3

Selection on Performance and Tracking[1]

Abstract: Tracking is widely used in secondary school systems around the world. Some countries put more emphasis on the use of performance to place students into tracks (e.g. the Netherlands), while in other countries parents have more influence on the track their child will go to (e.g. Germany). This chapter examines whether selection into tracks based on performance has an effect on the relation between tracking and student performance and educational opportunities. Using data from the Program for International Student Assessment for around 185,000 students in 31 countries, different estimation models are compared. The results indicate that a highly differentiated system is best for performance when schools always consider prior performance when deciding on student acceptance. In systems with a few tracks, there is no such impact. Equality of opportunity is best provided for in a system with many tracks when schools always consider prior performance.

1. This chapter is based on joint work with Jaap Dronkers. We would like to thank Lex Borghans, Rolf van der Velden, Hans Heijke, Oliver Marie, and Andreas Ammermueller for a number of useful comments and suggestions. We would also like to thank the seminar participants of the CPB Netherlands Bureau for Economic Policy Analysis, the Journées Louis-André Gérard-Varet 2012, ESPE 2012, the Maastricht University School of Business and Economics education lunch, and the NWO-PROO discussion group for fruitful discussions which improved the chapter.

3.1 Introduction

Tracking students in secondary school can have different effects depending on how students are selected into tracks. In general, tracking students in secondary school could have positive and negative effects on students in secondary school. One of the negative effects is that it can increase the effect of parental background (PB), while a positive effect could be increased learning due to homogenous peer groups. However, both these effects are dependent on how track placement is done: If track placement is done purely on ability levels then the effect of PB is reduced, while tracks are indeed homogenous in the ability composition. If track placement is not done based on ability, but for instance on PB, then increased learning might not happen since the tracks are not homogenous in ability and the effect of PB naturally increases.

The aim of this chapter is to investigate whether using performance to select students into tracks has an effect on the relation between tracking and student performance and between tracking and educational opportunities. Tracking does not vary within education systems. However in most countries school principals are often free in how they select students into tracks. We use data on whether school principals consider prior performance in accepting the students to the school from the Program of International Student Assessment (PISA) 2009 for 31 OECD countries to investigate our question. We find that students who attend schools whose principals consider prior performance in a highly differentiated system have higher test scores and a lower impact of PB than those in a comprehensive system. When comparing different estimation models we find it not likely that our results are driven by the selection which causes better able students to go to schools whose principals consider prior performance in accepting the students to the school.

Some countries have national policies regarding tracking (e.g. Germany, the Netherlands, Austria), while others let schools decide whether and how to implement informal tracking (e.g. United States, Sweden). The manner in which students are placed into tracks also differs widely across countries. For instance, in the Netherlands elementary school students all take an obligatory exit test and combined with the obligatory recommendation of the elementary school teacher on the most suited track, the secondary school accepts students to specific tracks mainly based on that test and that recommendation. In Germany, however, in most northern states parents have the right to persuade schools to accept their child into the highest tracks, while there is no exit test and the teacher's recommendation is only optional (Dollmann, 2011). In these two countries just mentioned, only the Netherlands places students into tracks based on performance (a proxy of ability combining ability with motivation), while it is to be expected that in Germany a strong effect of PB on track placement, and consequently on student performance, is to be found and this is indeed the case (Dustman, 2004).

These two examples show that in the relation of tracking and student performance or educational opportunities, the method of selecting students into tracks could influence outcomes.

To take into account that school principals who consider prior performance possibly do so to be able to accept only the best students to their school, we compare the results of different models which look at between-school between-countries variation, between-schools variation, and between-countries variation.[2] In the first model, we control for the best available internationally comparable track level of the individual students. In the second model, we use only the between-schools variation by using country fixed effects, which alleviates possible country heterogeneity. And thirdly, by using the national percentage of schools which consider prior performance, we look at between-countries variation only, to try to isolate the bias due to selective student acceptance. That tracking has a positive effect on student performance when the method of track placement is taken into account is seen in all three the models. The coefficients are the smallest in the within country model, while they are the largest, although insignificant, in the between countries model. If the relation was purely driven by a selection bias created by school principals, this would not be expected since then most of the variation would be between schools in a given country.

The findings of this chapter indicate that it is important to consider how school principals select students into tracks when looking at the effects of tracking on performance or educational opportunities. Or more general, school characteristics need to be taken into account when analyzing education systems. Perhaps this insight can also explain why the literature finds mixed effects of tracking on both student performance (Fuchs and Woessmann, 2007; Hanushek and Woessmann, 2006; Ariga and Brunello, 2007) and educational opportunities (Brunello and Checchi, 2007; Schuetz, Ursprung, and Woessmann (2008).

The structure of this chapter is as follows: Section 3.2 reviews the tracking literature. Section 3.3 discusses theoretical insights in tracking and how schools can affect tracking. Section 3.4 describes the data used in the analysis, while Section 3.5 lays out

2. Another reason why schools that consider prior performance might not use this for track placement is that the entry of students into the secondary school might not coincide with the start of tracking. However, in most countries the start of secondary and tracking coincide. In our sample of countries used, only 7 of the 21 tracking countries do not start tracking at the start of secondary school. Russia's secondary school starts at age 10, but tracking starts at 14.5 years. Student in Luxembourg start secondary school at age 12, but tracking starts at 13. In Lithuania it is 11 versus 15, in Italy 11 versus 14, and in Israel, Ireland, and Greece 12 versus 15. In the other countries the start of secondary school coincides with the start of tracking. Even for schools in these 7 countries where the start of secondary school and the start of tracking do not coincide, the obtained prior performance information on their students may assists schools later on in selecting students into tracks.

the empirical strategy. Sections 3.6 present the results. Finally, the last section concludes the chapter.

3.2 Literature

Although the separation of students into different tracks in secondary school has a bad reputation, the literature shows conflicting results on the effect of tracking on performance or on educational opportunities.

Hanushek and Woessmann (2006) use a difference-in-differences technique to study the effects of the timing of tracking. They find that early tracking increases distributional inequality and reduces mean performance. Thus, ultimately, nobody gains from tracking: Students in the lower percentiles of educational achievement do lose more than those in the high percentiles relative to students in non-tracking countries, but high-achievement students are also disadvantaged by early tracking.

Fuchs and Woessmann (2007) look at the relation between a number of institutional characteristics and student performance, taking into account school variables. The authors include the age of selection in their robustness checks but find no significant relation between age at first selection and performance.

Ariga and Brunello (2007) argue that the effects of tracking have not fully materialized for youngsters yet, especially in countries that start tracking when students are older. The authors use cross-country differences in tracking length, using variation due to age, curriculum differences, and drop outs, as well as reasons to drop out, as instrumental variables. They find a positive and significant local average treatment effect of time spent tracked on test scores for individuals between 16 years and their mid-20s.

Pekkarinen (2008) and Kerr *et al.* (2013) use the change of the Finnish education system from a system with multiple tracks and selection at the age of 10-11 to a system with one track and selection at the age of 15-16 as a natural experiment. Unfortunately the curriculum was also changed in the reform, clouding the results. Since not all municipalities in Finland implemented the change at the same time, both papers use differences in differences across municipalities and across time. Pekkarinen (2008) looks at the effect of this reform on educational choice but find no effect on the probability of choosing the academic track, the probability of continuing onto tertiary education, or taxable income. Kerr *et al.* (2013) use army entrance test scores to estimate the effect of the system shift on performance and find no effect of the reform on test scores. Analysing subsections of the tests, they find that the policy shift

improved performance on verbal test scores but had no effect on arithmetic or logical reasoning test scores. The authors show that the policy shift improved the test scores most of students with low parental education.

Horn (2009) examines the effectiveness and distributional inequality of institutional characteristics and finds that early selection is related to higher educational inequality. Both early selection and a high number of tracks in a system are related to lower effectiveness. These relations are in line with the causal effects found by Hanushek and Woessmann (2006).

