Maastricht University

Selection and tracking in secondary education; A cross country analysis of student performance and educational opportunities

Citation for published version (APA):

Korthals, R. A. (2013). *Selection and tracking in secondary education; A cross country analysis of student performance and educational opportunities*. Maastricht University, Graduate School of Business and Economics. GSBE Research Memoranda No. 054 https://doi.org/10.26481/umagsb.2013054

Document status and date:
Published: 01/01/2013

DOI:
10.26481/umagsb.2013054

Document Version:
Publisher's PDF, also known as Version of record

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy
If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Download date: 17 Apr. 2024

Maastricht University

Research Centre for Education and the Labour Market | ROA

Selection and tracking in secondary education: A cross country analysis of student performance and educational opportunities

Roxanne Korthals

ROA Research Memorandum

ROA-RM-2012/14

Research Centre for Education and the Labour Market
Maastricht University
P.O. Box 616, 6200 MD Maastricht, The Netherlands
T +31 43 3883647 F +31 43 3884914

secretary-roa-sbe@maastrichtuniversity.nl
www.roa.nl

Selection and tracking in secondary education: A cross country analysis of student performance and educational opportunities

Roxanne Korthals

ROA-RM-2012/14*
Updated version:
September 2013

Research Centre for Education and the Labour Market
Maastricht University
P.O. Box 616, 6200 MD Maastricht, The Netherlands
T +31 43 3883647 F +31 43 3884914

secretary-roa-sbe@maastrichtuniversity.nl
www.roa.nl

* The ROA Research Memorandum Series was created in order to make research results available for discussion, before those results are submitted for publication in journals.

Abstract

Selection and tracking in secondary education: A cross country analysis of student performance and educational opportunities**

This paper examines the effect of tracking in secondary school on student performance and educational opportunities, taking into account whether prior performance is considered when students are selected in the different tracks. The sample consists of data from the Programme for International Student Assessment 2009 for around 185,000 students in 31 comparable countries. The results are controlled for student- and school-level confounders. The results indicate that when tracking is implemented, it does not have a direct relation with performance. However, system and school interactions reveal that a highly differentiated system is best for student performance when schools always take into account prior performance to decide on student acceptance. In systems with a few tracks, admission rules have less of an impact and tracking is only mildly associated with performance. Equality of opportunity is best provided for in a system with many tracks, especially when schools always consider entrance requirements. However, caution is warranted in interpreting these results since selection issues could play a role.

JEL classification: I20, I21, I24
Keywords: economics of education, tracking, equality of opportunitie, education, PISA

Roxanne Korthals
ROA
Maastricht University
P.O. Box 616
6200 MD Maastricht
The Netherlands
r.korthals@maastrichtuniversity.nl

** I would like to thank Jaap Dronkers for his valuable comments and fresh perspective. Furthermore, I would like to thank Rolf van der Velden, Hans Heijke, Olivier Marie and Andreas Ammermüller for a number of useful comments and suggestions. I 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 paper.

1 Introduction

Separating secondary school students into different educational tracks based on ability is a common practice in the Western world. Some countries have national policies regarding tracking, while others let schools decide whether and how to implement informal tracking. This paper studies the effect of nationwide tracking policies on academic performance and inequality.

While there are numerous papers on the effects of tracking, most (e.g., Hanushek and Woessmann, 2006 and Ariga and Brunello, 2007) do not consider the school level, while schools could potentially have a strong influence on the implementation of education policies. In the case of tracking, however, one characteristic seems particularly important: whether schools track students based on prior performance. When the implementation of tracking at the school level is not in line with the idea of grouping students of different abilities (e.g., tracking by socioeconomic background, tracking is not implemented efficiently and the effect of tracking may not be accurately estimated. In other words, students and school characteristics, such as entry requirements, interact with system characteristics, such as tracking. It is therefore not straightforward to determine the effect of tracking on student performance.

A related issue is whether tracking changes educational opportunities, that is, whether it alters the effect of family background (FB) on performance. Tracking may be beneficial for certain student groups but may affect others negatively.

Determining the association of tracking with student achievement is important for two reasons: First, basic economic theory predicts that institutions, which influence incentives, have large effects on outcomes. Thus, tracking has a large potential impact on achievement and is therefore important in studying student performance. The second reason is that all students who receive formal education do so within a given education system. To fully understand the mechanisms behind the effects of FB and school characteristics on performance, it is important to consider the effect of education system characteristics on performance and possible interactions between school and system characteristics.

The results presented in this paper indicate, consistent with the literature, that tracking does not have a significant direct relation with performance. However, system and school interactions reveal that in a system with four or five tracks, it is best for student performance that schools always consider prior performance when selecting students. For systems with only two or three tracks, the use of entry requirements does not alter the association between tracking and performance. Thus, tracking only has a significant relation with performance when school characteristics are taken into account. This finding indicates that it is important to consider school characteristics when analysing education systems.

Equality of opportunity is best provided by a system with four or five tracks, especially when schools always consider entrance requirements. In these systems, tracking mitigates the positive relation of FB

for higher FB students, while it seems to have no effect on the association between FB and performance for students with a lower FB. Due to these opposing effects, the performance differences between students with high and low FB is lowest in a more differentiated system.

However, caution is warranted in interpreting the above-mentioned results since the selection of students into schools with entrance requirements may play a role.

The structure of this paper is as follows: Section 2 reviews the tracking literature. Section 3 discusses theoretical insights in tracking and how schools can affect tracking. Section 4 describes the data used in the analysis, while Section 5 lays out the empirical strategy. Sections 6 and 7 present the results and discuss possible selection problems, respectively. Section 8 discusses some robustness checks. Finally, the last section concludes the paper.

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 in secondary school. This paper most closely relates to the work of Hanushek and Woessmann (2006), Fuchs and Woessmann (2007), Ariga and Brunello (2007), Pekkarinen (2008), Pekkarinen, Uusitalo, and Kerr (2009), Horn (2009), and van Elk, van der Steeg, and Webbink (2011).[1]

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 materialised 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.

[1] Instead of focusing on formal tracking measured at the country level, studies often focus on informally grouping students into different classes within tracks, measured at the school level. This 'ability grouping', or 'streaming', however, is not the focus of this paper since its informal nature makes it more difficult to compare over multiple systems. For more on ability grouping, see, for instance, Figlio and Page (2002), Betts and Shkolnik (2005), or Duflo, Dupas, and Kremer (2011).

Pekkarinen (2008) and Pekkarinen et al. (2009) 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. The curriculum was also changed in the reform. 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. Pekkarinen et al. (2009) 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 of mostly 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 study is related to the work of Ammermueller (2005), Walldinger (2006), Brunello and Checchi (2007), and Schuetz, Ursprung, and Woessmann (2008). This paper defines inequality as the existence of FB influence on student performance.[2] Ammermueller (2005) finds that the more tracks available to students, the higher the positive effect of FB on performance growth from primary to secondary school. Walldinger (2006) finds that in the educational opportunities are lower in early tracking countries. However, this effect

[2] A second type of inequality is inequality in test scores, which arises when the test score distribution widens. Hanushek and Woessmann (2006) find that early tracking has negative effects on the variance of student test scores, consistent with the results of Horn (2009).

disappears when the effect of FB in primary 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 FB 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 FB 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.

