affected basic track and middle track students rather differently (see Section). The performance decline of basic track students can most likely be explained by peer effects as they lost better-performing peers in grade 5 and 6 (while curriculum and teachers did not change). In contrast, middle school students were separated two years earlier from lower-performing peers. It is rather unlikely that this caused a (strong) performance decline. Rather, it is likely that other features of the reform, especially the hiring of additional, inexperienced teachers and (unobservable) implementation problems related to the reform, caused the performance decline in the middle school track. Previous research, for example, has shown that inexperienced teachers during their first two or three years perform significantly worse than more experienced teachers (see, e.g., Murnane and Phillips (1981) and Rivkin et al. (2005)). A deterioration of school resources, such as student-teacher ratio or instruction time, in Bavarian middle schools, is, however, unlikely to explain the negative reform effects because school resources maintained rather stable around the reform years (see Section 6.5).

A decline in the performance level does nothing say about the type of student (low, middle, or high ability) who suffers from the reform. Columns (3) and (4) suggest that the performance level decreased for different reasons in the basic and middle track. In the basic track, especially the share of very low-performing students increased. In contrast, there

[24]Because the reform indicator is now identified off only one school track and because of the reduced sample sizes, standard errors are larger and therefore most coefficients are not statistically significant anymore.

seems to be no discernible increase in very low-performing students in the middle track. Instead, the share of very high-performing students declined in the middle track. This pattern of reform effects is plausible since very low-performing students are more likely to be encountered in basic schools[25], while top-performing students are more likely to attend middle schools. Finally, the effect on the performance dispersion is stronger in the basic track (Column 2). Since student achievement is strongly correlated with family background (see Hanushek and Woessmann, 2011), this suggests that the reform lowered the equality of opportunity especially in the basic school track.

6.4 Being on Grade-for-Age as Alternative Outcome

In addition to the test scores on the standardized PISA test—which is the preferred measure for assessing student performance—we can also investigate the effect of the reform on an alternative measure of student achievement: whether a student is below grade given her age. Failure to be on grade-for-age has already been used as an indicator of grade repetition and, more generally, of low performance (see, e.g., Oreopoulos, Page, and Stevens, 2006; Dahl and Moretti, 2008; Cascio, 2009). In Germany, students have to repeat a grade if certain subject grades (especially in the main subjects math and German) are below a pre-defined cutoff. Therefore, repeating a grade strongly depends on the student's academic performance. As children in Germany start school in the year after they turn six, the 15-year-old PISA participants should at least attend 9th grade; this is what the vast majority of PISA participants also does (see Table 2). I therefore classify a student as failing to be on grade if he or she attends only grade 7 or 8. The official PISA-E publications report for each school track in each state the grade-level distributions of the 15-year-old study participants, who basically attended grades 8 to 10 (tiny fractions of students attending grade 6 or 11 will be ignored).

Table 9 presents the effects on the share of students failing to be on grade-for-age, defined as students attending either grade 7 or grade 8 (again using only those three states with school structure similar to Bavaria as control group). On average, the reform increased the share of students below grade-for-age by 9.3 percentage points (Column 1). The descriptive statistics in Table 2 show that this effect is driven by an increase in the share of students below grade-for-age in the Bavarian basic and middle school tracks. At the same time, the share of students failing to be on grade-for-age even declined in Germany—both in the Bavarian Gymnasium (which was not affected by the reform) and in the Gymnasium as well as in the non-Gymnasium tracks in the control states.

The effect was much stronger for students in basic schools (13.6 percentage points; Column 2) than for students in middle schools (3.6 percentage points; Column 3). The strong effect on basic track students is sensible since basic track students are from the lower end

[25]The share of students in Bavaria achieving at most competency level 1 in PISA-E (across all subjects) is about 30 percent in basic schools but only about 2 percent in middle schools.

of the ability distribution and generally have a higher risk of repeating a grade than middle school students. Failing to be on grade-for-age can, of course, also arise if students enroll in primary school one year later than officially scheduled. In the triple-differences model, however, Bavarian-specific trends in enrollment behavior is eliminated. The descriptive statistics in Table 2 provide corroborating evidence: the share of students below grade 9 develops very similar in the Gymnasium track in Bavaria (not affected by the reform) between PISA 2003 and 2006 and in the Gymnasium tracks in the control states, thus suggesting similar trends in enrollment behavior across states.[26]

6.5 Potential Implementation Problems and Persistence of Reform Effects

As with any school reform, one is interested in the persistent effects of the reform and not in the immediate (transitory) effects that could arise due to implementation problems which gradually disappear over time. The Bavarian school reform might had detrimental effects on the academic achievement of cohorts attending the middle school track during the first reform years because there might have been a lack of teachers or because teachers might have had to adapt to the revised curriculum. (Note that these arguments are potential (transitory) reform problems for middle schools, but none applies to basic schools and can therefore not explain the performance decrease of basic school students.) Using official statistics, we can check whether a decline in spending per student or a shortage of teachers in middle schools, which could have led to larger classrooms and/or less instruction time, might explain the negative reform effects. Since the number of students in middle schools sharply increased after the reform, it is little surprising that spending per student declined somewhat (see Panel (a) in Figure 4). (Unfortunately, comparable figures are not available for the previous years.) Inversely, spending per student in the basic track, which lost students due to the reform, increased. The trends in spending per student, did, however, not translate in analogous changes in the student-teacher ratios. The ratio of students to teachers (in full-time equivalents) remained rather stable in middle schools and even improved somewhat in basic schools for students in grades 5 and 6 (see left hand of Panel (b)). In grades 7 through 9/10, the student-teacher ratio in middle schools increased only slightly over time. In line with this finding, Table 2 shows that average class size in the Bavarian non-Gymnasium tracks were very similar for the student cohorts tested in PISA 2003 and PISA 2006. Panel (c) shows that the average teacher gave slightly more hours of instruction

[26]The school-track-specific grade distribution of students, which partly reflects grade repetition, is not included as a control variable in the test score regressions because grade repetition itself is affected by the reform through its effect on academic achievement. Therefore, it is a bad control variable and should be omitted from the regression (Angrist and Krueger, 1999). Not surprisingly, including the school-track-specific grade distribution as additional control variable in the specifications of Table moves the reform coefficients towards zero (results not shown). This reflects the fact that low-performing students are more likely to repeat a grade, i.e. less likely to be on grade-for-age.

per week in the years after the reform, but this is similarly true for all tracks. Concerning the average weekly hours of instruction per class, Panel (d) shows that in all tracks the average class received the same amount of instruction time before and after the reform.[27]

While we can likely exclude that teacher shortage—implying increased class sizes, less instruction time per class, or more workload for teachers—caused student performance to decline, there might be other, unobservable, factors that lowered the performance of students in middle schools only in the first years after the reform. Especially, if beginning teachers are less able than experienced teachers, then students might be hurt by newly entering teachers. (Unfortunately, there are no official statistics on the number of newly hired teachers.) Similarly, middle school teachers might have had difficulties to adapt to the new, revised curriculum. Importantly, remember that these factors play no role for explaining the performance decline in the basic track.

To provide some more direct evidence that the negative effects of the Bavarian school reform are persistent, I add reading test scores of students who attended secondary school several years after the reform had started. Instead of nationally extending PISA 2009, the Secretariat of the Standing Conference of the Ministers of Education (*Kultusministerkonferenz*, or KMK) decided to test the reading competencies of ninth graders in 2009 as part of a quality check of the recently introduced educational standards (*Bildungsstandards*). These standards were implemented by the KMK in the school year 2003/2004 for students aiming at a middle school degree and are binding for all 16 German states. While the educational standards tested in 2009 apply only to students striving for a middle school degree, the aim of the student assessment was broader: to compare reading competencies of all ninth graders in all general education school tracks across German states.[28] Importantly, all Bavarian basic and middle school students participating in this test attended secondary school completely under the new school system. The educational standards test was linked to the international PISA 2009 survey in two ways. First, the student samples overlap. Two ninth grade classrooms in 201 general education schools that participated in PISA 2009 were tested in PISA on the first day. On the second day, these ninth graders participated in the educational standards test. Additional schools were furthermore sampled to ensure comparability across school tracks and German states. Altogether, 36,605 ninth grade students from 1,655 classrooms and 1,466 schools participated in the educational standards assessment. Second, the competency scale of the educational standards test was linked to

[27]Note that the increase in weekly instruction hours in Gymnasium in 2004 is due to the G8 reform, which shortened the Gymnasium duration from 9 to 8 years. Since the first cohort affected by the G8 reform finished school in 2011, this does not affect the reform estimates. Unfortunately, there are no official statistics on canceled classes.