Van Elk *et al.* (2011) use within-country variations in the age of selection into tracks to look at the performance of low-ability Dutch students. The authors control for delayed selection into tracking by instrumenting the age of selection into tracks by the regional variation in school availability. Early tracking has a negative effect for low-ability students on secondary school completion rates when controlled for by socioeconomic status and test scores, while the inclusion of the low-ability students in the track for high-ability students, resulting in a lower-ability peer group for high-ability students, has no effect on the high-ability students' completion rates.

In general, negative effects of tracking on performance are documented in the literature. However, these results are contradicted by certain positive effects or, most often, insignificant effects, especially when school factors are considered.

Regarding the possible trade-off between efficiency and equality due to tracking, this chapter defines inequality as the existence of PB influence on student performance. Ammermueller (2005) finds that the more tracks available to students, the higher the positive effect of PB on performance growth from primary to secondary school. Walldinger (2006) finds that educational opportunities are lower in early tracking countries. However, this effect disappears when the effect of PB in elementary schools is controlled for. Thus early tracking does not decrease educational opportunities.

Brunello and Checchi (2007) find that the length of tracking does not reinforce the effect of PB on later (literacy) performance and may even diminish it. However, since they measure performance for individuals aged 17 and above, any reinforcing effect of tracking may have disappeared over time. In addition, the authors find that regarding educational attainment and later earnings, tracking does decrease educational opportunities.

Schuetz *et al.* (2008) find that the effect of PB on performance is stronger when countries track early and that these countries do not perform better, on average. In other words, there is no trade-off between efficiency and equality: Tracking affects both negatively.

Summarizing the above literature, it seems that the PB effect is stronger in countries that track early or that have multiple tracks and thus there is less equality of opportunity. However, this reinforcing effect of tracking is by some authors explained by a larger effect of PB in elementary school in these countries or it seems to disappear over time.

3.3 Tracking

Tracking, as the formal practice of separating students into distinct educational programs is called, is a widely used education systems characteristics in secondary school systems around the world. Between the countries that track, the implementation of tracking can differ on the number of tracks available to 15 year old students (most frequently ranging from two to five) and on the age of selection into tracks (from 10 to 16) or alternatively on the length of the school system that is tracked. In the countries that have formalized tracking, tracks are institutionalized in different school types and often located in different buildings and administrative units, while in countries without formal tracking non-institutionalized tracking can occur within schools, either by ability grouping (different classes within schools) or seating (different curricula within classes).[3]

In some sense, tracking is a form of imposing peer homogeneity on students, while it also offers students a more targeted curriculum. Tracking separates students into groups based on observed ability and thus each track consists of a more or less homogeneous student population, depending on the number of tracks available. However, the effect of imposing peer homogeneity is not theoretically straightforward. First, by removing the better-performing students from the lower tracks, the mean performance of the lower tracks decreases and the resulting lower level of peer performance can harm the performance of the lower-ability students. In contrast, the performance of the high-ability students, who are now surrounded by more high-ability peers, improves with positive spillovers. If peer effects work through mean performance, as described above, we would expect to find no country effect of tracking since the positive and negative peer effects on performance cancel each other out. Tracking can then be seen as a zero sum game, although it may alter the distribution of performance. On the other hand, when peer effects are non-linear, tracking can have nationwide effects. The theoretical models of non-linear peer effects support either positive or negative effects of tracking (See Sacerdote, 2011). For

3. In countries with non-institutionalized forms of tracking (e.g. ability grouping or seating), placement in the higher performing group can also affect student performance and the effect of PB on performance. However, this non-institutionalized tracking lies outside the scope of this chapter.

instance, when especially high-ability students benefit from high-ability peers, tracking has a positive effect; when especially low-ability students benefit from high-ability peers, tracking has a negative effect.

Second, peer homogeneity in tracks can be good for both high- and low-ability students when teachers target their teaching to the average performance of the class. In highly differentiated systems (i.e., systems with a large number of tracks), the top and bottom pupils are closer to the average performance level and can thus benefit from peer homogeneity when it allows them to learn more from the teacher.

In addition to imposing peer homogeneity on students, tracking subjects students to specialized curricula, which means that students in different tracks are taught different things at different levels of difficulty. As long as the specialized curricula are optimally designed for the average characteristics of the students in the track, they should increase performance.

Overall, we would expect a positive effect of tracking due to improved teaching strategies and adjusted curricula, while the effect of tracking due to more homogeneous peer groups is theoretically uncertain. Unfortunately disentangling the different effects from peers, adjusted curricula and adjusted teacher strategies is not possible. We will therefore look at all three effects combined.

The arguments in the paragraphs above assume track placement is based on (prior) performance, which can be considered a proxy for ability. However, as described in the introduction of this chapter, track placement is not always based on performance. Dustmann (2004) shows for Germany that PB is a strong predictor of track choice and that there is strong intergenerational immobility in track choice. When parents are free to send their child to any of the available tracks, they may choose the track they attended and/or the track they are familiar with. Schools could also select students based on artistic performance or on other aspects as religion or residential area. When this happens tracks are no longer homogeneous in performance. We expect that schools that use an objective measure to place students into tracks have a greater performance homogeneity in tracks and this induces the expected positive effects of tracking as described above, and it lowers the influence of parents and thus ensure more equal opportunities.

3.3.1 Tracking and schools

Schools selecting students into tracks without an awareness of their observed abilities or basing their selection on non-academic criteria can severely limit the expected positive effects of tracking. Basing track placement on non-academic grounds will lower the envisioned class homogeneity of tracking. This problem can be mitigated by ensuring that the selection of students into tracks is always based (primarily) on ability. A proxy for ability can be prior performance, although it still is an imperfect measure of student ability. Schools that have

information on prior performance when they place students into tracks, may be better able to ensure homogeneous classes than schools that do not have information on prior performance.

3.3.2 Tracking and equality of opportunity

Naturally, we do not assume that when schools consider prior performance parents have no influence on school choice or performance. Parents always influence a child's ability, directly through genes and/or indirectly through the environment they create for their children. However, we assume that as long as observed ability limit non-ability-related parental influence on track choice, the effect of PB is decreased.

3.4 Data

The student- and school-level data used in this chapter are from the 2009 wave of the Program for International Student Assessment (PISA 2009), executed by the Organization for Economic Co-operation and Development (OECD). These data include internationally comparable test scores in reading, mathematics, and science, and information on students and schools. The country-level data are from the OECD and the World Bank.

The first wave of PISA was presented in 2000 and, since then, every three years a representative sample of students from all participating countries is subjected to tests on reading, mathematics, and science. The test results are standardized to a mean of 500 and a standard deviation of 100 on the PISA reading test in 2000 for the OECD countries.[4] In addition to the tests, the students and school principals are surveyed. A total of 75 countries participated in PISA 2009. Since these countries are very diverse in their economic development, this chapter uses a selection of comparable western countries to limit country heterogeneity. All 31 countries in this analysis have a gross domestic product (GDP) per capita above the minimum of the OECD and available data on national tracking policies.[5] These limitations on the sample are imposed to ensure that country differences do not drive the results, although also country fixed effects models are used to further take country difference into account.

A representative sample from each participating country is obtained by the OECD in two stages: First, schools are selected and, then, students of target age are selected within these schools. The target age is set to a range of 15 years and three months to 16 years and two months (OECD, 2010). Since not all selected schools and students were willing to participate and some schools and students were oversampled to obtain extra information on these groups, the OECD provides weights to ensure sample representation. The student sample in this analysis consists of all native students in

4. The OECD provided five plausible values estimated using item response theory for the test scores since students do not receive all questions. In this chapter only one plausible value is used.
5. Countries such as Australia, Canada, France, and the United Kingdom, are not included since these have missing country-level data or too much missing school-level data. Mexico is excluded since many of its student and school variables are outliers.

(pre-) vocational or general education who were in schools where more than five students participated in PISA 2009.[6] This amounts to 187,768 students in 7,489 schools in 31 countries.