In general, it seems that the FB 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 could be explained by a larger effect of FB in primary school in these countries or its disappearance over time.

This paper contributes to the literature as follows. First, by studying the number of tracks available to students, it examines the extent of tracking, whereby students are placed into more or less homogeneous peer groups. Since I expect the possible effects of tracking to be strongest due to its influence on the homogeneity of classes, this tracking characteristic, as opposed to the year of first selection, is especially interesting.

Second, building on the work of Dunne (2010) and Dronkers (2010) but unlike most other studies, this paper measures tracking at the country level but also includes school-level variables. Since I use a country-level measure of tracking, I capture the effect of tracking as a formal education system characteristic. A consequence of including school-level variables is that part of the effect of tracking is captured by these variables Although this weakens the country-level effect of tracking, this approach provides a more accurate investigation of tracking since schools can have a large impact on student performance (see, e.g., Hoxby (2001) for the impact on peers and Fryer (2011) for the impact on teacher incentives).

Therefore the effects of school and system characteristics on performance also provide insights for comparing the effects of school characteristics across education systems and thus across countries. By looking at schools and education systems, it is possible to analyse the relation between school policies and performance, *given* system characteristics. This method facilitates the comparison of the effect of tracking across countries.

A third, related contribution is that by including school policies, I can study whether the association of tracking with performance changes when school policies intervene with system characteristics, in other words, when tracking is not effectively implemented. I show that system characteristics such as tracking cannot be considered separately from school entrance selection practices. The omission of school practices in previous studies may explain their inconclusive results obtained. This paper tries to bridge this gap.

3 Tracking

In a sense, tracking is a form of imposing peer homogeneity on students. Tracking separates students into groups based primarily on 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 achievement 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, I 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.[3] For 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, performance homogeneity in tracks can be good for both high- and low-ability students when teachers adjust 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 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 specialised curricula, which means that students in different tracks are taught different things at different levels of difficulty. As long as the specialised curricula are optimally designed for the average characteristics of the students in the track, they should increase performance.[4]

Overall, I would expect a positive effect of tracking due to improved teaching strategies and adjusted curricula. The effect of tracking due to more homogeneous peer groups is theoretically uncertain; however, the above arguments assume that the tracking is implemented efficiently. The inadequate allocation of students to tracks, that is, not based on ability or with selection taking place before ability can be observed, can lead to lower performance Dustmann (2004) shows that parental background 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. When this happens, however, tracks are no longer homogeneous in ability. Thus I would expect that schools using an objective measure to select students would improve

[3] For an overview of these models, see Sacerdote (2011).

[4] Whether specialised curricula increase the minimum level of general knowledge society expects all students to have is unclear. Thus there are no focused theoretical expectations on the effect of specialised curricula on performance when performance is measured by comparable international test scores in reading, mathematics, and science.

homogeneity in tracks and thus the efficiency of tracking.

Theoretically, the selection of students into multiple tracks has a positive effect on performance, disregarding uncertain peer effects. This positive effect is expected to be strongest in systems with more tracks since the classes are then even more homogeneous and the curricula and teaching are even more within reach for all students than in systems with fewer tracks. However, if tracking is not implemented efficiently (i.e., not based on observed ability), this positive effect cannot (fully) manifest itself or tracking can even have a negative effect. Depending on the cause of the inefficient student allocation, tracking can increase the effect of FB and thus decrease equality of opportunity.

3.1 Tracking and schools

School selection of students into different tracks can be done efficiently if based primarily on performance (i.e., ability and motivation). The fact that selection by schools can be efficient or inefficient reveals two potential problems with the selection process: First, when students are not selected based only on observed ability, they can be assigned to the wrong tracks. Second, since ability is difficult to observe, not all students are correctly classified. The first problem can be mitigated by ensuring that the selection of students is always based primarily on observed ability. A proxy for ability can be prior performance and/or the previous (primary school) teacher's opinion of the student's ability. Test scores and the teacher's assessment of the student are the best available but still imperfect measures of student ability. Naturally, I do not assume when schools use test scores and teacher assessments that 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.[5] However, I assume that as long as objective ability measures limit non-ability–related parental influence on track choice, tracking efficiency in improved.

The implementation of tracking is most efficient when schools use objective student entrance requirements. Schools placing students in different tracks without an awareness of their abilities or basing their selection on non-academic criteria can severely limit the possible positive effects of tracking.[6]

[5]This environment created by parents includes choices made earlier in the child's life, such as those on pre-school attainment and in primary school.

[6]The analogue of tracking in schools is ability grouping, where schools can groups students based on ability within classes or tracks for some or all subjects. This informal tracking allows schools to differentiate between students with different abilities. Especially in a comprehensive system, ability grouping is often used to prepare certain students for tertiary education. However, since ability grouping is not a formal part of the education system, the manner in which students are grouped differs across schools. These differences are often unobservable in ability grouping measures, which complicates the analysis of the effects of ability grouping. Since ability grouping may have an influence, this study includes it as a control variable; however, since ability grouping is subject to differences in interpretation, it is not part of the main discussion.

3.2 Tracking and equality of opportunity

FB combined with tracking can influence student performance in a number of ways. First, as mentioned above, when ability is not perfectly observable or ability measures are not consulted by schools, parents may have a large influence on a child's track placement. High FB students are then more likely to be placed in the higher tracks and being in the higher tracks can increase student performance, irrespective of ability, since students are exposed to more demanding curricula. A comprehensive education system could then be expected to lessen the impact of FB on student performance since there is no track choice to be selected into. However, if selection into schools takes place in a comprehensive system (e.g., when parents move to neighbourhoods with good schools or send their children to private schools), thereby grouping together students with high FB, the effect of FB can remain strong.