[28]The educational standards (only in German) can be downloaded from http://www.kmk.org/bildung-schule/qualitaetssicherung-in-schulen/bildungsstandards/dokumente.html. In addition to reading, listening and orthography competencies in the subject of German and reading and listening competencies in the subject of English were tested in 2009. For details and results of the educational-standards-based assessment, see Köller, Knigge, and Tesch (2010).

the PISA scale to ensure maximal comparability with the competency levels in PISA. The scale was standardized such that the mean (496 points) and the standard deviation (92 points) are identical to those of the reading results of German ninth graders in PISA 2000. The fact that the three best-performing states in the PISA 2006 reading assessment are also the three best-performing states on the educational-standards-based reading assessment in 2009 is one indication that the test scores of the educational-standards assessment and the PISA test scores reflect similar reading competencies, and are therefore comparable.[29]

Table 10 presents evidence that the negative effects of the reform on students' reading performance are persistent. The dependent variable is the mean reading performance of non-Gymnasium students and Gymnasium students, respectively, within each state. Note, however, that the reading performance is based on slightly different student groups across the surveys.[30] The coefficient on *Reform 2009*, which equals 1 for non-Gymnasium students in Bavaria in 2009 and 0 otherwise, indicates that the performance of non-Gymnasium students was lower in 2009 compared to the average performance in 2000 and 2003 (see Column 1 in Table 10). Column (1) also suggests that reading performance in 2009 was slightly better than in 2006.[31] Since the reform was not implemented immediately, I exclude the two intermediate surveys, PISA 2003 and PISA 2006, such that the sample contains only students who either attended (almost) entirely the old school system (2000 cohort) or the new system (2009 cohort). Excluding the two intermediate surveys yields a similar, slightly smaller, coefficient on the reform 2009 indicator (Column 2). In sum, the results based on the reading performance of ninth graders in 2009, who exclusively attended the new school system, suggest that the detrimental impact of the reform is not only due to difficulties associated with the reform implementation, but rather indicate that the reform has persistent negative effects on student performance.

7 Conclusion

Countries differ greatly in the timing of separating students into vocationally-oriented and academic school tracks. While several countries have postponed tracking during the last decades, the German state of Bavaria has gone the opposite way: students in the basic and

[29]Overall, the state ranking is similar between PISA 2006 and the assessment in 2009. For example, two of the three worst-performing states in 2009 were the two worst-performing states in PISA-E 2006. Furthermore, the difference between the best-performing state (512 points) and the worst-performing state (474) in PISA 2006 is strikingly similar to the respective difference in 2009 (509 vs. 469 points).

[30]For example, the reading performance refers to 15-year-old students in 2003 and 2006 and refers to ninth grade students in 2009. Similar to the PISA-E surveys, the mean reading performance for non-Gymnasium students in 2009 is computed on the basis of the mean state-specific reading performance, the mean reading performance of Gymnasium students, and the share of Gymnasium students. See also notes of Table 10.

[31]Two states, North Rhine-Westphalia and Schleswig-Holstein, introduced central exit exams in all secondary school types between PISA 2006 and the IQB testing in 2009. Adding a dummy variable for both non-Gymnasium and Gymnasium tracks in these two states in 2009 does, of course, not change the reform coefficients in the triple-differences model.

middle track were separated at the end of grade 6 until 2000, but are separated at the end of grade 4 since the reform. This paper evaluates the impact of the Bavarian reform on student performance, based on PISA test scores of 15-year-olds. Effects are estimated in a difference-in-differences-in-differences model, using students in the most academic track (Gymnasium), who were not affected by the reform, and students in other German states as control groups.

The results indicate that the reform considerably decreased student performance both in the basic and middle track. While the performance decline in the middle track might have been caused by short-run implementation problems, the performance decline in the basic track is most likely due to peer effects. An alternative measure of student performance—whether a student is below grade given her age—also indicates detrimental effects of the reform, especially for basic track students. The increase in performance dispersion furthermore suggests that the reform increased inequality of opportunity since student performance is strongly correlated with family background (Hanushek and Woessmann, 2011). With early tracking, more students in basic schools reach only a very low performance level. Further results indicate that the reform had a stronger effect on reading than on math or science. This is in line with Kerr et al. (2012) who find that postponing tracking in Finland increased especially verbal test scores.

Because students still attended school when they participated in PISA, factors other than the school system are unlikely to explain the negative effects on test scores. The reform implied various changes for students in the middle track: they lost lower-performing peers in grade 5 and 6, faced a different curriculum, and were taught by different, partly inexperienced, teachers. As with other major educational reforms, it is impossible to disentangle the relative impact of each of these factors. Although official statistics indicate that observable resources, such as spending per student, class size, and instruction time, changed only little, we cannot rule out that the reform caused (unobservable) disturbances in middle schools in the first years which, in turn, negatively affected student performance. In contrast, the impact of the reform was very different for students in the basic track: these students (only) lost better-performing peers in grade 5 and 6. Therefore, the effects estimated for basic track students likely reflect peer effects only. The findings of this paper therefore indicate that early tracking has negative effects on the performance of low-ability students as they separated from more able peers.

References

Aakvik, A., K. G. Salvanes, and K. Vaage (2010): "Measuring Heterogeneity in the Returns to Education Using an Education Reform," *European Economic Review*, 54, 483–500.

Ammermüller, A. (2005): "Educational Opportunities and the Role of Institutions," ZEW Discussion Paper No. 44.

Angrist, J. and A. Krueger (1999): "Empirical Strategies in Labor Economics," in *Handbook of Labor Economics*, ed. by O. Ashenfelter and D. Card, Elsevier, vol. 3A, 1277–1366.

Angrist, J. and J. Pischke (2009): *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton, NJ: Princeton University Press.

Baik, K.-J. (2011): "Die Einführung der sechsstufigen Realschule im Freistaat Bayern - eine kritische Analyse," Dissertation, University of Regensburg.

Baumert, J., C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, K.-J. Tillmann, and M. Weiss, eds. (2002): *PISA 2000: Die Länder der Bundesrepublik Deutschland im Vergleich*, Opladen: Leske+Budrich.

Baumert, J., M. Becker, M. Neumann, and R. Nikolova (2009): "Frühübergang in ein grundständiges Gymnasium - Übergang in ein privilegiertes Entwicklungsmilieu? Ein Vergleich von Regressionsanalyse und Propensity Score Matching," *Zeitschrift für Erziehungswissenschaft*, 12, 189–215.

Bavarian State Ministry of Education (2008): *Schule und Bildung in Bayern 2008: Zahlen und Fakten*, Munich: Bavarian State Ministry of Education.

Bavarian Statistical Office (1996): *Die allgemeinbildenden Schulen in Bayern Schuljahr 1995/96: Realschulen, Realschulen für Behinderte, Abendrealschulen*, Munich: Bavarian Statistical Office.

Betts, J. (2011): "The Economics of Tracking in Education," in *Handbook of the Economics of Education*, ed. by E. Hanushek, S. Machin, and L. Woessmann, Elsevier, vol. 3, 341–381.

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

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

Bund-Länder-Kommission (2003): *BLK-Bildungsfinanzbericht 2001/2002*, vol. 108-II, Bonn: Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung.

Cascio, E. (2009): "Do Investments in Universal Early Education Pay Off? Long-term Effects of Introducing Kindergartens into Public Schools," NBER Working Paper No. 14951.

Dahl, G. and E. Moretti (2008): "The Demand for Sons," *Review of Economic Studies*, 75, 1085–1120.

Duflo, E., P. Dupas, and M. Kremer (2011): "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya," *American Economic Review*, 101, 1739–74.

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

Epple, D. and R. E. Romano (2011): "Peer Effects in Education: A Survey of the Theory and Evidence," in *Handbook of Social Economics*, ed. by A. B. Jess Benhabib and M. O. Jackson, Elsevier B.V., vol. 1B, chap. 20, 1053–1163.

Federal Statistical Office (2006): *Bildung und Kultur. Allgemeinbildende Schulen: Schuljahr 2005/06. Fachserie 11 Reihe 1*, Wiesbaden: Statistisches Bundesamt.