3.4.1 Tracking and selection by schools

This chapter defines tracking as the separation of students into tracks that differ in academic orientation and curricula. The extent of tracking is measured using the 'number of school types or distinct educational programs available to 15-year-olds', taken from Table 5.2 of OECD (2007), shown in the first column of Table 1. This measure of tracking is different from those used in some other papers. For instance, Fuchs and Woessmann (2007) and Van Elk *et al.* (2011) use the age at which a student is first selected into a track as a measure of tracking, while Hanushek and Woessmann (2006) and Schuetz *et al.* (2008) divide countries into early versus late trackers. The advantage of using the number of tracks available to students is that it better captures the possible effects due to the degree of class homogeneity. Its disadvantage, however, is that it does not take into account the amount of time students spend in the tracks. Section 6.4 addresses this issue.

Whether schools use prior performance to select students into tracks is proxied by an index based on a PISA 2009 question to school principals on how often consideration was given to a student's record of academic performance (including placements tests) and to feeder school recommendations in admitting the student to the school. Schools are divided into three categories: schools where neither of the two factors is considered, schools where at least one of these factors is sometimes used to decide acceptance, and schools where at least one of the two factors is always considered. In this chapter whether schools consider prior performance on acceptance to the school is used synonymous to whether schools have performance criteria for track placement of students. This assumes that secondary schools that have prior performance information use this to decide on the track placement of students.

Table 1 gives an overview of the percentage of schools per country that consider prior performance. Countries differ substantially on the percentage of schools that do or do not consider prior performance: from 79.6 percent of students in schools in Spain that never consider them to 93.5 percent that always consider them in Croatia. The type of school that considers prior performance also differs across countries: for instance, in the Czech Republic and Hungary schools that do are often upper secondary schools, while in Austria and Poland it are mostly schools that students with a high PB attend. In general, village schools or schools without neighboring schools are less likely to consider prior performance; schools that service more girls, vocational students or

6. This chapter includes only native students because previous literature has shown that native and migrant students react differently to system characteristics (e.g., Dronkers *et al*. 2012).

students in upper secondary school, higher PB schools, and schools which have more teacher shortages are more likely to consider prior performance.

As can be seen from Table 1, in almost every country there are schools in all three categories, that is, schools that never, sometimes, and always consider prior performance. Thus, even in systems with a high number of tracks, some schools do not use prior academic performance or teacher recommendations to decide school admittance. In the seven countries with four tracks, only 52.1 percent of the students go to schools that always consider prior performance; in the two countries with five tracks, 21 percent of students go to schools that never consider prior performance. Maybe more surprising, even in a comprehensive system, some schools consider prior performance when accepting students: In the 10 countries with only one track, 45 percent of students went to schools that consider prior academic performance. Given the observable characteristics, there seems to be no reason why school policies deviate from implied nationwide system characteristics, that is, why some schools in comprehensive systems select students and why some schools in highly differentiated systems do not. However, also in countries without tracking, non-institutionalized forms of tracking (ability grouping or seating) exist, which could induce schools to select students based on prior performance. The mechanisms in those countries could work in similar ways as described here. For instance, Lucas (1999) has shown that various methods of placement in non-institutionalized tracks in schools in the US can produce variation in the strength of the effect of early ability and PB on student performance.

Table 1: Tracking and selection by schools

Country	Number of school tracks available to 15-year-olds	Percentage of students in schools that consider prior performance for student acceptance		
		Never	*Sometimes*	*Always*
Argentina	3	47.4	31.3	21.3
Austria	4	21.3	18.2	60.5
Belgium	4	38.3	45.1	16.6
Chile	2	17.4	43.1	39.5
Croatia	3	0.0	6.5	93.5
Czech Republic	5	23.7	27.0	49.3
Denmark	1	50.6	44.6	4.8
Estonia	1	13.5	56.6	30.0
Finland	1	69.2	27.0	3.8
Germany	4	10.3	15.7	74.1
Greece	2	55.0	38.8	6.2
Hungary	3	5.4	4.8	89.8
Iceland	1	69.8	27.4	2.8
Ireland	4	43.7	36.8	19.6
Israel	2	11.1	35.1	53.8
Italy	3	28.6	30.1	41.3
Latvia	3	44.6	30.4	25.0
Lithuania	3	38.9	48.1	13.0
Luxembourg	4	0.5	57.5	42.0
Netherlands	4	0.6	11.7	87.7
New Zealand	1	35.3	39.6	25.2
Norway	1	75.8	18.3	5.8
Poland	1	34.9	49.0	16.1
Portugal	3	71.1	27.7	1.2
Russian Federation	3	36.3	40.8	22.9
Slovak Republic	5	18.3	18.7	63.1
Slovenia	3	23.0	46.6	30.4
Spain	1	79.6	18.0	2.5
Sweden	1	78.6	18.4	3.1
Switzerland	4	22.6	13.1	64.3
United States	1	45.8	26.4	27.8
One track	1 (61,535 students)	55.3 (37,794)	32.5 (17,731)	12.2 (6,010)
Two tracks	2 (14,336 students)	27.8 (3,748)	39.0 (5,634)	33.2 (4,954)
Three tracks	3 (66,376 students)	32.8 (21,322)	29.6 (19,943)	37.6 (25,111)
Four tracks	4 (35,217 students)	19.6 (8,326)	28.3 (9,184)	52.1 (17,707)
Five tracks	5 (10,304 students)	21.0 (1,902)	22.8 (2,375)	56.2 (6,027)

Source: OECD (2007) (first column) and PISA 2009 (second till fourth column). Whether schools consider prior performance is obtained from a questionnaire filled out by the school principal (see text for more details).

3.4.2 Control variables

All student variables are collected through student surveys. This study controls for gender, age, PB, a dummy for (pre-) vocational education as opposed to general education, and a dummy for upper secondary school as opposed to lower secondary education. This division is based on the International Standard Classification of Education (ISCED) level, which provides internationally comparable standards for

comparing education levels. PB is captured by a widely used index composed by the OECD that describes the student's economic, social, and cultural status.

Table 2: Descriptive statistics

Variable	Mean	Std. dev.	Min.	Max.
PISA reading score	493.11	93.52	1.97	817.42
PISA mathematics score	495.33	94.61	42.85	901.86
PISA science score	502.88	94.98	10.95	863.24
Parental background	0.07	0.94	-5.34	3.41
Gender	0.50	0.50	0.00	1.00
Age	15.77	0.29	15.25	16.33
Student in (pre-)vocational education	0.16	0.37	0.00	1.00
Student in upper secondary school	0.47	0.50	0.00	1.00
Mean school PB	0.04	0.56	-2.55	1.61
SD of school PB	0.77	0.16	0.17	1.58
Percentage in class for whom the first language is not the test language	4.90	1.25	1.00	6.00
Student-teacher ratio	12.72	6.45	0.39	264.33
Teacher shortage	-0.14	0.88	-1.02	3.34
Hinder by a shortage instruct material	1.82	0.83	1.00	4.00
School responsible for curriculum and assessment	-0.14	0.94	-1.37	1.36
Index of achievement tracked by authority	0.34	0.47	0.00	1.00
School type	1.18	0.44	1.00	3.00
School competition	1.67	0.85	1.00	3.00
School location	2.88	1.09	1.00	5.00
School size	659.48	491.29	10.00	6,850.00
Ability grouping	1.78	0.67	1.00	3.00
School considers prior performance	1.98	0.83	1.00	3.00
Number of tracks	1.61	1.31	0.00	4.00
GDP per capita, 2008	$28,856.03	$12,410.14	$13,275.68	$73,349.64

Source: PISA (2009), OECD (2007: Number of tracks available to 15-year old students) and World Bank (2012: GDP per capita 2008).