Second, peer effects may increase or decrease the effect of FB on student performance. For instance, being surrounded by high-ability or high FB peers can compensate for a student's lower FB when there is a spillover effect of positive FB across students. In a tracked system, such circumstances would then increase the positive FB effect on student performance. Since FB is positively correlated to performance, this increased effect would be positive for students in the higher tracks, who are more often surrounded by high FB students, and negative for low-ability students, who are more often surrounded by low FB students. In a comprehensive system where all students from different socioeconomic backgrounds and ability levels are grouped together, there is less chance of being surrounded by only high or low FB peers and thus the beneficial or negative effects of peers' FB are mitigated.[7]

4 Data

The student- and school-level data used in this paper are from the 2009 wave of the Programme for International Student Assessment (PISA 2009), executed by the Organisation for Economic Co-operation and Development (OECD). These data include internationally comparable test scores in reading, mathematics, and science and information on students FB 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 from all participating countries is subjected to tests on reading, mathematics, and science. The test results are standardised to a mean of 500 and a standard deviation of 100 for the OECD countries. In addition to the tests, the students and school principals (and in some countries also the parents) are surveyed. A total of 34 OECD countries and 41 partner countries participated in PISA 2009. Since these countries are very diverse in their economic development, this paper uses a selection of comparable countries to limit country heterogeneity that may influence the results. Of the 75 countries in PISA 2009, this study covers 31. All 31 countries have a gross domestic product (GDP) per capita above the

[7]The peer effect of FB may still exist to a lesser extent due to spatial segregation, whereby high and low FB families live in separate areas (see also the argument above).

minimum of the OECD (i.e., Turkey, with $11,655.23) and available data on tracking.[8]

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 (pre-)vocational or general education who were in schools where more than five students participated in PISA 2009.[9] This amounts to 187,768 students in 7,489 schools in 31 countries.[10]

4.1 Tracking and selection by schools

This paper defines tracking as the separation of students into tracks that differ in ability levels and curricula. The measure of the extent of tracking is the 'number of school types or distinct educational programmes available to 15-year-olds', taken from OECD (2007).[11] Table 1 shows the number of tracks for the sample of countries. This measure of tracking is different from those used in 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. This consequence of tracking is expected to have the largest effect on performance. Its disadvantage, however,is that it does not take into account the amount of time students spend in the tracks. Section 8.2 addresses this issue. However, the correlation between the age of first selection and the number of tracks in a system is very high, around -0.80 among the countries in this analysis. These results could thus be extrapolated to age of first selection.

Efficient implementation of tracking is proxied by an index of academic school selectivity. This index is based on a PISA 2009 question for 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 ac-

[8] 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. I could have included Mexico but since many of its school variables have very different distributions from those of the other countries, I excluded it.

[9] This paper includes only native students because previous literature has shown that native and migrant students react differently to system characteristics (e.g., Dronkers, van der Velden, and Dunne, 2012.

[10] The 31 countries included in this analysis together tested 214,302 students. Of those 214,302, 25,404 individuals were first- or second-generation immigrants, 1,046 were students in schools where fewer than five students participated in PISA 2009, and 84 students in Luxembourg were in modular education. Excluding these students in this order leads to a sample of 187,768 individuals in 7,489 schools in 31 countries.

[11] This variable is given in table 5.2 of OECD (2007).

ceptance, and schools where at least one of the two factors is always considered. Table 1 gives an overview of school policies with respect to entrance requirements per country. Countries differ substantially on the percentage of schools that use entry requirements, from 79.6% of students in schools in Spain that never use entrance requirements to 93.5% that always use them in Croatia. The type of school that has entry requirements also differs across countries: for instance, in the Czech Republic and Hungary schools that select are often upper secondary schools, while in Austria and Poland it are mostly schools that students with a high FB attend. Overall, no large correlations exists between entry requirements and school characteristics. Appendix A provides more information on the type of school that has entrance requirements per country.

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 abide by the admission rules. Thus, even in systems with a high number of tracks, some schools do not use prior academic achievement or teacher recommendations to decide school admittance. In the seven countries with four tracks, only 55% of the students go to schools that always consider prior performance; in the two countries with five tracks, 18% of students go to schools that never use entry requirements. Maybe more surprising, even in a comprehensive system, some schools consider (either sometimes or always) entrance requirements for accepting students: In the 10 countries with only one track, 45% of students had to supply information on 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. This lack of explanation is unsatisfactory and this paper regards this deviation as a proxy for the inefficient implementation of a nationwide education policy. Section 7 returns to this topic.

[Table 1 about here.]

4.2 Control variables

Student variables are collected through student surveys. The study's control variables are gender, age, a dummy equal to one if the student is in (pre-)vocational education and zero if in general education, a dummy equal to one if the student is in upper secondary school and zero if in lower secondary education[12], and FB (. FB is captured by an index that describes the student's economic, social, and cultural status and consists of the parents' highest occupational status, their highest educational level, and home possessions.[13, 14]

[12]Lower and upper secondary education is based upon International Standard Classification of Education (ISCED) levels. The ISCED levels provide internationally comparable standards for comparing education levels.

[13]Home possessions comprises family wealth possessions, cultural possessions, home educational resources, and the number of books in the home.

[14]The preferred measure of FB is family income. Unfortunately, family income is not available from PISA; however, the student's economic, social, and cultural status is a widely used proxy for FB.

The school-level variables are collected through a survey completed by the school principal and can be sorted into three categories: school composition, school inputs, and other school characteristics.[15] School composition is captured by the school average and standard deviation of the FB of all the students per school and by the percentage of them who speak a language other than the test language at home.[16] 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. The other school variables 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 uses ability grouping. I also control for the GDP per capita, which is for 2008 from the World Bank (2012).Table 2 provides descriptive statistics for all variables.

[Table 2 about here.]

5 Empirical strategy

5.1 Tracking and performance

The aim of this paper is to investigate the relation between tracking and performance, taking into account school entrance requirement policies. The education production function used in this paper is a linear function:

$$Test_{ijk} = \beta_0 + Student_{ijk}\beta_1 + School_{jk}\beta_2 + EntReq_{jk}\beta_3 + \#ofTracks_k\beta_4 + GDPpc_k\beta 5 + \eta_{ijk} \quad (1)$$

where $Test_{ijk}$ is the individual PISA test score in reading, mathematics, or science of student i in school j in country k, $Student_{ijk}$ is a matrix of student and FB variables, $School_{jk}$ is a matrix of school variables, $EntReq_{jk}$ is a matrix containing the dummies on school entrance requirement policies, $Tracking_k$ is a vector containing the number of tracks available to students in each country, $GDPpc_k$ is a vector containing GDP per capita data, and η_{ijk} is the error term.

Since students are nested within schools within countries, estimating equation 1 leads to downwards-biased estimates for the standard errors. By ignoring the nested data structure, it is implicitly assumed that all observations are independent from each other and thus each observation contains new information on the full set of variables. However, this is not the case. By directly estimating equation 1, the researcher

[15] There is discussion about whether or not including the school level is optimal in analysing student performance across countries. By including school-level variables or even only school-level variance, part of the effect of the educational system will be captured by these variables and other coefficients may also be affected. However, since this approach provides a more accurate picture of tracking, this study includes school-le vel variables. However, a two-level fixed effects model where the entrance requirements are included at the individual level delivers similar results.