Galindo-Rueda, F. and A. Vignoles (2007): "The Heterogeneous Effect of Selection in UK Secondary Schools," in *Schools and the Equal Opportunity Problem*, ed. by L. Woessmann and P. Peterson, Cambridge, Mass.: The MIT Press, 103–128.

Graham, B. S. (2008): "Identifying Social Interactions Through Conditional Variance Restrictions," *Econometrica*, 76, 643–660.

Guyon, N., E. Maurin, and S. McNally (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.

Hamermesh, D. and S. Trejo (2000): "The Demand for Hours of Labor: Direct Evidence from California," *Review of Economics and Statistics*, 82, 38–47.

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

——— (2008): "The Role of Cognitive Skills in Economic Development," *Journal of Economic Literature*, 46, 607–668.

——— (2011): "The Economics of International Differences in Educational Achievement," in *Handbook of the Economics of Education*, ed. by E. Hanushek, S. Machin, and L. Woessmann, Elsevier, vol. 3, 89–200.

Jürges, H., K. Schneider, and F. Büchel (2005): "The Effect of Central Exit Examinations on Student Achievement: Quasi-Experimental Evidence from TIMSS Germany," *Journal of the European Economic Association*, 3, 1134–1155.

Kang, C. (2007): "Classroom Peer Effects and Academic Achievement: Quasi-Randomization Evidence from South Korea," *Journal of Urban Economics*, 61, 458–495.

Kerr, S., T. Pekkarinen, and R. Uusitalo (2012): "School Tracking and Development of Cognitive Skills," mimeo; forthcoming in: Journal of Labor Economics.

Köller, O., M. Knigge, and B. Tesch, eds. (2010): *Sprachliche Kompetenzen im Ländervergleich*, Münster: Waxmann.

Kultusministerkonferenz (2007): *Schüler, Klassen, Lehrer und Absolventen der Schulen 1997 bis 2006. Dokumentation Nr. 184*, Bonn: Secretariat of the Standing Conference of the Ministers of Education and Cultural Affairs.

Lüdemann, E. (2011): *Schooling and the Formation of Cognitive and Non-cognitive Outcomes*, no. 39 in ifo Beiträge zur Wirtschaftsforschung, Ifo Institute for Economic Research at the University of Munich.

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

Levin, J. (2001): "For Whom the Reductions Count: A quantile Regression Analysis of Class Size and Peer Effects on Scholastic Achievement," *Empirical Economics*, 26, 221–246.

Lohmar, B. and T. Eckhardt (2010): "The Education System in the Federal Republic of Germany 2008: A Description of the Responsibilities, Structures and Developments in Education Policy for the Exchange of Information in Europe," Bonn: Secretariat of the Standing Conference of the Ministers of Education and Cultural Affairs.

Malamud, O. and C. Pop-Eleches (2010): "General Education versus Vocational Training: Evidence from an Economy in Transition," *Review of Economics and Statistics*, 92, 43–60.

——— (2011): "School Tracking and Access to Higher Education among Disadvantaged Groups," *Journal of Public Economics*, 95, 1538–1549.

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

Meier, V. and G. Schütz (2008): "The Economics of Tracking and Non-Tracking," *Zeitschrift für Betriebswirtschaft*, 78, 23–43.

Mühlenweg, A. (2008): "Educational Effects of Alternative Secondary School Tracking Regimes in Germany," *Schmollers Jahrbuch (Journal of Applied Social Science Studies)*, 128, 351–379.

Murnane, R. J. and B. R. Phillips (1981): "Learning by Doing, Vintage, and Selection: Three Pieces of the Puzzle Relating Teaching Experience and Teaching Performance," *Economics of Education Review*, 1, 453–465.

OECD (2002): *PISA 2000 Technical Report*, Paris: OECD.

——— (2004): *Learning for Tomorrow's World: First Results from PISA 2003*, Paris: OECD.

Oreopoulos, P., M. Page, and A. Stevens (2006): "The Intergenerational Effects of Compulsory Schooling," *Journal of Labor Economics*, 24, 729–760.

Pekkarinen, T., R. Uusitalo, and S. Pekkala (2009): "School Tracking and Intergenerational Income Mobility: Evidence from the Finnish Comprehensive School Reform," *Journal of Public Economics*, 93, 965–973.

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

Prenzel, M., C. Artelt, J. Baumert, W. Blum, M. Hammann, E. Klieme, and R. Pekrun, eds. (2008): *PISA 2006 in Deutschland: Die Kompetenzen der Jugendlichen im dritten Ländervergleich*, Münster: Waxmann.

Prenzel, M., J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, J. Rost, and U. Schiefele, eds. (2005): *PISA 2003: Der zweite Vergleich der Länder in Deutschland: Was wissen und können Jugendliche?*, Münster: Waxmann.

Rivkin, S., E. Hanushek, and J. Kain (2005): "Teachers, Schools, and Academic Achievement," *Econometrica*, 73, 417–458.

Sacerdote, B. (2011): "Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?" in *Handbook of the Economics of Education*, ed. by E. Hanushek, S. Machin, and L. Woessmann, Elsevier, vol. 3, 249–277.

Schneeweis, N. and R. Winter-Ebmer (2007): "Peer Effects in Austrian Schools," *Empirical Economics*, 32, 387–409.

Schütz, G., H. Ursprung, and L. Woessmann (2008): "Education Policy and Equality of Opportunity," *Kyklos*, 61, 279–308.

Slavin, R. E. (1990): "Achievement Effects of Ability Grouping in Secondary Schools: A Best-Evidence Synthesis," *Review of Educational Research*, 60, 471–499.

van Elk, R., M. van der Steeg, and D. Webbink (2011): "Does the Timing of Tracking Affect Higher Education Completion?" *Economics of Education Review*, 30, 1009–1021.

Vigdor, J. and T. Nechyba (2007): "Peer Effects in North Carolina Public Schools," in *Schools and the Equal Opportunity Problem*, ed. by L. Woessmann and P. Peterson, Cambridge, Mass.: MIT Press.

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

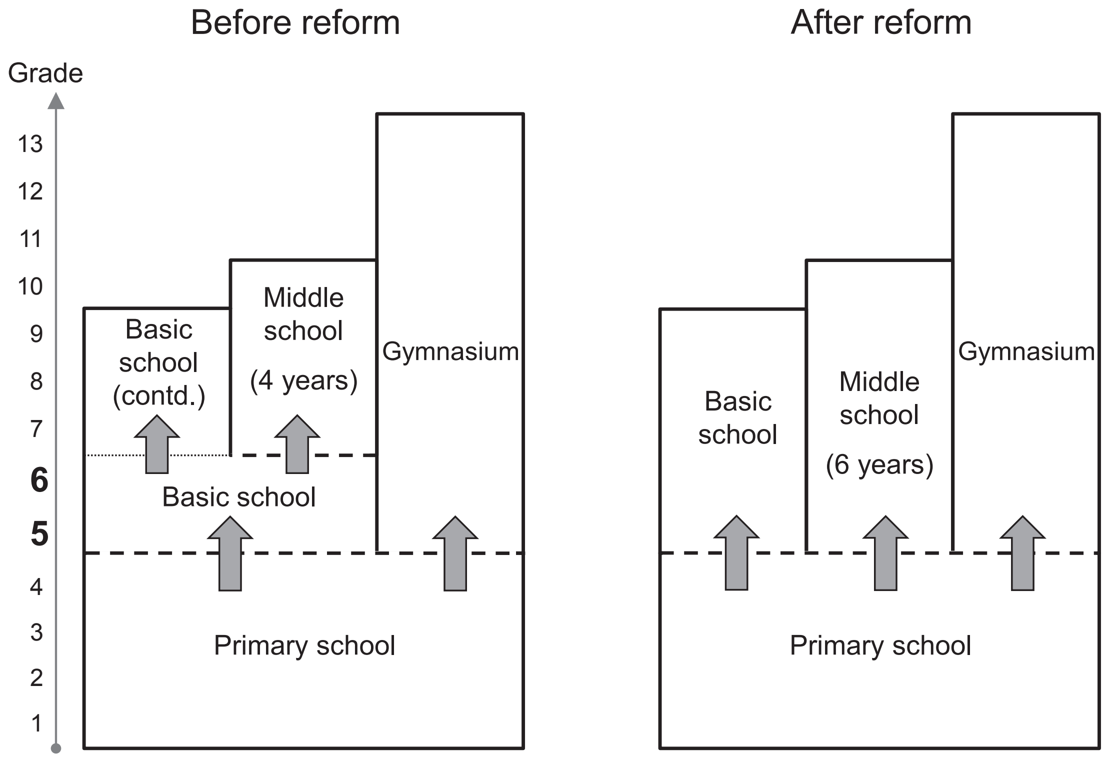
Whitmore, D. (2005): "Resource and Peer Impacts on Girls' Academic Achievement: Evidence from a Randomized Experiment," *American Economic Review*, 95, 199–203.