The school-level variables are collected through a survey completed by the school principals. School composition is captured by the school average and standard deviation of the PB of all the students per school and by the percentage of them who speak a language other than the test language at home.[7] School inputs are captured by the student-teacher ratio, an index of possible teacher shortages, dummy variables indicating whether the school is hindered by a shortage in instruction material, and an index indicating whether the school is responsible for the curriculum and assessment. Other school characteristics indicate whether school achievement is tracked by an educational authority; whether the school is a public, private government-dependent, or private government-independent school; whether the school has to compete with other schools for students; the school location; school size; and whether the school

7. For the school average and standard deviation of the PB, all the students in the sample schools are used, both natives and immigrants.

uses ability grouping[8]. We also control for the GDP per capita, which is for 2008 from the World Bank (2012). Table 2 provides descriptive statistics for all variables.

3.5 Empirical strategy

The aim of this chapter is to investigate whether using performance to select students into tracks has an effect on the relation between tracking and student performance and educational opportunities. To answer this question we make use of three models: a between schools between countries model depicted in equation (1), a within country model depicted in equation (2), and a between countries model depicted in equation (3).

$$Test_{isc} = \beta_0 + Student_{isc}\beta_1 + School_{sc}\beta_2 + EntrReq_{sc}\beta_3 + \#ofTracks_c\beta_4 + EntrReq_{sc} * \#ofTracks_c'\beta_5 + GDPpc_c\beta_6 + u_c + u_{sc} + \epsilon_{isc} \quad (1)$$

$$Test_{isc} = \delta_0 + Student_{isc}\delta_1 + School_{sc}\delta_2 + EntrReq_{sc}\delta_3 + C_c\delta_4 + EntrReq_{sc} * \#ofTracks_c'\delta_5 + w_{sc} + \vartheta_{isc} \quad (2)$$

$$Test_{isc} = \theta_0 + Student_{isc}\theta_1 + School_{sc}\theta_2 + Nat\ \%\ EntrReq_c\theta_3 + \#ofTracks_c\theta_4 + Nat\ \%\ EntrReq_c * \#ofTracks_c'\theta_5 + GDPpc_c\theta_6 + z_c + z_{sc} + \mu_{isc} \quad (3)$$

In these equations, $Test_{isc}$ is the individual PISA test score in reading, mathematics, or science of student i in school s in country c. $Student_{isc}$is a matrix of student variables, $School_{sc}$ is a matrix of school variables, while $EntrReq_{sc}$is a matrix containing the dummies on whether schools consider prior performance. $\#ofTracks_c$ is a vector containing the number of tracks available to students in each country. As compared to the main model in equation (1), model (2) adds country fixed effects, C_c, while in model (3) we use the national percentage of school whose principals sometimes or always consider prior performance on accepting the student to the school, $Nat\ \%\ EntrReq_c$, as opposed to the school level variable. $GDPpc_c$ is a vector containing GDP per capita. We use random effect models, which are estimated using maximum likelihood, to take into account error terms for countries, schools, and individuals. Separate error terms for countries, schools, and individuals are necessary, since students are nested within schools within countries. If we would ignore the nested data structure, we would implicitly assume that all observations are independent from each other. This would lead to a downwards bias of the standard errors. Weighting is used to ensure representative samples.[9] The control variables are like discussed in Section 3.4.2.

8. Excluding the control for ability grouping leaves the results unchanged. These results are not shown, but are available from the author upon request.
9. The sum of the weights per country are equalised over all countries such that each country has an equal weight in the estimation. All statistics and estimations are executed using weights, unless otherwise specified.

To study the hypotheses on student performance, the main focus lies on the interaction between tracking and whether schools consider prior performance, $EntrReq_{sc} * \#ofTracks'_c$. Our hypothesis is that this interaction will be positive: School principals that consider prior performance in selecting students into tracks are more likely to achieve positive tracking effects due to more homogenous classes. To look at whether the effect of PB is lower when track placement is done based on prior performance we use equation (4) and (5). Model (4) contains an interaction between PB and the number of tracks in a country and will show whether in countries with more tracks there is a larger effect of PB. Model (5) then adds the interaction between number of tracks, PB, and whether school principals consider prior performance. This triple interaction will show whether the effect of PB is lowered when track placement is done based on ability. The PB of students is included in all models in the vector $Student_{isc}$.

$$Test_{isc} = \gamma_0 + Student_{isc}\gamma_1 + School_{sc}\gamma_2 + EntrReq_{sc}\gamma_3 + \#ofTracks_c\gamma_4 + PB_{isc} * \#ofTracks'_c\gamma_5 + GDPpc_c\gamma_6 + v_c + v_{sc} + \varepsilon_{isc} \quad (4)$$

$$Test_{isc} = \rho_0 + Student_{isc}\rho_1 + School_{sc}\rho_2 + EntrReq_{sc}\rho_3 + \#ofTracks_c\rho_4 + PB_{isc} * \#ofTracks'_c\rho_5 + PB_{isc} * \#ofTracks'_c * EntrReq_{sc}\rho_6 + GDPpc_c\rho_7 + t_c + t_{sc} + \pi_{isc} \quad (5)$$

Almost all student- and school-level variables have some missing observations. Although most variables have below 3 percent missing values, deleting all observations with missing variables would lead to a drop in observations from around 185,000 to around 130,000. To delete observations with missing values would be to implicitly assume that the missing values are missing at random, which is a questionable assumption. Another reason for not deleting all observations is that it leads to distorted weighting. Therefore, the missing values in the sample are replaced by group averages.[10] To control for possible bias introduced by the method for replacing missing values, imputation dummies and imputation interactions are used in all models.[11]

3.6 Results

First, we replicate the standard cross country analysis of the effect of tracking on student performance. Then we turn to investigating whether using performance to select students into tracks has an effect on the relation between tracking and student performance. We first present the results on the between countries between schools

10. The student variables are replaced by the average value of the students in the same school, the school variables are replaced by the country average. Country variables are never missing.
11. The results are robust to the exclusion of the imputation variable interaction terms and to the exclusion of the imputation dummies.

models and subsequently show the between countries and the within country results. And thirdly, we focus on whether using performance to select students into tracks has an effect on the relation between tracking and educational opportunities.

3.6.1 Direct relation between tracking and student performance

The analysis starts by investigating whether the number of tracks has a direct and significant relation with student performance. Since we include a wide variety of school background variables, which may capture part of the tracking effect, we do not expect a large coefficient for tracking. Table 3 confirms our expectations: The association between the number of tracks and performance is insignificant, while it is negative for reading and positive for mathematics and science. For all three test subjects, the relation between schools that sometimes consider prior performance and student performance is negative, while for schools that always consider prior performance this relation is positive (and significant).

All the control variables, which are excluded from the table, have the expected sign. PB has a strong and positive relation with student performance. Females do better in reading and males are better in mathematics and science. Older children do somewhat better in reading and science but not in mathematics. Students in vocational education perform worse than students in general education and students in upper secondary education perform better than those in lower secondary school. The school average in PB has a positive relation with student performance. The negative relation between the percentage of students with a different first language and performance only becomes significant when more than 20 percent of the students speak a different language at home (for mathematics and science) or more than 40 percent do so (for reading). Shortages in teachers and materials have a negative relation with student performance and students in private schools that do not rely on the government for financial support perform considerably worse. Finally, school size, location, and whether it uses ability grouping are also important factors. Students in larger schools or in smaller agglomerations perform better and students in schools that use ability grouping perform worse. The results of the full models are available upon request.

Table 3: The effect of tracking on performance

Dependent variable:	(1) Reading	(2) Mathematics	(3) Science
School considers prior performance			
Sometimes	-2.02	-1.37	-1.13
	(1.95)	(1.89)	(2.06)
Always	6.34*	5.51*	7.45**
	(3.27)	(3.07)	(3.79)
Number of tracks (0-4)	-2.56	1.94	0.75
	(2.55)	(3.67)	(3.30)
(pseudo-)LL	-43,761	-43,739	-44,094

# of students	187,768	187,768	187,768
# of countries	31	31	31

Notes: The table presents coefficients from random effects models (standard errors in parenthesis) on the relation between student performance and the number of tracks in a country, controlling for whether or not schools consider prior performance when selecting students. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The control variables are as described in the text. The models include imputation dummies and imputation variable interaction terms.