[16] For the school average and standard deviation of the FB, all the students in the sample schools are used, both natives and immigrants.

assumes the data contain more information than they actually do and thus the standard errors are lower. These lower standard errors lead to easy acceptance of every significance test, especially since the data contain around 185,000 observations. To obtain correct standard errors, the estimation technique should take into account the nested data structure. The analysis uses random effects to mimic as closely as possible the data collection process.[17,18]

Another change to equation 1 is that since I am primarily interested in the relation between tracking and performance and the effects schools have on this, I also include a school system interaction term. The random effects model for the education production function is then

$$Test_{ijk} = \beta_0 + Student_{ijk}\beta_1 + School_{jk}\beta_2 + EntReq_{jk}\beta_3 + \#ofTracks_k\beta_4 + GDPpc_k\beta_5 + EntReq_{jk}\#ofTracks'_k\beta_6 + u_k + u_{jk} + \eta_{ijk} \quad (2)$$

where there is now a total of three error terms: for individuals, schools, and countries, respectively. This random effects model more accurately estimates the standard errors and thus the significance levels are a better indicator of the existence of the relations found.

5.2 Tracking and inequality

The main relation between FB and performance is included in the vector $Student_{ijk}$. I can examine the relation between FB and student performance and the influence of tracking thereon by including an interaction between FB and tracking. The model used to study this relation is

$$Test_{ijk} = \beta_0 + Student_{ijk}\beta_1 + School_{jk}\beta_2 + EntReq_{jk}\beta_3 + \#ofTracks_k\beta_4 + GDPpc_k\beta_5 + FB_{ijk}\#ofTracks'_k\beta_6 + u_k + u_{jk} + \eta_{ijk} \quad (3)$$

5.3 Missing values

Almost all student- and school-level variables have some missing observations (see last column of Table 2). Although most variables have at most 3% 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, a questionable assumption. Another reason for not deleting all observations is that it leads to distorted weighting. The sample weights assigned to all students are constructed to obtain a representative sample. However, if some students are deleted but are not randomly distributed among the groups for which the weights

[17] Random effects are often used for panel data sets. There random effects control for the fact that observations over time are nested within one observational unit. Here random effects are employed to take into account another form of nested data: Students are nested in schools in countries.

[18] The sum of the weights per country is equalised over all countries such that each country has an equal weight in the estimation. The weights are also normalised to sum up to the total number of observations. The random effects model requires weights for each data level, that is, conditional student weights and school weights to control for the two-stage sample design. The PISA data provide unconditional student weights and school weights. Conditional student weights are obtained by dividing the unconditional student weight by the school weight. All statistics and estimations are executed using weights, unless otherwise specified.

correct, the sample is no longer representative, even when weights are used. For these two reasons the missing values in the sample are replaced by group averages.

The method for replacing missing values is simple: 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, and the country variables do not contain missing variables. To control for possible bias introduced by such a simple method for replacing missing values, imputation dummies and imputation interactions are used.[19] Thus the final education production function for investigating the relation between school policies and tracking is

$$Test_{ijk} = \beta_0 + Student_{ijk}\beta_1 + School_{jk}\beta_2 + EntReq_{jk}\beta_3 + \#ofTracks_k\beta_4 + EntReq_{jk}\#ofTracks'_k\beta_5 + D_{ijk}\beta_6 + D_{jk}\beta_7 + D_{ijk}Student'_{ijk}\beta_8 + D_{jk}School'_{jk}\beta_9 + u_k + u_{jk} + \eta_{ijk} \quad (4)$$

where D_{ijk} and D_{jk} are dummies for cases where the student has missing values for student and/or school variables. Similar changes are made to equation 3.

6 Results

The analysis starts byinvestigating whether the number of tracks has a direct and significant relation with student performance. Since I include a wide variety of school background variables, which may capture part of the tracking effect, I do not expect a large coefficient for tracking and, indeed, do not find one. For all three test subjects, the relation between schools that sometimes use entry requirements and performance is negative, while for schools that always use entrance requirements this relation is positive (and significant). The insignificant association between the number of tracks and performance is negative for reading and positive for mathematics and science. Table 3 shows the results.

All the control variables have the expected sign. FB 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 do worse than students in general education and students in upper secondary education perform better than those in lower secondary school. The school average in FB 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% of the students speak a different language at home (for mathematics and science) or more than 40% 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

[19]The results are robust to the exclusion of the imputation variable interaction terms and to the inclusion of the imputation dummies.

agglomerations perform better and students in schools that use ability grouping perform worse. The coefficients of the full models are given in Appendix B.

[Table 3 about here.]

6.1 Tracking and performance

Table 4 shows the main results from linear models (first to third columns) and non-linear models (fourth to sixth columns). These main models include the interactions between the number of tracks and school policies.[20] The linear models use the number of tracks in a country as a continuous variable ranging from zero to four, while the non-linear models 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). As can be seen from the first three columns, including the interaction terms results in significant relations between tracking, school policies, and student performance. The results for reading, mathematics and science are very similar except that for reading the main effect of the number of tracks is large and significant. For the non-linear models, in the fourth to sixth columns, only the interactions between the number of tracks and school entry requirement policies are significant when countries have a large number of tracks (four or five).

Since the models include interaction terms, the main effects no longer have the standard interpretation. To simplify the interpretation of the main effects and the interactions, Figure 1 illustrates the combined relations between entrance requirements, tracking, and student performance. The figures show for each combination of number of tracks and school policy what the relation between the two and student achievement is, relative to students in a system with only one track in schools that never consider entrance requirements. Looking first at the panels on the left, one sees the same trend for all three subjects (reading, mathematics, and science): Schools in multiple track systems do better when they consider admission rules more often, while schools in comprehensive systems perform better when they do not consider entrance requirements. When only the significant differences in the graphs are considered, it becomes clear that for two or more tracks school policies regarding entrance admission only change the results when schools always consider prior performance. The coefficients for the Never and Sometimes entry requirements are not significantly different from each other when the number of tracks is two or more.[21] This result is confirmed by the graphs displaying the combined coefficients of the non-linear models, seen in the panels on the right in Figure 1. Again the combined coefficients are not significantly different across school policies when a system has only a few (two or three) tracks and two or three tracks are negatively associated with student performance. However, schools that always consider entry requirements in a system with a higher number of tracks (four or five) perform significantly better compared to other schools.

[20] The control variables all have the same sign and similar coefficients as in the model without the interactions.

[21] Except for science, where the combined coefficients for the Never and Sometimes entry requirements are significantly different in a system with five tracks.

These results indicate that determining whether tracking is good or bad for performance is not a straightforward process. Compensation by school policies can be quite large in terms of points on achievements tests. In a five-track system, the estimated relations show that the difference in student performance between being in a school that never considers entrance requirements and being in one that always considers them is equal to around 22–28 points on the PISA test, equivalent to a 1–1.5 standard deviation increase in FB.[22]

The finding that students in a multitrack system do better when their school considers prior performance seems to confirm the hypotheses in this paper: Tracking itself is negatively associated with student performance, but for schools that always consider prior performance in a system with a high number of tracks, the relation between the number of tracks and student performance is either zero or positive. The schools that have entry requirements can select the right students into the available tracks and thus in these schools tracking works most efficiently.