Woessmann, L. (2010): "Institutional Determinants of School Efficiency and Equity: German States as a Microcosm for OECD Countries," *Jahrbücher für Nationalökonomie und Statistik*, 230, 234–270.

Woessmann, L., E. Lüdemann, G. Schütz, and M. West (2009): *School Accountability, Autonomy, and Choice around the World*, Cheltenham: Edward Elgar.

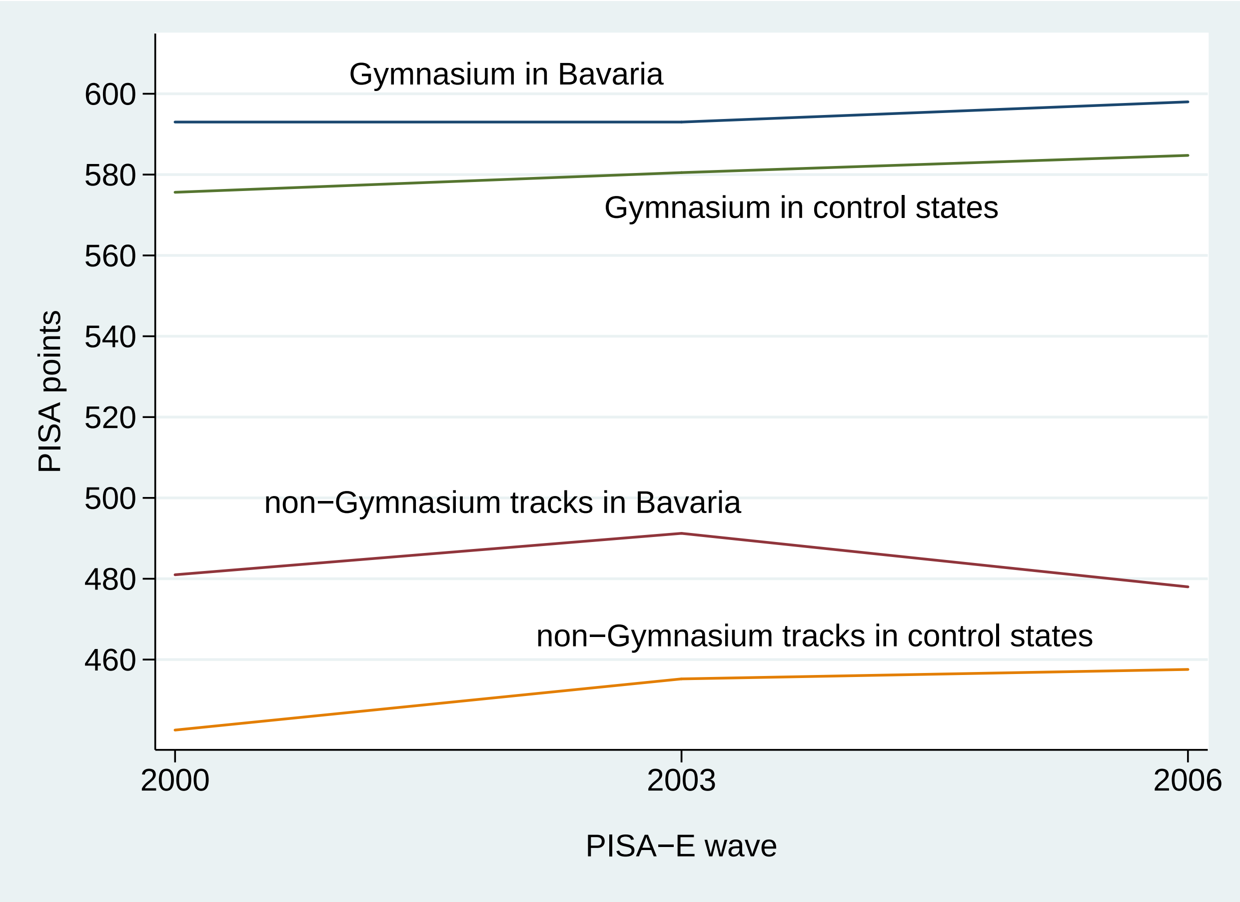
Figures and Tables

Figure 1
School System in Bavaria before and after Reform



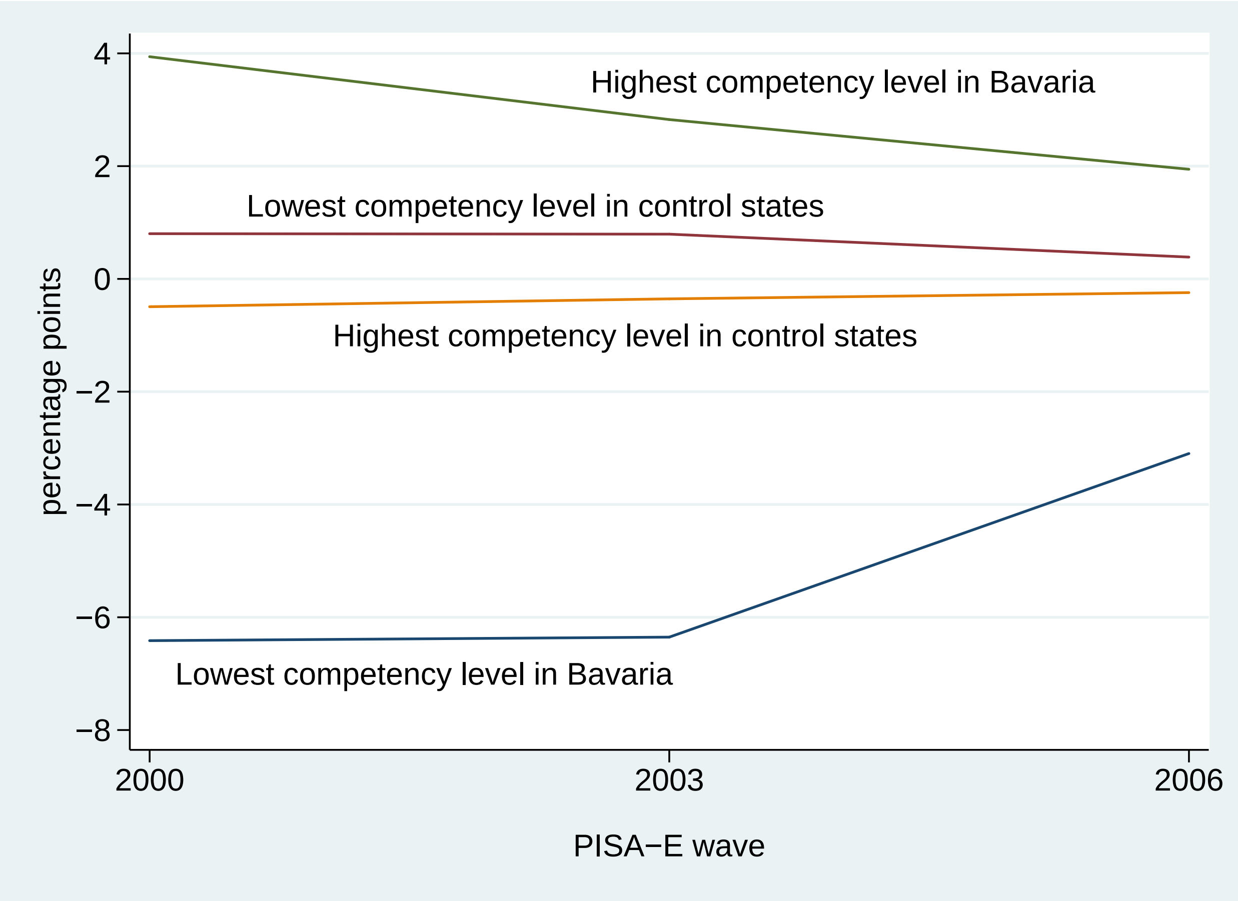
Source: own presentation.