3.6.2 Tracking and performance

To test whether considering prior performance to select students into tracks has an effect on the relation between tracking and student performance, we include interactions between whether schools consider prior performance and the number of tracks, as described by equation (1). The results can be seen in Table 4. More tracks in an education system are positive for student's performance if students attend schools where the principals consider prior performance on accepting the student to the school. For reading there is still a significant negative effect of more tracks (-5.74**), but this is compensated when schools always consider prior performance and there are 4 or 5 tracks to choose from. Since for mathematics and science no significant negative coefficient of tracking exists (-0.88 and -2.89), more tracks are even better for students when they attend schools which always considers prior performance.

Table 4: The effect of tracking on performance and the influence of selection on performance

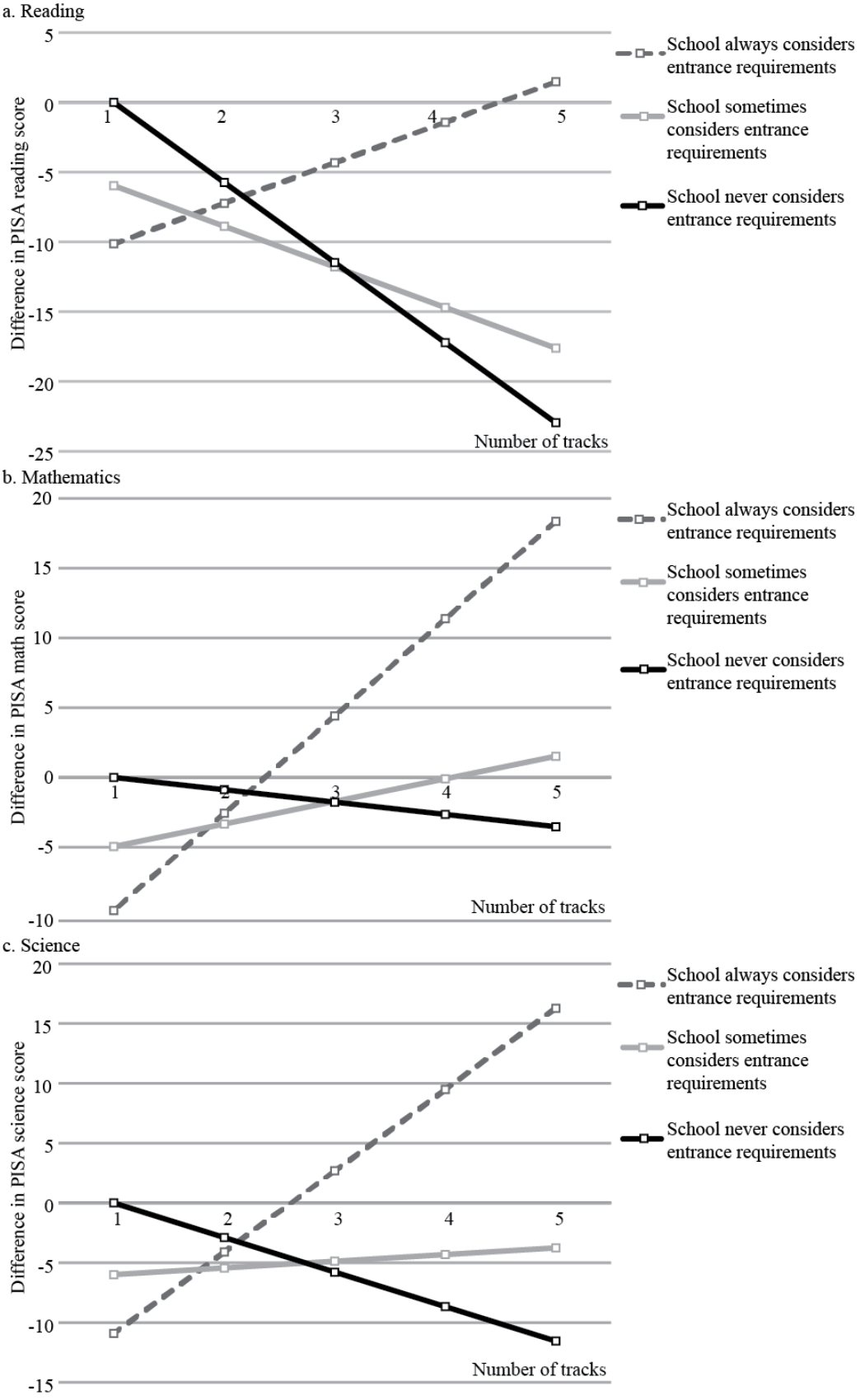
Dependent variable	(1) Reading	(2) Mathematics	(3) Science
School considers prior performance			
Sometimes	-5.97**	-4.95**	-5.99**
	(2.54)	(2.40)	(2.44)
Always	-10.13**	-9.54**	-10.90**
	(4.11)	(4.25)	(5.06)
Number of tracks (0-4)	-5.74**	-0.88	-2.89
	(2.70)	(3.75)	(3.31)
Sometimes*Number of tracks	2.83**	2.50**	3.45***
	(1.15)	(1.16)	(1.08)
Always*Number of tracks	8.64***	7.85***	9.68***
	(1.82)	(1.90)	(2.04)
(pseudo)LL	-43,759	-43,738	-44,092
# of students	187,768	187,768	187,768
# of countries	31	31	31

Notes: The table presents coefficients from random effects models (standard errors in parenthesis) on the relation between student performance and whether or not schools consider prior performance when selecting students and the number of tracks in a country. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The control variables are as described in the text. The models include imputation dummies and imputation variable interaction terms.

To facilitate the interpretation of the interaction terms, Figure 1 shows the combined coefficients for the three models. Figure 1 show for each combination of number of tracks and whether schools consider prior performance what the relation between the two and student performance is, relative to students in a system with only one track in schools that never consider prior performance. Looking at the figures, one sees the same trend for all three subjects (reading, mathematics, and science): Schools in multiple track systems do better when they consider prior performance more often, while schools in comprehensive systems perform better when they do not consider prior performance. When only the significant differences in the graphs are considered, it becomes clear that for two or more tracks whether schools consider prior performance only changes the results when schools always consider prior performance. The coefficients for Never and Sometimes are not significantly different from each other when the number of tracks is two or more.[12]

12. Except for science, where the combined coefficients for the Never and Sometimes are significantly different in a system with five tracks.

Figure 1: Differences in student performance for students in different education systems that attend schools that do or do not consider prior performance



Notes: The figures depict the differences in PISA test scores relative to students in a comprehensive system in schools that never consider prior performance.
The models in Table 4 use the number of tracks in a country as a continuous variable ranging from zero to four. Appendix H contains non-linear models which include the number of tracks in a country as three dummy variables: no tracks (only one track available to 15-year-olds), a few tracks (two or three tracks available), and a large number of tracks (four or five tracks). The results are qualitatively similar.

From Table 4 it can be concluded that students in a system with a high number of tracks do better when their school always considers prior performance, while for students in a system with a low number of tracks (two or three) whether schools consider prior performance do not seem to matter. In a system with four or five tracks, the schools that consider prior performance can place students into the available tracks based on this information and thus in these schools class homogeneity is higher as compared to schools that do not obtain information on prior performance of their students. The data suggests that to place students into only two or three tracks is not beneficial for student performance regardless of whether schools consider prior performance. A possible explanation for this is that two or three tracks do not allow for enough differentiation between students with heterogeneous ability. In a system with one track whether schools consider prior performance still matters for student performance: Although students in schools that sometimes or always consider prior performance do not perform differently from each other, schools that never consider prior performance perform (marginally) better.