[Table 4 about here.]

[Figure 1 about here.]

6.2 Tracking and inequality

The results for the models of equality of opportunity are displayed in Table 5. As expected and consistent with the literature, (centred) FB has a positive and substantial relation with student performance. This effect is similar over the three PISA test subjects, reading, mathematics, and science. If we look at the interaction between FB and tracking, it can be seen that tracking mitigates the association with FB: The interaction of the number of tracks and FB is negative and highly significant. The more tracks a system has, the lower the effect of FB. The first to third columns show that in a system with five tracks the association of FB is lowered by 17.3 points in reading, 16.2 in mathematics, and 18.2 in science. Tracking thus mitigates the association with FB in systems with a high number of tracks and therefore reduces inequality of opportunity.

From Table 5 one can deduce that the association between socioeconomic status, tracking, and entry requirements on performance is different for students with different FBs. Students with low socioeconomic backgrounds (one standard deviation below the mean FB) benefit from a highly tracked system: Comparing low FB students between systems with one and five tracks shows that students in a five-track system have higher scores for reading, mathematics, and science: 3.0, 20.3, and 17.0 points, respectively. On the other hand, students of high socioeconomic background (one standard deviation above the mean) do better in a comprehensive education system: Comparing their scores in a one-track system against those in a five-track system shows that in a one-track system they obtained higher scores for reading,

[22] The PISA test scores are standardised so that the OECD countries have a mean of 500 points and a standard deviation of 100. The OECD standardises FB on the OECD countries. Here FB has a mean of zero and a standard deviation of 0.94, which means it is almost fully standardised and thus the interpretation of the coefficients can be translated to standard deviation increases or decreases.

mathematics, and science: 23.0, 4.0, and 10.3 points, respectively. A similar trend is seen in the non-linear models in the fourth to sixth columns. High FB students do better in a system with no tracks, while low FB students do better in a system with a large number of tracks.

[Table 5 about here.]

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, I show in Table 6 the results of the same models as in the first three columns of Table 5, but now including also interactions between FB, the number of tracks, and school entrance requirement policies. Table 6 shows that the interaction between the number of tracks and FB is no longer significant; however, the triple interactions (number of tracks, FB, and entry requirements) 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 FB with performance. Thus it is not tracking itself that diminishes the association of FB and performance but, rather, tracking combined with entry requirements. This is consistent with my expectation that when schools use entrance requirements, parents have less influence on their child's track choice and subsequent performance.

[Table 6 about here.]

7 Sorting: Tracking and schools

The results presented above will be biased if the variation used is affected by the selection of students into schools. This may be the case when schools use the selection process to select the best students and not to allocate students into tracks. If schools use entrance requirements in this manner, then 1) these schools are the most attractive for the better students and will therefore experience a higher inflow of better-performing students and 2) these schools may change the behaviour of their upcoming student population since students may have to work harder to be allowed into these schools and may therefore also perform better in later tests.

It is therefore important to look into the characteristics of students and schools that have entry requirements to ensure that these students and schools are not significantly different in other aspects from students and schools that do not have entry requirements. If this is the case, student selection into these schools can influence the results presented above. A first hint that this is not fully the case in my sample is the fact that entrance requirements are not always beneficial for students: In systems with only one track, it is better to be in a school without entry requirements than to be in one with entrance requirements. Therefore, schools with admission rules do not perform better by definition. This section provides further indirect evidence of the validity of my results.

7.1 Descriptive evidence

As already mentioned in section 4.1 (and further expanded on in Appendix A), in distinct countries there might be selection into schools. However, overall schools with and without entry requirements do not seem to differ much on observables. Table 7 shows the correlations between entry requirements and other factors in which students and schools can differ substantially.

The fact that some schools in comprehensive systems have entrance requirements cannot completely be explained by the (perhaps more restrictive) private schools in these systems: The correlation between entrance requirements and school type is 0.09 and excluding these schools from the sample does not change the basic results much. In addition, the correlation between student socioeconomic background and admission rules (0.07) cannot explain the pattern in entrance requirements. Thus it is not true that more selective schools necessarily have the most students from better FBs. As can be seen from Table 7, the correlations between entrance requirements and other school or student characteristics are not large and thus cross country analysis is not affected by selection.

[Table 7 about here.]

7.2 Macro analyses

Table 8 presents (weak) evidence that selection does not drive the results presented in this paper. It displays the results for the same linear models as in Table 4, but now the school-level dummies on entrance requirements are replaced by variables depicting the national percentage of students in schools that sometimes or always have entry requirements. This model excludes the possible selection of students since this micro-phenomenon cannot intervene with the estimation when the entrance requirements are measured at the country level. Unfortunately, the table does not show very significant results, partly due to the reduction in the number of degrees of freedom; however, it does show the same trends as before: For an average country, where 29% of students are in schools that sometimes use entry requirements and where 32% are in schools that always uses entrance requirements, 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 have entry requirements, students perform best in a five-track system, regardless of the subject.[23] For countries with high numbers of schools that never or sometimes have entrance requirements, students perform best with two tracks, also regardless of the subject.[24]

These results seem to indicate that, although selection could be a problem, it is unlikely that it alone

[23] A country with a very high percentage of students in schools that always have entry requirements is defined as a country with 16% of its students in schools that sometimes use entry requirements (the mean minus one standard deviation) and 58% in schools that always use entrance requirements (the mean plus one standard deviation).

[24] The definition of such a country is one where 42% of students attend schools that sometimes use entry requirements (the mean plus one standard deviation) and 6% attend schools that always uses entrance requirements (the mean minus one standard deviation).

drives my results.

[Table 8 about here.]

8 Robustness

A concern with the models in Section 6 is they may suffer from endogeneity. Possible bias due to simultaneity seems a minor issue: The institutional variables that are the focus of this analysis are often constant or similar over large periods of time and it is therefore reasonable to assume that these are not affected by current student performance. However, it is possible that country heterogeneity introduces endogeneity and thus influences the results. To limit the possibility of omitted variable bias, this analysis includes an extensive set of student and school variables. Furthermore, the selection of countries in this analysis comprises, in many aspects, a homogeneous sample of countries, while they differ substantially in tracking policies. However, even with similar countries, country heterogeneity still exists. For this reason, I also present the results from country fixed effects models.

Other possible distorting factors in my results are the measure used to capture tracking, the sample of countries and the control variables. I report all robustness checks with the PISA reading score as the dependent variables, since reading was the main focus in PISA 2009. Unless otherwise stated, the results for mathematics and science are robust to the various checks. As shown below, the main results are robust to a large number of changes.[25]

8.1 Country fixed effects

As 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 admission rules and the number of tracks. All other effects due to country heterogeneity are controlled for. 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 model estimated here can be written

$$Test_{ijk} = \beta_0 + Student_{ijk}\beta_1 + School_{jk}\beta_2 + EntReq_{jk}\beta_3 + EntReq_{jk}\#ofTracks'_k\beta_4 + C_k\beta_4 + u_{jk} + \eta_{ijk} \quad (5)$$

where C_k are the country dummies.