Figure 2
Reading Performance in Non-Gymnasium Tracks and Gymnasium in Bavaria and Control States



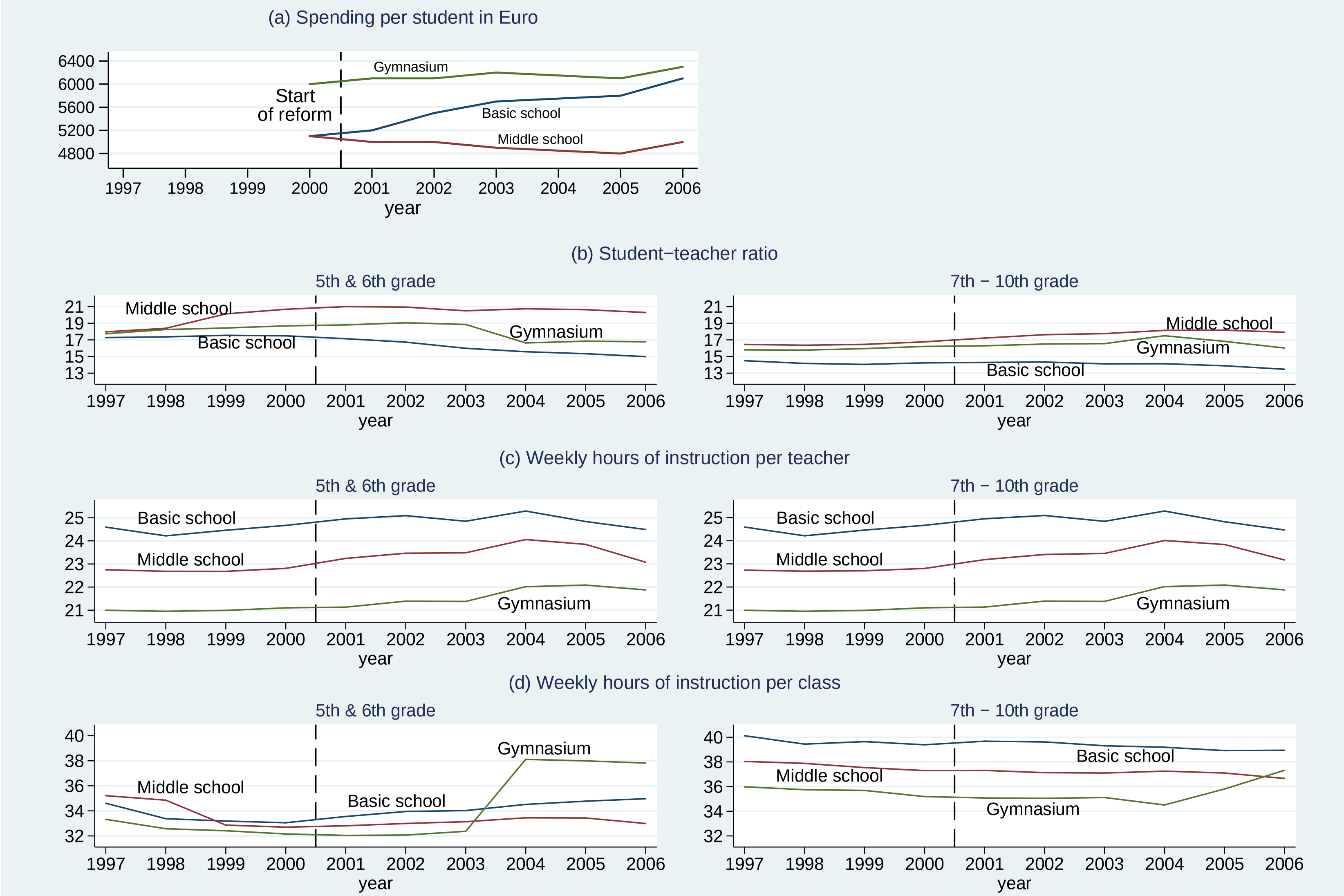
Notes: mean reading performance of PISA-E waves 2000, 2003, and 2006 plotted. Mean reading performance in non-Gymnasium tracks is computed on the basis of the mean reading performance of all 15-year-old students in the state, the mean reading performance of Gymnasium students, and the share of Gymnasium students in the state; the latter two measures refer to ninth graders in PISA-E 2000 and to 15-year-old students in PISA-E 2003 and 2006. See text for definition of control states.
Data source: Baumert et al. (2002); Prenzel et al. (2005, 2008).

Figure 3
Share of Low Performers and Top Performers in Bavaria and Control States



Notes: residuals from regressions of performance measures (pooled across math, reading, and science) on survey year dummies plotted. Performance measures are the share of 15-year-old students achieving at most competency level 1 and the share of students achieving the highest competency level, respectively. See text for definition of control states.
Data source: Baumert et al. (2002); Prenzel et al. (2005, 2008).

Figure 4
Development of Resources in Bavarian School Tracks



Notes: spending per student is based on expenditures on public schools, including spending on personnel, material costs and investments. Comparable figures for previous years are not available. Number of teachers is measured in full-time equivalents, i.e. including full-time and part-time teachers as well as teachers in the preparation phase (Referendariat). Weekly hours of instruction per teacher is the average number of hours of instruction per teacher and week (similar for weekly hours of instruction per class). Source: Bund-Länder-Kommission (2003, Table 3.16) and following issues for spending per student. Kultusministerkonferenz (2007) for other statistics.

Table 1
Descriptive Statistics of Outcomes at State Level

	Bavaria			Control states		
PISA-E wave	2000	2003	2006	2000	2003	2006
	(1)	(2)	(3)	(4)	(5)	(6)
PISA-E performance (across all subjects)						
Mean	511.3	527.0	522.0	487.2	497.2	504.8
Share of students achieving competency level 1 or below	17.5	13.9	14.3	24.7	21.0	17.8
Share of students achieving highest competency level	13.0	9.0	11.0	8.5	5.9	8.8
Percentiles						
p5	330.7	341.7	341.7	302.9	316.3	334.5
p10	373.3	392.0	386.0	347.9	360.1	374.6
p90	639.0	646.7	644.7	617.8	626.1	632.7
p95	667.3	674.0	672.7	649.0	655.0	662.2
Share of 15-year-olds affected by reform in Bavaria(*)						
Basic schools	9.2	26.2	77.4			
Middle schools	8.5	25.4	74.7			
Share of 15-year-olds attending ...(**)						
Basic schools	30.9	33.8	33.1	28.1	27.7	25.9
Middle schools	34.7	32.0	32.1	30.3	31.0	31.2
Gymnasium	28.9	28.9	29.9	28.0	27.9	29.3
Special education schools	4.7	4.7	4.2	5.0	5.2	5.2

Notes: all outcomes are measured at the state level. Outcomes for control states are unweighted averages across all control states; for definition of control states, see text. All performance measures refer to 15-year-old students. *PISA-E performance (across all subjects)* is the simple average across the three subjects math, reading, and science. *Share of 15-year-olds attending* different school types refer to the school years of the PISA tests. School type shares for the control group are based on only those three states that have a school structure similar to Bavaria (Baden-Wuerttemberg, Northrhine-Westphalia, and Schleswig-Holstein). The gap to 100% (especially for control states) is due to students attending comprehensive schools.
Data source: Baumert et al. (2002); Prenzel et al. (2005, 2008); (*) Figures are also based on Bavarian State Ministry of Education (2008); own calculations. (**) Federal Statistical Office (2006) and previous issues.