The results presented above will be biased if the variation in the data is affected by sorting of students into schools. This may be the case when schools that consider prior performance on accepting students do so to be able to select the most able students and not to allocate students into tracks. Under the assumption of full sorting into schools, it is to be expected that schools that consider prior performance will have students that always perform better. However, this is not the case: In systems with only one track, whether schools consider prior performance does not seem to matter much (it is only marginally better to be in a school that does not consider prior performance than to be in one which does) and in a system with only a few tracks (2 or 3), whether schools consider prior performance do not seem to matter at all. Therefore, schools which consider prior performance when deciding on accepting the student to the school do not perform better by definition. Assuming sorting into schools could also lead to diverging hypotheses for systems with and without tracking: In an one track system, schools that consider prior performance should select the best students and be therefore the best performing schools. As said before, we do not see this. In systems with two or more tracks, we could expect under the assumption of full sorting into schools, that schools that consider prior performance are the schools in the higher

tracks (for instance, Gymnasia). In this case, whether schools consider prior performance is equal to track level. However, track level is controlled for in this analysis by using the ISCED level of the students. And as said before, for students in a system with 2 or 3 tracks, whether schools consider prior performance does not seem to matter for student performance.

That the hypotheses following full sorting do not seem to be confirmed by the data does not mean sorting is not a problem in these analyses. Sorting in lesser extent can still exist and could potentially bias the results. To investigate this we present results using only between countries or within country variation, as shown for mathematics in Table 5. Appendix I shows the same comparison for reading and science. The first column of Table 5 replicate column (2) from Table 4 for comparative purposes. The second column shows the same model but now including country fixed effects, as described by equation (2). The interactions are a bit smaller (2.28** vs 2.50** and 5.11** vs 7.85**), but qualitatively very similar. For all subjects, students perform best in systems with a high number of tracks when schools always consider prior performance. As also pointed out by Walldinger (2006), Brunello and Checchi (2007), and Schuetz *et al.* (2008) a model with country fixed effects provides unbiased results for cross-country analysis, assuming that the existing country heterogeneity does not influence the interaction between whether schools consider prior performance and the number of tracks. Although this model still does not allow for a strict causal interpretation, the assumption required is considerably weaker than the assumption that no unobserved country heterogeneity exists, even with a sample of very similar countries.

The third column of Table 5 replaces the school-level dummies on whether schools consider prior performance by variables depicting the national proportion of students in schools that sometimes or always consider prior performance, as described by equation (3). This model excludes the possible sorting of students into schools since this micro-phenomenon cannot intervene with the estimation when whether schools consider prior performance is measured at the country level. Schools that always consider prior performance have a negative impact on student performance (-134.23***), however this is compensated in countries with more than 3 tracks due to the positive interaction term with the number of tracks (49.89***).[13] For an average country, where 29 percent of students are in schools that sometimes consider prior performance and where 32 percent are in schools that always consider prior

13. The variables and their interactions do not show very significant results for reading and science, partly due to the reduction in the number of degrees of freedom. However, a F test is performed to see whether the main and interaction effects are jointly significant and they are for mathematics (p-value= 0.02) and reading (p=0.00). For science the five variables are not jointly significant, but the three variables relating to schools that always consider prior performance are jointly significant at the five percent level (p-value= 0.04).

performance, performance for reading is best if there is one track and performance for mathematics, and science is best if there are five tracks. For a country with a very high percentage of schools that always consider prior performance, students perform best in a five-track system, regardless of the subject.[14] For countries with high numbers of schools that never or sometimes consider prior performance, students perform best with two tracks, also regardless of the subject.[15] There results seem to indicate that, although sorting into schools could be a problem, it is unlikely that it alone drives our results.

Table 5: Comparing different models on student performance (mathematics as dependent variable)

	(1) Between countries, between schools	(2) Within country	(3) Between countries
School considers prior performance			
Sometimes	-4.95** (2.40)	-3.82* (2.03)	99.29* (54.46)
Always	-9.54** (4.25)	-2.03 (3.99)	-134.23*** (40.72)
Number of tracks (0-4)	-0.88 (3.75)	-	9.60 (17.72)
Sometimes*Number of tracks	2.50** (1.16)	2.28** (1.01)	-58.77 (37.00)
Always*Number of tracks	7.85*** (1.90)	5.11** (2.01)	49.89** (22.7)
Country FE	-	√	-
(pseudo)LL	-43,738	-43,695	-43,733
# of students	187,768	187,768	187,768
# of countries	31	31	31

Notes: The table presents coefficients from random effects models (standard errors in parenthesis) on the relation between student performance and whether or not schools consider prior performance when selecting students and the number of tracks in a country using three specifications. Column (1) shows the main model as depicted in column (2) of Table 4. Column (2) shows the same models but with country fixed effects included. Column (3) measures the school variables "school principal consider prior performance" on a national level and thus depicts the proportion of schools (between 0 and 1) in the country with school which say they always or sometimes consider prior performance. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The control variables are as described in the text. The models include imputation dummies and imputation variable interaction terms.

14. A country with a very high percentage of students in schools that always consider prior performance is defined as a country with 16 percent of its students in schools that sometimes consider prior performance (the mean minus one standard deviation) and 58 percent in schools that always consider prior performance for student acceptance (the mean plus one standard deviation).
15. The definition of such a country is one where 42 percent of students attend schools that sometimes consider prior performance (the mean plus one standard deviation) and 6 percent attend schools that always consider prior performance (the mean minus one standard deviation).

3.6.3 Tracking and inequality

The results for whether using performance to select students into tracks has an effect on the relation between tracking and educational opportunities are displayed in Table 6, and described in equations (4) and (5). As expected and consistent with the literature, PB has a positive and substantial relation with test scores. This effect is similar over the three PISA test subjects, reading, mathematics, and science and comprise about a quarter of a standard deviation in the test scores. If we look at the interaction between PB and tracking in columns (1), (2), and (3), it can be seen that tracking mitigates the association with PB and therefore reduces inequality of opportunity: The interaction of the number of tracks and PB is negative and highly significant. The first three columns show that in a system with five tracks the association of PB is lowered by 17.3 points in reading, 16.2 in mathematics, and 18.2 in science.

Table 6: The effect of tracking on inequality and the influence of selection on performance

Dependent variable	(1) Reading	(2) Math	(3) Science	(4) Reading	(5) Math	(6) Science
Parental background	23.68***	24.45***	24.76***	23.44***	24.26***	24.53***
	(1.86)	(2.06)	(2.02)	(1.91)	(2.10)	(2.07)
School considers prior performance						
Sometimes	-1.91	-1.35	-1.09	-2.51	-1.98	-1.75
	(1.94)	(1.89)	(2.05)	(2.01)	(2.01)	(2.06)
Always	6.72*	5.7*	7.75*	5.91*	4.96	6.90*
	(3.39)	(3.07)	(3.74)	(3.35)	(3.11)	(3.71)
Number of tracks (0-4)	-2.50	2.03	0.84	-2.27	2.23	1.06
	(2.62)	(3.73)	(3.38)	(2.59)	(3.71)	(3.37)
Parental background*	-3.45***	-3.23***	-3.64***	-1.93	-1.80	-2.05*
Number of tracks	(0.99)	(1.11)	(1.09)	(1.21)	(1.25)	(1.22)
Parental background*Number of tracks*Sometimes				-0.90**	-1.13**	-1.13**
				(0.37)	(0.55)	(0.52)
Parental background*Number of tracks*Always				-2.49***	-2.18***	-2.50***
				(0.72)	(0.72)	(0.67)
(pseudo-)LL	-43,752	-43,731	-44,085	-43,750	-43,730	-44,083
# of students	187,768	187,768	187,768	187,768	187,768	187,768
# of countries	31	31	31	31	31	31

Notes: The table presents coefficients from random effects models (standard errors in parenthesis) on the relation between student performance and parental background, whether or not schools consider prior performance when selecting students and the number of tracks in a country. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The control variables are as described in the text. The models include imputation dummies and imputation variable interaction terms.