The first three columns of Table 9 show the results for the models with country fixed effects: the first

[25] Most of the small changes are reflected through the change in the coefficient for the dummy indicating that the school always considers entry requirements for admittance. The main result, that entrance requirements mitigate the relation between strong tracking and performance, holds in any specification. All tables not shown in this section are available from the author upon request.

column for the model on student performance and the second and third columns for the models analysing educational opportunities. As shown in the first column, the main effects of the dummies on entry requirements become less significant and smaller but otherwise the results are quite similar. Since this model does not include the main effect of tracking, since now this is replaced by the country dummies, the interpretation is slightly different.[26] In the country fixed effects models it is more difficult to judge student performance for schools that do not consider entrance requirements, since the students in these schools are treated as one group, regardless of the number of tracks available. However, the inclusion of country fixed effects mitigates this to some extent by taking into account, among other things, the effect of the number of tracks in a country. The results with respect to the other categories in tracking and school policies show the same picture as before: For all subjects, students perform best in systems with a high number of tracks when schools always consider prior 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. The second and third columns of Table 9 show that country fixed effects models do not change the estimates much: The interactions are very similar, but the dummy for the Always entry requirements becomes significant and also increases.

8.2 Age of first selection

Unlike other papers on this topic, I use the number of tracks available to students at the age of 15 to characterise 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 these late selection countries may be too weak to be picked up by my analysis. To control for this possible confounding effect, I redo the analysis by using only late trackers.[27] The results can be seen in fourth column of Table 9. The results for the late-tracking countries are robust to some extent, although insignificantly so for mathematics and science.

8.3 Heterogeneous educational opportunities

Table 5 indicates that the differences between students of low and high FB are minimised in a highly differentiated system. To investigate whether the effects of tracking and entry requirements on equality of opportunity are indeed different for students of different socioeconomic background, I estimate models for the subsamples of low and high socioeconomic background students (last two columns of Table 9): For high FB students the results are very similar to those in Table 5. The main effect of FB is a bit weaker and the negative coefficient for the FB–tracking interaction is even larger than before for these

[26] Because the main effect for the number of tracks is not included, the combined coefficients for reading in the model on student performance are very similar to those for mathematics and science.

[27] The late-tracking countries, which select at the age of 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.

students, indicating that the positive relation between their high FB and performance is greatly mitigated by tracking.

The results are less robust to limiting the sample to the low FB students: The negative interaction term is smaller and insignificant. In other words, for low FB students the number of tracks does not alter educational opportunities. However, the triple interactions (FB, number of tracks, and entry requirements) again show that when schools always have entrance requirements, a high number of tracks is beneficial for low FB students.[28] It may be that the parents of the low FB students do not even attempt to influence the school in accepting their child and these students are thus not hindered by any objective entrance selection measure, as the high FB students are. It is also possible that students with a low FB, as opposed to students with high FB, already have an efficient match between their ability and track placement. Future research should look into these possible explanations.

Consistent with the main model, the results on the split sample suggest that systems with a high number of tracks provide more equal opportunity to students, achieved through a more efficient matching of ability and tracking, thus mitigating the relation of FB with performance for high FB students.

[Table 9 about here.]

8.4 Other robustness checks

Including the categorical dummies with the number of tracks reveals that especially countries with only a few tracks perform worse (see also Figure 1). 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. The conclusions stated before are robust to excluding these low-performing countries.

Other countries that can possibly distort the results are Sweden and the United States. These countries officially have a comprehensive system. However, since ability grouping is very common in both countries and often strictly executed, their education systems may more closely resemble systems with multiple tracks than comprehensive systems. To investigate this possibility, I excluded the United States and then Sweden. The results are also robust to this exclusion. Either the independent effects of the United States and of Sweden are too small to make a difference in an analysis with 31 countries or their education systems can indeed be characterised as comprehensive systems, regardless of frequent ability grouping. I also excluded each of the other countries, one by one, to determine if one of these drives the results and found that they do not.

I also adjusted the control variables used in the models on performance. Excluding the student background variables does not change the results much, although the results are less significant. Excluding

[28] The results are not shown. Tables are available from the author upon request.

all the school variables would also exclude the school system interaction terms I am interested in, but I excluded either school composition, school inputs, or other school characteristics; I found no large changes.

I also introduced two other control variables: the interaction between the school standard deviation in FB and the number of tracks and the interaction between the ISCED level and the number of tracks. However, these control variables did not change the results much either and their coefficients are not significant.

The models on inequality are also robust to excluding all the school variables (but keeping the school error term) and to excluding either the school composition, school inputs, or school characteristics variables. When school composition is excluded, the dummy for whether schools always consider entry requirements becomes larger and is always highly significant, while excluding school characteristics lowers the coefficient of the dummy for whether schools always consider entry requirements and renders it insignificant.

To see if differences in FB per country matter, I standardise FB by country, include country dummies, and redo the analysis. The interaction between FB and the number of tracks decreases a little but is still highly significant.

9 Conclusion

The variation in tracking policies in education systems throughout the Western world is quite large: Many countries have no formal tracking, while some countries distinguish up to five tracks for students in secondary school. To study the association between tracking and student performance, this paper uses the data of around 185.000 students in 31 comparable countries from PISA 2009, while taking into account school characteristics such as whether schools consider prior performance when placing students in different tracks. Prior performance can be an important proxy for student ability when students are to be placed in tracks based on ability. Therefore, schools can use entrance requirements on prior performance to help them decide student track placement.

The analyses in this paper show that when tracking in general does not have a direct relation with performance. On the other hand, when interactions between system and school characteristics are considered, students in highly differentiated systems perform best when schools always take into account prior performance to decide student acceptance. In systems with a low number of tracks, admission rules have less of an impact and tracking is only mildly associated with performance. Thus when tracking is consistently executed by schools in a consistent way, it can be beneficial for students, primarily in highly differentiated systems. These results contradictsome of the results in the literature, where no or negative effects are found. However, unlike this paper, those papers did not take into account how schools translate nationwide policy on tracking into everyday practice.

The association between FB, tracking, and student performance shows that equality of opportunity is best provided by a system with a high number of tracks, especially when schools have admission rules. It turns out that for high FB students in these systems, tracking weakens the positive relation between FB and performance, whereas for low FB students, tracking seems to have no effect on this relation.

However, the sorting of students into schools with and without entrance requirements is possible. Although indirect evidence to the contrary is presented, caution is warranted in interpreting the results since the possibility of selection can never be completely dismissed. Future research, using supplemental data sets, should investigate these issues further.