Table 2
Descriptive Statistics of Outcomes at School-Track Level

	Bavaria				Control states			
PISA-E wave	2003		2006		2003		2006	
	non-Gym	Gym	non-Gym	Gym	non-Gym	Gym	non-Gym	Gym
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PISA-E performance (across all subjects)								
Mean	502.7	605.7	496.3	608.3	470.8	588.9	477.4	594.1
Standard deviation	90.5	62.9	90.3	62.0	86.5	65.0	83.1	65.3
Share achieving competency level 1 or below	17.2	0.5	18.0	0.0	26.3	0.8	21.9	0.4
Share achieving competency level 4 or higher	31.9	80.8	28.4	81.7	18.2	72.4	18.7	74.2
Share of 15-year-olds attending school track	59.4	26.3	60.0	27.5	56.7	28.3	54.3	30.5
Share of migrants	24.6	13.2	23.8	12.2	19.2	12.5	17.9	9.9
PISA-E participation rate	90.7	96.0	89.5	95.0	92.7	95.6	94.0	95.9
Average class size in grades 5-6 (*)	25.9	28.4	26.0	28.5	25.0	27.5	25.1	27.4
Average class size in grades 7-9/10 (*)	25.1	27.0	25.0	27.3	23.5	25.5	23.2	25.6
Share of 15-year-olds in 7th grade	1.2	0.7	1.8	0.2	2.8	0.1	2.2	0.1
Share of 15-year-olds in 8th grade	16.4	9.0	18.5	6.7	19.5	5.7	16.1	3.1
Share of 15-year-olds in 9th grade	68.3	59.6	61.7	54.2	57.4	61.3	57.6	54.8
Share of 15-year-olds in 10th grade	14.1	30.6	18.1	38.6	20.1	32.6	24.1	41.4
Average grade attended	8.95	9.19	8.97	9.32	8.94	9.27	9.04	9.39
Share of students below 9th grade	17.6	9.7	20.3	6.9	22.4	5.8	18.3	3.1

Notes: all outcomes are measured at the school-track*state level. Outcomes for the control states are unweighted averages across all control states; see text for definition of control states. *Non-Gym* outcomes are aggregated across all non-Gymnasium school types within each state (excluding comprehensive, vocational, and special education schools), with each track weighted by the respective share of 15-year-old students; *Gym* refers to outcomes of Gymnasium students in each state. All performance measures refer to 15-year-old students, averaged across the three subjects math, reading, and science. *Average class size in grades 5-6* refers to two school years prior to the PISA-E test, while *average class size in grades 7-9/10* refers to the PISA-E school year. A negligible share of PISA-E students attended grades 6 and 11, respectively.
Data source: Prenzel et al. (2005, 2008); (*) Kultusministerkonferenz (2007).

Table 3
Reform Effects on Student Performance

	Double Differences				Triple Differences			
Dependent variable:	Mean	SD	Share low performers	Share top performers	Mean	SD	Share low performers	Share top performers
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Reform	–13.01***	4.02***	5.10***	–4.05***	–10.46***	5.25***	5.19***	–3.13**
	(2.81)	(0.82)	(1.14)	(1.13)	(2.55)	(1.16)	(1.06)	(1.09)
Bavaria	31.88***	–4.82**	–9.02***	13.73***	16.79***	–2.06*	–0.31**	8.38***
	(4.97)	(1.89)	(2.32)	(1.23)	(2.45)	(0.93)	(0.10)	(1.29)
PISA 2006	6.61**	–3.81***	–4.31***	0.56	5.21***	0.33	–0.38**	1.79***
	(2.81)	(0.82)	(1.14)	(1.13)	(1.28)	(0.49)	(0.12)	(0.49)
Non-Gymnasium tracks					–118.06***	12.62***	25.44***	–54.27***
					(3.52)	(2.16)	(2.27)	(0.66)
Bavaria*PISA 2006					–2.54*	–1.23**	–0.09	–0.92*
					(1.28)	(0.49)	(0.12)	(0.49)
Bavaria*non-Gymnasium					15.09***	–2.76	–8.71***	5.35***
					(3.52)	(2.16)	(2.27)	(0.66)
PISA 2006*non-Gymnasium					1.40	–4.14***	–3.93***	–1.22
					(2.55)	(1.16)	(1.06)	(1.09)
Adj. R-Squared	0.006	0.034	–0.006	0.048	0.770	0.445	0.460	0.884
Observations	90	90	90	90	144	144	144	144

Notes: *double-differences* estimations are based on non-Gymnasium school tracks only; *triple-differences* estimations also include Gymnasium tracks. Dependent variable: mean performance (Columns 1+5), standard deviation (Columns 2+6), share of students achieving at most competency level 1 (Columns 3+7), and share of students achieving competency level 4 or higher (Columns 4+8) within each school track and state. All dependent variables are pooled across the subjects math, reading, and science. *Reform* equals 1 for basic and middle school track in Bavaria in PISA-E wave 2006 and equals 0 otherwise. Ordinary least squares regressions with robust standard errors clustered at the state level in parentheses. Non-Gymnasium school tracks are weighted by the share of students attending the respective track relative to all non-Gymnasium students. For definition of control states, see text. PISA-E waves 2003 and 2006 included. Significance levels: * p<0.10, ** p<0.05,*** p<0.01.

Table 4
Robustness Check: Alternative States as Control Group

	Double Differences				Triple Differences			
Dependent variable:	Mean (1)	SD (2)	Share low performers (3)	Share top performers (4)	Mean (5)	SD (6)	Share low performers (7)	Share top performers (8)
Panel A: Main control group								
Reform	–13.01***	4.02***	5.10***	–4.05***	–10.46***	5.25***	5.19***	–3.13**
	(2.81)	(0.82)	(1.14)	(1.13)	(2.55)	(1.16)	(1.06)	(1.09)
Observations	90	90	90	90	144	144	144	144
Panel B: All states								
Reform	–14.52***	4.00***	5.73***	–4.59***	–10.63***	6.19***	5.86***	–2.99**
	(1.79)	(0.64)	(0.69)	(0.73)	(2.41)	(1.08)	(0.71)	(1.08)
Observations	165	165	165	165	261	261	261	261
Panel C: Best 8 performing states in 2003								
Reform	–11.79***	4.42***	4.32***	–3.94**	–9.60**	5.51***	4.51***	–2.94**
	(2.92)	(0.82)	(0.94)	(1.30)	(2.77)	(1.30)	(0.93)	(1.24)
Observations	78	78	78	78	126	126	126	126
Panel D: Best 4 performing states in 2003								
Reform	–11.59**	4.53*	4.36	–3.83***	–12.48*	5.98	4.68*	–3.80**
	(3.47)	(1.75)	(1.91)	(0.43)	(4.23)	(2.87)	(1.89)	(0.79)
Observations	36	36	36	36	60	60	60	60
Panel E: States with similar school structure								
Reform	–10.26**	3.54**	3.84*	–3.17***	–13.48	5.80*	4.31*	–5.36
	(2.60)	(0.69)	(1.52)	(0.39)	(7.34)	(2.20)	(1.73)	(2.85)
Observations	48	48	48	48	72	72	72	72

Notes: *double-differences* estimations are based on non-Gymnasium school tracks only; *triple-differences* estimations also include Gymnasium tracks. Dependent variable: mean performance (Columns 1+5), standard deviation (Columns 2+6), share of students achieving at most competency level 1 (Columns 3+7), and share of students achieving competency level 4 or higher (Columns 4+8) within each school track and state. All dependent variables are pooled across the subjects math, reading, and science. *Reform* equals 1 for basic and middle school track in Bavaria in PISA-E wave 2006 and equals 0 otherwise. All regressions include the control variables indicated in Table 3. Ordinary least squares regressions with robust standard errors clustered at the state level in parentheses. Non-Gymnasium school tracks are weighted by the share of students attending the respective track relative to all non-Gymnasium students. Control states: panel A (as in Table 3); panel C (D): best-performing 8 (4) states (excl. Bavaria) in PISA 2003; panel E: states with same dominant three school tracks as Bavaria (Baden-Württemberg, Lower Saxony, and Schleswig-Holstein). PISA-E waves 2003 and 2006 included. Significance levels: * p<0.10, ** p<0.05,*** p<0.01.