Finding that tracking reduces inequality is not fully consistent with the literature, which most often finds that tracking increases inequality or has no effect. To further investigate the drivers of this positive effect of tracking, we show in columns (4), (5), and (6) in Table 6 the results of the same models, but now including also interactions between PB, the number of tracks, and whether schools consider prior performance on deciding to accept the student to the school. This allows us to check if for schools that consider prior performance, and are assumed to use this information for track placement, the effect of PB is lower. The last three columns show that the interaction between the number of tracks and PB is no longer significant; however, the triple interactions (number of tracks, PB, and whether schools consider prior performance) are. The table indicates that it is primarily the schools that always consider prior performance in systems with multiple tracks that mitigate the relation of PB with performance. Thus it is not tracking itself that diminishes the association of PB and performance but, rather, tracking combined with whether schools consider prior performance. This is consistent with our expectation that when schools consider prior performance, parents have less influence on their child's track choice and subsequent performance.

Similarly to the models on student performance, the two models on equality of opportunity with fixed effects do not seem to suffer from country heterogeneity as can be seen in columns (1) and (2) of Table 7. The interactions are very similar, but the dummy for the Always considering prior performance becomes significant and also increases.

Table 6 indicates that the differences between students of low and high PB are minimized in a highly differentiated system. To investigate whether the effects of tracking and whether schools consider prior performance on accepting students on equality of opportunity are indeed different for students of different socioeconomic background, we estimate models for the subsamples of low and high socioeconomic background students. As Table 2 already showed, PB is measured by a variable with a mean close to zero. Consequently the high PB students have positive values of PB, while the low PB students have negative values of PB. For the low PB students a lowering of the relation between PB and performance is therefore positive for their performance, while for the high PB students it is negative. Table 7 shows the results. For students with high PB (column (4)) the relation between tracks and performance is negative, irrespective of whether schools consider prior performance. However, for low PB students the number of tracks does not alter educational opportunities (column (3)). However, the triple interactions (PB, number of tracks, and whether schools consider prior performance) show that when schools always consider prior

performance, a high number of tracks is beneficial. It may be that the parents of the low PB students do not even attempt to influence the school in accepting their child and these students are thus not hindered by any objective prior performance measure, as the high PB students are. It is also possible that students with a low PB, as opposed to students with high PB, already have an efficient match between their ability and track placement.

Table 7: Robustness checks (mathematics as the dependent variable)

Model:	(1) FE	(2) FE	(3) Low PB	(4) High PB
School considers prior performance				
Sometimes	-0.56	-1.14	-2.86	-0.70
	(1.78)	(1,88)	(2.12)	(1.76)
Always	7.65***	6.95***	3.65	7.64**
	(2.70)	(2.70)	(3.16)	(3.37)
Number of tracks (0-4)			2.84	4.04
			(4.09)	(3.72)
Parental background* Number of tracks	-3.24***	-1.90	-0.50	-3.08**
	(1.13)	(1.29)	(1.78)	(1.31)
Parental background* Number of tracks*		-1.05*	-1.17	-0.36
Sometimes		(0.55)	(0.74)	(1.13)
Parental background* Number of tracks*		-2.04***	-2.54**	-0.86
Always		(0.75)	(1.19)	(0.93)
Country FE	√	√	-	-
(Pseudo) LL	-43,677	-43,687	-43,311	-43,918
# of students	187,768	187,768	95,406	92,362
# of countries	31	31	31	31

Notes: The table presents coefficients from random effects models (standard errors in parenthesis). Column (1) and column (2) replicate column (2) and column (5) of Table 6 but include country fixed effects instead of the number of tracks per country. Column (3) and column (4) replicate column (5) of Table 6 but column (3) uses only low PB students, while column (4) uses only high PB students. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The control variables are as described in the text. The models include imputation dummies and imputation variable interaction terms.

3.6.4 Robustness

Other possible distorting factors are our measure of tracking, the sample of countries and the control variables. We report all robustness checks with the PISA mathematics score, unless otherwise stated the results for reading and science are robust to the

various checks.[16] The results which are not shown in this chapter are available on request.

Unlike some other papers on this topic, we use the number of tracks available to students at the age of 15 to characterize a country's tracking regime. Since the PISA test is conducted when the students are between 15 and 16, it is possible that although multiple tracks are available to students at the age of 15, students have not yet been tracked for a substantial amount of time. If this is the case, the association with tracking in late selection countries may be too weak to be picked up. To check for this, we redo the analysis using only late trackers.[17] The results are presented in Appendix J. The results for the late-tracking countries have the same sign and are similar in coefficient size, although the interactions are insignificant for mathematics and science, possible due to the large drop in observations.

Using dummies for the number of tracks as in Appendix H reveals that especially countries with only a few tracks perform worse. Looking at the countries with only a few tracks, the three countries with two tracks (Israel, Greece, and Chile) are among the worst-performing countries with regard to PISA test scores in reading, mathematics, and science. However, excluding these countries does not influence the results much. We also excluded each country, one by one, to determine if one of these drives the results and found that they do not.

We also excluded one by one parts of the control variables: Excluding the student variables does not change the results much, although the results are less significant. Excluding all the school variables would also exclude the school system interaction terms we are interested in, but we excluded either school composition, school inputs, or other school characteristics; we found no large changes in the models on performance. When school composition is excluded in the models on inequality, the dummy for whether schools always consider prior performance becomes larger and is always highly significant, while excluding school characteristics lowers the coefficient of this dummy and renders it insignificant.

16. Most of the changes are changes in the coefficient for the dummy indicating that the school always considers prior performance. The main result, that schools that consider prior performance for student acceptance mitigate the relation between strong tracking and performance, holds in any specification.
17. Late-tracking countries, which select at age 14 or later, are Croatia, Denmark, Estonia, Finland, Greece, Iceland, Ireland, Israel, Italy, Latvia, Lithuania, New Zealand, Norway, Poland, Portugal, the Russian Federation, Slovenia, Spain, Sweden, and the United States.

3.7 Conclusion

The variation in tracking in education systems throughout the Western world is quite large: Many countries have no formal tracking in secondary school (although most have some form of non-institutionalized tracking), while some countries distinguish up to five tracks for students. Also the manner in which tracking is implemented on the school level differs widely. Some countries put more emphasis on the use of performance to place students into the available tracks (e.g. the Netherlands), while in other countries parents have more influence on the track their child will go (e.g. Germany). In this chapter it is argued that the inconsistencies between the theory of tracking and some of the empirical results in the literature could be explained by country differences in track placement. When track placement is done not based on an ability measure but mainly on PB, or, related, residential areas, the theoretical benefits of tracking might not arise. These theoretical benefits rely heavily on the idea that tracking leads to more homogenous ability classes which the teacher is better able to teach and where students benefit from a curriculum tailored to their needs and abilities. When students are placed into tracks based not on prior performance but on PB, classes will be more heterogeneous and students might not be taught a fitting curriculum. Furthermore, when PB is used to decide on track placement, educational opportunities are affected and inequality is likely to grow.

To study whether using performance to select students into tracks has an effect on the relation between tracking and student performance and educational opportunities, this chapter uses the data of around 185,000 students in 31 comparable countries from PISA 2009. Prior performance can be thought of as an important measure of observed student ability. Therefore, for schools have the information on prior performance it can help them allocate students across tracks, allowing for a better match between student ability level and track level which benefits the student, both by allowing the student to learn more and by limiting the effect of PB on student performance. The analyses in this chapter show that tracking in general does not have a direct relation with performance. On the other hand, interactions between tracking and whether schools consider prior performance reveal that students in highly differentiated systems perform best when schools always take into account prior performance to decide on student acceptance. In systems with a low number of tracks, whether schools consider prior performance has less of an impact.

The association between PB, tracking, and student performance shows that equality of opportunity is best provided for in a system with a high number of tracks combined with schools always consider prior performance on accepting the student to the school. It turns out that for high PB students in these systems, tracking weakens the positive relation between PB and performance, whereas for low PB students the (for them

negative) relation between tracking and performance is lowered primarily when they attend schools that always consider prior performance. Thus it seems that high PB students might be harmed by tracking when schools consider prior performance.