The literature is not conclusive on the effect of tracking (either captured by the age of first selection or the number of tracks available) on student performance. A number of studies find tracking to have negative effects on performance; however, their results are contradicted by positive results and, most often, insignificant results. I argue that it is not straightforward to determine whether tracking in itself has a positive or negative effect on performance. Not only is it important to study the extent of tracking while controlling for potential school-level confounders, but also it is of foremost importance that the execution by schools is taken into account when examining country-wide characteristics such as tracking.

The conflicting results of the previous literature may be explained by the fact that a large portion of this literature does not take into account school characteristics when studying tracking policies. These conflicting results may be reconciled in this paper, which shows that the association of tracking with performance and equality is dependent on school policies.

References

Andreas Ammermueller. Educational Opportunities and the Role of Institutions. *ZEW Discussion Paper*, 05-44, 2005.

Kenn Ariga and Giorgio Brunello. Does Secondary School Tracking Affect Performance? Evidence from IALS. *IZA*, 2643, 2007.

Julian Betts and Jamie Shkolnik. The effects of ability grouping on student achievement and resource allocation in secondary schools. *Economics of Education Review*, 19(2000):1–15, 2005.

Giorgio Brunello and Daniele Checchi. Does school tracking affect equality of opportunity. *Economic policy*, 22(52):781–861, 2007.

Jaap Dronkers. Positieve maar ook negatieve effecten van etnische diversiteit in scholen op onderwijsprestaties? Een empirische toets met internationale PISA-data. *Tijdschrift voor Onderwijsrecht en Onderwijsbeleid*, 6:483–499, 2010.

Jaap Dronkers, Rolf van der Velden, and Allison Dunne. Why are Migrant Students Better Off in Certain Types of Educational Systems or Schools than in Others? *European Educational Research Journal*, 11(1):11–44, 2012.

Esther Duflo, Pascaline Dupas, and Micheal Kremer. Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *American Economic Review*, 101: 1739–1774, 2011.

Allison Dunne. *Dividing Lines: Examining the Relative Importance of between- and within-School Differentiation during lower Secondary Education.* PhD thesis, European University Institute, 2010.

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

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

Roland Fryer. Teacher Incentives and Student Achievement: Evidence from New York City Public Schools. *NBER Working Paper*, No. 16850, 2011.

Thomas Fuchs and Ludger Woessmann. What accounts for international differences in student performance? A re-examination using PISA data. *Empirical Economics*, 32:433–464, 2007.

Eric Hanushek and Ludger Woessmann. Does Educational Tracking Affect Performance and Inequality? Difference-in-Difference Evidence Across Countries. *Economic Journal*, 116:C63–C76, 2006.

Daniel Horn. Age of selection counts: a cross-country analysis of educational institutions. *Educational Research and Evaluation*, 15(4):343–366, 2009.

Caroline Hoxby. Peer Effects in the Classroom: Learning from Gender and Race variation. *NBER Working Paper*, No. 7867, 2001.

OECD. *Science competenties for tomorrows world.* OECD, 2007.

OECD. *PISA Data. Analysis Manual. SPSS.* OECD, second edition, 2009.

OECD. *PISA 2009 at a glance.* OECD, 2010.

Tuomas Pekkarinen. Gender Differences in Educational Attainment: Evidence on the Role of Tracking from a Finnish Quasi-experiment. *Scandinavian Journal of Economics*, 110(4):807–825, 2008.

Tuomas Pekkarinen, Roope Uusitalo, and Sari Kerr. School Tracking and Development of Cognitive Skills. *IZA Discussion Paper*, 4058, 2009.

Bruce Sacerdote. Peer effects in education: How Might They Work How Big Are They and How Much Do We Know Thus Far. In Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, editors, *Handbook of the Economics of Education, Volume 3.* Elsevier, 2011.

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

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

Fabian Walldinger. Does Tracking Affect the Importance of Family Background on Students' Test Scores? 2006. URL `http://www.cepr.org/meets/wkcn/3/3527/papers/Waldinger.pdf`. http://www.cepr.org/meets/wkcn/3/3527/papers/Waldinger.pdf. Last accessed: 05-12-2012.

WorldBank. *World Development Indicators & Global Development Finance Database.* World Bank, 2012. URL `http://databank.worldbank.org/data/home.aspx`.

Table 1: Tracking

Country	Number of school tracks available to 15-year-olds	Percentage of students in schools that consider entrance requirements		
		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) and PISA 2009. The entrance requirements data are from the surveyquestionnaire filled out by the school principal (see text for more details). Weights are used to obtain representative statistics, causing the percentages and absolute student numbers to deviate from each other.

Table 2: Descriptive statistics

Variable	Mean	Std. dev.	Minimum	Maximum	Type	% missing
PISA reading score	493.11	93.52	1.97	817.42	cont	0.00
PISA mathematics score	495.33	94.61	42.85	901.86	cont	0.00
PISA science score	502.88	94.98	10.95	863.24	cont	0.00
Family background	0.07	0.94	-5.34	3.41	cont	0.38
Gender	0.50	0.50	0.00	1.00	dum	0.00
Age	15.77	0.29	15.25	16.33	cont	0.00
Student in (pre-)vocational education	0.16	0.37	0.00	1.00	dum	0.01
Student in upper secondary school	0.47	0.50	0.00	1.00	dum	0.01
Mean school fam. background	0.04	0.56	-2.55	1.61	cont	nvt
SD of school fam. background	0.77	0.16	0.17	1.58	cont	nvt
Percentage in class for whom the first language is not the test language	4.90	1.25	1.00	6.00	cat	2.70
Student–teacher ratio	12.72	6.45	0.39	264.33	cont	8.24
Teacher shortage	-0.14	0.88	-1.02	3.34	cont	2.08
Hinder by a shortage instruct material	1.82	0.83	1.00	4.00	cat	2.30
School responsible for curriculum and assessment	-0.14	0.94	-1.37	1.36	cont	1.56
Index of achievement tracked by authority	0.34	0.47	0.00	1.00	dum	3.36
School type	1.18	0.44	1.00	3.00	cat	3.57
School competition	1.67	0.85	1.00	3.00	cat	2.07
School location	2.88	1.09	1.00	5.00	cat	1.67
School size	659.48	491.29	10.00	6850.00	cont	3.27
Ability grouping	1.78	0.67	1.00	3.00	cat	2.36
Entrance requirements	1.98	0.83	1.00	3.00	cat	2.77
Number of tracks	1.61	1.31	0.00	4.00	cont	0.00
GDP per capita, 2008	$28,856.03	$12,410.14	$13,275.68	$73,349.64	cont	0.00

Source: OECD (2007, 2009) and WorldBank (2012). Variables can be continuous (cont), dummy variables (dum), or categorical variables (cat). The mean and standard deviation of the school FB is constructed from the imputed FB data. The table presents uncentred data.