Table 5
Robustness Check: Adding Control Variables at School-Track Level

Dependent variable:	Mean (1)	SD (2)	Share low performers (3)	Share top performers (4)
Reform	–12.65***	4.99**	5.70***	–4.19**
	(3.19)	(1.63)	(1.18)	(1.38)
Bavaria	17.18***	–1.01	–1.23	7.54***
	(3.86)	(1.13)	(1.44)	(1.49)
PISA 2006	4.62*	0.13	–0.50	1.29
	(2.45)	(1.01)	(0.75)	(1.01)
Non-Gymnasium tracks	–198.44***	12.03*	61.43***	–72.84***
	(16.66)	(6.10)	(5.01)	(6.17)
Bavaria*PISA 2006	–2.72	–0.73	–0.21	–1.28
	(1.82)	(0.95)	(0.53)	(0.85)
Bavaria*non-Gymnasium	38.52***	–1.60	–19.36***	10.60***
	(7.56)	(2.04)	(2.94)	(2.51)
PISA 2006*non-Gymnasium	0.87	–3.42*	–3.10*	–0.65
	(4.43)	(1.77)	(1.64)	(1.74)
Share of 15-year-olds in track	–0.04	0.38*	–0.07	–0.10
	(0.66)	(0.18)	(0.25)	(0.23)
Share of migrants	–0.11	0.24	–0.15	–0.23
	(0.62)	(0.18)	(0.20)	(0.24)
PISA-E participation rate	–0.44	0.37**	–0.08	–0.45*
	(0.71)	(0.12)	(0.29)	(0.24)
Average class size in grades 5-6	–3.95***	0.35	1.18**	–1.48***
	(1.12)	(0.21)	(0.42)	(0.38)
Average class size in grades 7-9/10	2.17	–0.61	–0.09	1.50**
	(1.72)	(0.42)	(0.78)	(0.52)
School type dummies	Yes	Yes	Yes	Yes
Adj. R-Squared	0.947	0.622	0.907	0.960
Observations	144	144	144	144

Notes: dependent variable: mean performance (Col. 1), standard deviation (Col. 2), share of students achieving at most competency level 1 (Col. 3), and share of students achieving competency level 4 or higher (Col. 4) within each school track and state. All dependent variables are pooled across the subjects math, reading, and science. *Reform* equals 1 for basic and middle school track in Bavaria in PISA-E wave 2006 and equals 0 otherwise. *Share of 15-year-olds in track* is the share of all 15-year-old students in a state attending the respective school track. *Average class size in grades 5-6* refers to two school years prior to the PISA test, while *average class size in grades 7-9/10* refers to PISA-E school year. Ordinary least squares regressions with robust standard errors clustered at the state level in parentheses. Non-Gymnasium school tracks are weighted by the share of students attending the respective track relative to all non-Gymnasium students. PISA-E waves 2003 and 2006 included. Significance levels: * $p<0.10$, ** $p<0.05$,*** $p<0.01$.

Table 6
Robustness Check: Pre-Reform Trends and Reform Effects Based on Performance Measures at State Level

Dependent variable:	5th – 95th percentile		10th – 90th percentile		Share low performers – share top performers		Mean reading performance: non-Gym – Gym	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Pre-reform trends								
Bavaria*PISA 2003	–2.96	–3.08	7.04	7.81	1.33	0.98	2.46	2.63
	(5.38)	(4.61)	(4.89)	(4.42)	(1.41)	(1.43)	(3.07)	(2.99)
Bavaria	9.42*	11.12**	4.29	7.85	–11.65***	–11.50***	21.05***	19.36***
	(4.74)	(4.58)	(5.35)	(4.65)	(1.78)	(1.60)	(4.33)	(4.40)
PISA 2003	7.29	11.68**	3.96	9.35**	–0.99	0.64	7.78**	3.27
	(5.38)	(4.19)	(4.89)	(3.56)	(1.41)	(1.64)	(3.07)	(4.16)
Share of 15-year-olds in Gymnasium		–1.97		–1.62		–1.00		2.06
		(1.92)		(2.14)		(0.62)		(1.43)
Share of 15-year-olds in special schools		4.86		6.65**		1.59		–4.97
		(2.67)		(2.66)		(1.80)		(3.31)
Adj. R-Squared	0.016	0.037	0.003	0.078	0.234	0.271	0.403	0.476
Observations	54	54	54	54	54	54	18	18
Panel B: Reform effects								
Bavaria*PISA 2006	–11.19**	–11.11***	–8.40**	–7.53**	5.32***	4.11***	–15.09***	–13.64***
	(3.61)	(3.06)	(2.96)	(2.79)	(0.88)	(1.06)	(2.57)	(3.65)
Bavaria	7.94**	9.85*	7.81*	11.89**	–10.99***	–11.09***	22.28***	20.05***
	(3.08)	(4.52)	(3.86)	(4.85)	(1.52)	(1.52)	(3.09)	(3.43)
PISA 2003	6.96	9.21*	4.74	8.31*	–0.84	1.15	8.06**	2.73
	(4.77)	(4.17)	(4.41)	(3.64)	(1.26)	(1.64)	(2.71)	(3.44)
PISA 2006	18.17***	21.45***	12.27**	16.45**	–7.14***	–2.41	6.01	–4.11
	(4.69)	(4.27)	(4.32)	(5.34)	(1.42)	(2.68)	(3.31)	(6.84)
Share of 15-year-olds in Gymnasium		–0.53		–0.37		–1.28**		2.30*
		(1.46)		(1.68)		(0.50)		(1.10)
Share of 15-year-olds in special schools		2.81		4.78		1.96		–5.99*
		(2.68)		(2.67)		(1.73)		(3.06)
Adj. R-Squared	0.104	0.090	0.062	0.092	0.279	0.340	0.311	0.437
Observations	81	81	81	81	81	81	27	27

Notes: dependent variable: 5th minus 95th percentile (Cols. 1+2); 10th minus 90th percentile (Cols. 3+4); share of students achieving at most competency level 1 minus share of students achieving the highest competency level (Cols. 5+6). All performance outcomes in Cols. (1)-(6) are measured at the state level and pooled across the subjects math, reading, and science. Dependent variable in Cols. (7+8): mean reading performance in non-Gymnasium tracks minus mean reading performance in Gymnasium track. Non-Gymnasium performance is computed on the basis of the performance of all 15-year-old students in the state, the performance of Gymnasium students, and the share of Gymnasium students in the state; the latter two measures refer to ninth graders in PISA-E 2000 and to 15-year-old students in PISA-E 2003 and 2006. Ordinary least squares regressions with robust standard errors clustered at the state level in parentheses. Panel A includes PISA-E waves 2000 and 2003, and Panel B includes PISA-E waves 2000, 2003, and 2006. Significance levels: * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 7
Reform Effects by PISA Subject

	Double Differences				Triple Differences			
Dependent variable:	Mean (1)	SD (2)	Share low performers (3)	Share top performers (4)	Mean (5)	SD (6)	Share low performers (7)	Share top performers (8)
Panel A: Reading								
Reform	–14.87***	2.03	5.67***	–4.54***	–15.62***	4.80*	5.83***	–5.33***
	(4.03)	(1.32)	(1.53)	(1.24)	(3.07)	(2.54)	(1.47)	(1.47)
Observations	30	30	30	30	48	48	48	48
Panel B: Math								
Reform	–11.94***	0.86	3.17**	–5.08***	–6.06*	3.32**	3.51**	–3.45*
	(3.18)	(1.12)	(1.37)	(1.47)	(2.73)	(1.39)	(1.29)	(1.53)
Observations	30	30	30	30	48	48	48	48
Panel C: Science								
Reform	–12.20***	9.19***	6.47***	–2.53*	–9.70**	7.64***	6.23***	–0.62
	(2.76)	(1.27)	(1.22)	(1.14)	(3.20)	(1.90)	(1.21)	(1.14)
Observations	30	30	30	30	48	48	48	48

Notes: *double-differences* estimations are based on non-Gymnasium school tracks only; *triple-differences* estimations also include Gymnasium tracks. Dependent variable: mean performance (Columns 1+5), standard deviation (Columns 2+6), share of students achieving at most competency level 1 (Columns 3+7), and share of students achieving competency level 4 or higher (Columns 4+8) within each school track and state. *Reform* equals 1 for basic and middle school track in Bavaria in PISA-E wave 2006 and equals 0 otherwise. All regressions include the control variables indicated in Table 3. Ordinary least squares regressions with robust standard errors clustered at the state level in parentheses. Non-Gymnasium school tracks are weighted by the share of students attending the respective track relative to all non-Gymnasium students. For definition of control states, see text. PISA-E waves 2003 and 2006 included. Significance levels: * $p<0.10$, ** $p<0.05$,*** $p<0.01$.