We argue that it is not straightforward to determine whether tracking in itself has a positive or negative effect on performance. When education system characteristics are studied, it should be taking into account that schools can have large influence on the implementation of these system characteristics and thus heterogeneous effects across schools can arise. In this chapter we show that when tracking is combined with whether schools consider prior performance in accepting the student, tracking benefits both student performance and educational opportunities.

Appendix H: Nonlinear models (Chapter 3)

This Appendix presents the same models as Table 4 and 6 in the main text of Chapter 3 but now including the number of tracks in a country as dummy variables instead of including it as a continuous variable. To avoid too many dummies and, especially, too many interactions between the number of tracks and whether school principals consider prior performance when accepting students to the school, we split up the categorical variable in three: No tracking, two or three tracks, or four or five tracks. The results on the relation between tracking, selection and student performance are qualitatively the same as in the main text, as can be seen in Table H1: The interactions between number of tracks and whether school principals consider prior performance sometimes or always are positive and significant. The combined coefficients of the main effects and the interaction effects show that for students in schools where the principal considers prior performance in a country with four or five tracks, tracking is either positive or neutral as compared to students in a school where the principal does not consider prior performance in a country without tracking. Different to the main results is that *students in a country with two or three tracks always perform much worse than students in any of the education* systems, irrespective whether their school principal considers prior performance. This is further examined in Section 3.6.4.

Table H1: Non-linear models looking at student performance

Dependent variable	Reading	Mathematics	Science
School considers prior performance			
Sometimes	-6.56	-4.69**	-5.83***
	(2.53)	(2.36)	(2.15)
Always	-4.97	-4.04	-5.47
	(3.21)	(3.77)	(4.93)
2-3 tracks	-32.61***	-39.10***	-36.67***
	(8.87)	(10.38)	(11.77)
4-5 tracks	-18.08**	-2.98	-10.45
	(8.32)	(10.23)	(9.33)
Sometimes*2-3 tracks	6.11	3.82	6.00
	(3.99)	(4.11)	(4.23)
Always*2-3 tracks	9.24	3.39	9.58
	(6.83)	(6.07)	(8.22)
Sometimes*4-5 tracks	10.58***	8.78**	2.45***
	(3.95)	(4.18)	(4.11)
Always*4-5 tracks	21.79***	21.92***	26.74***
	(5.48)	(5.14)	(6.28)
(pseudo)LL	-43755	-43732	-44088
# of students	187768	187,768	187,768
# of countries	31	31	31

Notes: Coefficients with standard errors in parenthesis. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. See also notes of Table 3 of Chapter 3.

Table H2 shows similar models to those in the first 3 columns of Table 6 but also here including the number of tracks in a country as three dummy variables. The interactions between parental background and the number of tracks are significantly negative. This means

that the relation between parental background and student performance is lower in countries with more tracks. Also this result is similar to those presented in the main text.

Table H2: Non-linear models looking at equality of opportunity.

Dependent variable	Reading	Mathematics	Science
School considers prior performance			
Sometimes	-1.68	-1.08	-0.62
	(2.02)	(1.88)	(2.08)
Always	6.87*	5.76	8.24*
	(3.28)	(2.98)	(3.67)
2-3 tracks	-30.51***	-39.43***	-35.00***
	(8.52)	(10.42)	(10.99)
4-5 tracks	-6.95	8.40	3.12
	(7.73)	(10.08)	(9.42)
Parental background* 2-3 tracks	-12.50***	-12.66***	-14.18***
	(2.34)	(2.56)	(2.55)
Parental background* 4-5 tracks	-10.90***	-10.59***	-11.33***
	(3.29)	(3.30)	(3.40)
(pseudo-)LL	-43742	-43719	-44074
# of students	187768	187,768	187,768
# of countries	31	31	31

Notes: Coefficients with standard errors in parenthesis. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. See also notes of Table 3 of Chapter 3.

Appendix I: Comparing different models for reading and science (Chapter 3)

This appendix compares the different estimation models for reading and science, like is done in the main text for mathematics in Table 5 of Chapter 3. The main difference is that the models showing results using only the between country variation is not significant for reading or science. However they do show the same trend and it must be remembered that since these models use country level data only the number of observations is effectively reduced to 31. An F test is performed to see whether the main and interaction effects are jointly significant and they are for mathematics (p-value= 0.02) and reading (p=0.00). For science the five variables are not jointly significant, but the three variables relating to schools that always consider prior performance are jointly significant at the five percent level (p-value= 0.04).

Table I1: Comparing different models on student performance (reading as dependent variable)

	(1) Between countries, between schools	(2) Within country	(3) Between countries
School considers prior performance			
Sometimes	-5.97**	-4.37**	35.24
	(2.54)	(2.17)	(39.05)
Always	-10.13**	-2.67	-66.67**
	(4.11)	(4.10)	(30.42)
Number of tracks (0-4)	-5.74**		5.47
	(2.70)	-	(12.54)
Sometimes*Number of tracks	2.83**	2.54**	-36.50
	(1.15)	(1.05)	(23.61)
Always*Number of tracks	8.64***	6.03***	19.95
	(1.82)	(1.96)	(18.06)
Country FE	-	√	-
(pseudo)LL	-43759	-43727	-43758
# of students	187768	187768	187768
# of countries	31	31	31

Notes: Coefficients with standard errors in parenthesis. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Column (1) shows the main model as depicted in column (2) of Table 4. Column (3) measures the school variables "school principal consider prior performance" on a national level and thus depicts the proportion of schools (between 0 and 1) in the country with school which say they always or sometimes consider prior performance. See also notes of Table 3 of Chapter 3.

Table I2: Comparing different models on student performance (science as dependent variable)

	(1) Between countries, between schools	(2) Within country	(3) Between countries
School considers prior performance			
Sometimes	-5.99**	-5.40**	78.61
	(2.44)	(2.20)	(55.61)
Always	-10.90**	-5.07	-86.87*
	(5.06)	(4.88)	(46.45)
Number of tracks (0-4)	-2.89		6.40
	(3.31)	-	(16.08)
Sometimes*Number of tracks	3.45***	3.33***	-42.89
	(1.08)	(0.99)	(32.60)
Always*Number of tracks	9.68***	7.39***	34.62
	(2.04)	(2.08)	(24.08)
Country FE	-	√	-
(pseudo)LL	-44092	-44054	-44092
# of students	187768	187,768	187,768
# of countries	31	31	31

Notes: Coefficients with standard errors in parenthesis. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Column (1) shows the main model as depicted in column (2) of Table 4. Column (3) measures the school variables "school principal consider prior performance" on a national level and thus depicts the proportion of schools (between 0 and 1) in the country with school which say they always or sometimes consider prior performance. See also notes of Table 3 of Chapter 3.

Appendix J: Subsample of early tracking countries (Chapter 3)

Table J1 presents similar results to Table 4 of Chapter 3 for late tracking countries only, since in these countries the time in a tracked system may have been too short to have an effect on student performance. The results are for the largest part comparable to those using the full sample, but the significance of the results is lower, especially for mathematics and science.

Table J1: Results comparable to Table 4 of Chapter 3 for late tracking countries only

	(1) Reading	(2) Math	(3) Science
School considers entrance requirements			
Sometimes	-6.86***	-4.51***	-5.37***
	(2.20)	(2.16)	(2.01)
Always	-7.61	-5.84	-6.34
	(4.78)	(5.25)	(6.34)
Number of tracks (0-4)	-7.85**	-7.94**	-5.15
	(3.71)	(3.96)	(4.64)
Sometimes*Number of tracks	3.32**	2.78	2.17
	(1.62)	(1.90)	(1.77)
Always*Number of tracks	6.31**	3.63	5.01*
	(2.65)	(2.25)	(2.76)
(Pseudo) LL	-28,138	-28,132	-28,358
# of students	131,423	131,423	131,423
# of countries	14	14	14

Notes: The table presents coefficients from random effects models (standard errors in parenthesis) on the relation between student performance and whether or not schools consider prior performance when selecting students and the number of tracks in a country. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The control variables are as described in the text. The models include imputation dummies and imputation variable interaction terms.