Table 8
Reform Effects on Students in Basic and Middle School Track

Dependent variable:	Mean	SD	Share low performers	Share top performers
	(1)	(2)	(3)	(4)
Panel A: Reform effects on students in basic track				
Reform	–14.11	8.97***	6.23	–3.43
	(13.59)	(1.48)	(4.36)	(3.61)
Adj. R-Squared	0.976	0.637	0.955	0.987
Observations	48	48	48	48
Panel B: Reform effects on students in middle track				
Reform	–12.56	3.93	1.38	–8.36*
	(6.40)	(3.34)	(1.46)	(2.91)
Adj. R-Squared	0.909	0.182	0.758	0.932
Observations	48	48	48	48

Notes: dependent variable: mean performance (Column 1), standard deviation (Column 2), share of students achieving at most competency level 1 (Column 3), and share of students achieving competency level 4 or higher (Column 4) within each school track and state. All dependent variables are pooled across the subjects math, reading, and science. *Panel A* includes only the basic school and Gymnasium tracks, while *Panel B* includes only the middle school and Gymnasium tracks.*Reform* equals 1 for basic and middle school track in Bavaria in PISA-E wave 2006 and equals 0 otherwise. All regressions include the control variables of the triple-differences models indicated in Table 3. Ordinary least squares regressions with robust standard errors clustered at the state level in parentheses. Control states are states with the same dominant three school tracks as Bavaria (Baden-Württemberg, Lower Saxony, and Schleswig-Holstein). PISA-E waves 2003 and 2006 included. Significance levels: * $p<0.10$, ** $p<0.05$,*** $p<0.01$.

Table 9
Reform Effects on Probability of Being Below Grade-for-Age

Reform effect on students in:	Both tracks (1)	Basic track (2)	Middle track (3)
Reform	9.26**	13.57***	3.57
	(1.98)	(1.85)	(3.10)
Bavaria	2.40	2.40	2.40
	(3.51)	(4.01)	(4.01)
PISA 2006	–3.60	–3.60	–3.60
	(3.05)	(3.49)	(3.49)
Non-Gymnasium track	19.82***	31.33***	9.43***
	(2.23)	(5.34)	(1.26)
Bavaria*PISA 2006	0.80	0.80	0.80
	(3.05)	(3.49)	(3.49)
Bavaria*non-Gymnasium	–11.92**	–25.63**	1.07
	(2.23)	(5.34)	(1.26)
PISA 2006*non-Gymnasium	–3.80	–5.47*	–1.17
	(1.98)	(1.85)	(3.10)
Adj. R-Squared	0.359	0.682	0.572
Observations	24	16	16

Notes: dependent variable: share of 15-year-old students below 9th grade. *Reform* equals 1 for basic and middle school tracks in Bavaria in 2006; 0 otherwise. Column (2) includes basic school and Gymnasium tracks; Column (3) includes middle school and Gymnasium tracks. Ordinary least squares regressions with robust standard errors clustered at the state level in parentheses. In all columns, control states are the states with the same dominant three school tracks as Bavaria (Baden-Württemberg, Lower Saxony, and Schleswig-Holstein). PISA-E waves 2003 and 2006 included. Significance levels: * p<0.10, ** p<0.05,*** p<0.01.

Table 10
Persistence of Reform Effects on Reading Performance

Included survey years:	2000+2003+2006+2009 (1)	2000+2009 (2)
Reform 2009	–12.03**	–10.80*
	(4.93)	(5.76)
Reform 2006	–15.09***	
	(2.64)	
Bavaria	16.76***	17.38***
	(4.43)	(4.16)
Non-Gymnasium	–133.20***	–133.06***
	(4.24)	(4.40)
Bavaria*Non-Gymnasium	22.28***	21.05***
	(3.17)	(4.40)
PISA 2003	4.74	
	(2.82)	
PISA 2003*Bavaria	–3.64	
	(3.30)	
PISA 2003*non-Gymnasium	8.06**	
	(2.78)	
PISA 2006	9.06**	
	(3.28)	
PISA 2006*Bavaria	–3.51	
	(3.48)	
PISA 2006*non-Gymnasium	6.01	
	(3.40)	
Year 2009	–5.57	–5.50
	(5.13)	(5.13)
Bavaria 2009	–6.88	–7.50
	(5.27)	(5.13)
Non-Gymnasium 2009	21.09***	20.95***
	(5.64)	(5.76)
Adj. R-Squared	0.974	0.966
Observations	72	36

Notes: dependent variable: mean reading performance of non-Gymnasium students and Gymnasium students within each state pooled. In all PISA-E waves, non-Gymnasium reading performance is computed on the basis of the mean reading performance of all 15-year-old students in the state (including students in vocational and special education schools), the mean reading performance in Gymnasium, and the share of Gymnasium students in the state; the latter two measures refer to ninth graders in PISA-E 2000 and to 15-year-old students in PISA-E 2003 and 2006. Non-Gymnasium reading performance for 2009 is based on the mean reading performance of ninth graders in the state (without students in vocational and special education schools), the mean reading performance of ninth graders in Gymnasium, and the share of ninth grade Gymnasium students in the state. *Reform 2009* equals 1 for non-Gymnasium performance in Bavaria in 2009 and equals 0 otherwise (similarly for *Reform 2006*). Ordinary least squares regressions with robust standard errors clustered at the state level in parentheses. PISA-E waves 2000, 2003, and 2006 included. Performance in 2009 is the reading performance from the "Educational Standards" (*Bildungsstandards*) survey. Significance levels: * $p<0.10$, ** $p<0.05$,*** $p<0.01$.

Ifo Working Papers

No. 152 Battisti, M., Individual Wage Growth: The Role of Industry Experience, January 2013.

No. 151 Röpke, L., The Development of Renewable Energies and Supply Security: A Trade-Off Analysis, December 2012.

No. 150 Benz, S., Trading Tasks: A Dynamic Theory of Offshoring, December 2012.

No. 149 Sinn, H.-W. und T. Wollmershäuser, Target-Salden und die deutsche Kapitalbilanz im Zeichen der europäischen Zahlungsbilanzkrise, Dezember 2012.

No. 148 Nagl, W., Better Safe than Sorry? The Effects of Income Risk, Unemployment Risk and the Interaction of these Risks on Wages, November 2012.

No. 147 Mang, C., Online Job Search and Matching Quality, November 2012.

No. 146 Link S., Single-Sex Schooling and Student Performance: Quasi-Experimental Evidence from South Korea, October 2012.

No. 145 Nagl, W., Wage Compensations Due to Risk Aversion and Skewness Affection – German Evidence, October 2012.

No. 144 Triebs, T.P. and S.C. Kumbhakar, Productivity with General Indices of Management and Technical Change, October 2012.

No. 143 Ketterer, J.C., The Impact of Wind Power Generation on the Electricity Price in Germany, October 2012.

No. 142 Triebs, T.P., D.S. Saal, P. Arocena and S.C. Kumbhakar, Estimating Economies of Scale and Scope with Flexible Technology, October 2012.

No. 141 Potrafke, N. und M. Reischmann, Fiscal Equalization Schemes and Fiscal Sustainability, September 2012.

No. 140 Fidrmuc, J. and C. Hainz, The Effect of Banking Regulation on Cross-Border Lending, September 2012.

No. 139 Sala, D. and E. Yalcin, Export Experience of Managers and the Internationalization of Firms, September 2012.

No. 138 Seiler, C., The Data Sets of the LMU-ifo Economics & Business Data Center – A Guide for Researchers, September 2012.

No. 137 Crayen, D., C. Hainz and C. Ströh de Martínez, Remittances, Banking Status and the Usage of Insurance Schemes, September 2012.

No. 136 Crivelli, P. and J. Gröschl, The Impact of Sanitary and Phytosanitary Measures on Market Entry and Trade Flows, August 2012.

No. 135 Slavtchev, V. and S. Wiederhold, Technological Intensity of Government Demand and Innovation, August 2012.

No. 134 Felbermayr, G.J., M. Larch and W. Lechthaler, The Shimer-Puzzle of International Trade: A Quantitative Analysis, August 2012.

No. 133 Beltz, P., S. Link and A. Ostermaier, Incentives for Students: Evidence from Two Natural Experiments, August 2012.

No. 132 Felbermayr, G.J. and I. Reczkowski, International Student Mobility and High-Skilled Migration: The Evidence, July 2012.

No. 131 Sinn, H.-W., Die Europäische Fiskalunion – Gedanken zur Entwicklung der Eurozone, Juli 2012.

No. 130 Felbermayr, G.J., A. Hauptmann and H.-J. Schmerer, International Trade and Collective Bargaining Outcomes. Evidence from German Employer-Employee Data, March 2012.

No. 129 Triebs, T.P. and S.C. Kumbhakar, Management Practice in Production, March 2012.

No. 128 Arent, S., Expectations and Saving Behavior: An Empirical Analysis, March, 2012.

No. 127 Hornung, E., Railroads and Micro-regional Growth in Prussia, March, 2012.

No. 126 Seiler, C., On the Robustness of the Balance Statistics with respect to Nonresponse, March 2012.