Table 3: Basic models

Dependent variable	Reading		Mathematics		Science	
	Coef	SE	Coef	SE	Coef	SE
School considers entrance requirements						
Never (ref)						
Sometimes	-2.02	1.95	-1.37	1.89	-1.13	2.06
Always	6.34*	3.37	5.51*	3.07	7.45**	3.76
Number of tracks (0-4)	-2.56	2.55	1.94	3.67	0.75	3.30
N	187768		187768		187768	
(pseudo-)LL	-43761		-43739		-44094	

Notes: The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The control variables are a constant, student background variables (sex, centred age, centred FB, lower/upper secondary education, vocational/general education), school composition variables (centred school FB, centred standard deviation of school FB, percentage of students speaking a different first language), school input variables (centred student–teacher ratio, centred teacher shortage index, extent of hindrance due to shortage in instruction material, centred index of the extent of responsibility of the school for the curricula and assessment), school characteristics (school type, number of schools competing for students, school location, centred school size, whether the school uses ability grouping), and the 2008 GDP per capita. All models include imputation dummies and imputation variable interaction terms.

Table 4: Main models of student performance

Dependent variable	Reading		Mathematics		Science		Reading		Mathematics		Science	
	Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE
School considers entrance requirements												
Never (ref)												
Sometimes	-5.97**	2.54	-4.95**	2.40	-5.99**	2.44	-6.56***	2.53	-4.69**	2.36	-5.83***	2.15
A	-10.13**	4.11	-9.54**	4.25	-10.90**	5.06	-4.97	3.21	-4.04	3.77	-5.47	4.93
Number of tracks (0-4)	-5.74**	2.70	-0.88	3.75	-2.89	3.31						
Sometimes*Number of tracks	2.83**	1.15	2.50**	1.16	3.45***	1.08						
Always*Number of tracks	8.64***	1.82	7.85***	1.90	9.68***	2.04						
2-3 tracks							-32.61***	8.87	-39.10***	10.38	-36.67***	11.77
4-5 tracks							-18.08**	8.32	-2.98	10.23	-10.45	9.33
Sometimes*2-3 tracks							6.11	3.99	3.82	4.11	6.00	4.23
Always*2-3 tracks							9.24	6.83	3.39	6.07	9.58	8.22
Sometimes*4-5 tracks							10.58***	3.95	8.78**	4.18	12.45***	4.11
Always*4-5 tracks							21.79***	5.48	21.92***	5.14	26.74***	6.28
N	187768		187768		187768		187768		187768		187768	
(pseudo-)LL	-43759		-43738		-44092		-43755		-43732		-44088	

Notes: The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. All models include imputation dummies and imputation variable interaction terms. The control variables are equal to those in Table 3.

Table 5: Main models of equality of opportunity.

Dependent variable	Reading		Mathematics		Science		Reading		Mathematics		Science	
	Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE
Family background	23.68***	1.86	24.45***	2.06	24.76***	2.02	26.14***	1.89	27.24***	1.99	27.69***	1.92
School considers entrance requirements												
Never (ref)												
Sometimes	-1.91	1.94	-1.35	1.89	-1.09	2.05	-1.68	2.02	-1.08	1.88	-0.62	2.08
Always	6.72**	3.39	5.74*	3.07	7.75**	3.74	6.87**	3.28	5.76*	2.98	8.24**	3.67
Number of tracks (0-4)	-2.50	2.62	2.03	3.73	0.84	3.38						
Family background*Number of tracks	-3.45***	0.99	-3.23***	1.11	-3.64***	1.09						
2-3 tracks							-30.51***	8.52	-39.43***	10.42	-35.00***	10.99
4-5 tracks							-6.95	7.73	8.40	10.08	3.12	9.42
Family background* 2-3 tracks							-12.50***	2.34	-12.66***	2.56	-14.18***	2.55
Family background* 4-5 tracks							-10.90***	3.29	-10.59***	3.30	-11.33***	3.40
N	187768		187768		187768		187768		187768		187768	
(pseudo-)LL	-43752		-43731		-44085		-43742		-43719		-44074	

Notes: The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The control variables are equal to those in Table 3. All models include imputation dummies and imputation variable interaction terms.

Table 6: Educational opportunities, tracking, and entrance requirements.

Dependent variable	Reading		Mathematics		Science	
	Coef	SE	Coef	SE	Coef	SE
Family background	23.44***	1.91	24.26***	2.10	24.53***	2.07
School considers entrance requirements						
Never						
Sometimes	-2.51	2.01	-1.98	2.01	-1.75	2.06
Always	5.91*	3.35	4.96	3.11	6.90*	3.71
Number of tracks (0-4)	-2.27	2.59	2.23	3.71	1.06	3.37
Family background*Number of tracks	-1.93	1.21	-1.80	1.25	-2.05*	1.22
Family background*Number of tracks*Sometimes	-0.90**	0.37	-1.13**	0.55	-1.13**	0.52
Family background*Number of tracks*Always	-2.49***	0.72	-2.18***	0.72	-2.50***	0.67
N	187768		187768		187768	
(Pseudo-)LL	-43750		-43730		-44083	

Notes: The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The control variables are equal to those in Table 3. All models include imputation dummies and imputation variable interaction terms.

Table 7: Correlations with entrance requirements

	Correlation		Correlation
Student in upper secondary school	0.26	Student–teacher ratio	0.10
School location	0.21	School type	0.09
Student in (pre-)vocational education	0.19	Ability grouping index	0.07
School competition	-0.17	Family background	0.07
Teacher shortage	0.11	School responsible for curriculum and assessment	0.06
School size	0.11	Percentage in class for whom the first language is not the test language	0.03
Mean school fam. background	0.10	Achievement tracked by authority	-0.03
SD of school fam. background	-0.10	Hinder by a shortage instruct material	0.01

Source: OECD (2007) and PISA 2009. This table shows the (weighted) correlations between the categorical variable of entrance requirements and the respective variables.

Table 8: National percentage of schools that sometimes/always consider entrance requirements
(Full sample, i.e., 187,768 students in 31 countries)

Dependent variable	Reading		Mathematics		Science	
	Coef	SE	Coef	SE	Coef	SE
National percentage of schools which consider entrance requirements						
Never						
Sometimes	35.24	39.05	99.29*	54.46	78.61	55.61
Always	−66.67**	30.42	−134.23***	40.72	−86.87*	46.45
Number of tracks (0-4)	5.47	12.54	9.60	17.72	6.40	16.08
Sometimes*Number of tracks	−36.50	23.61	−58.77	37.00	−42.89	32.6
Always*Number of tracks	19.95	18.06	49.89**	22.77	34.62	24.08
N	187768		187768		187768	
(pseudo-)LL	-43758		-43733		-44092	

Notes: The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The control variables are equal to those in Table 3. All models include imputation dummies and imputation variable interaction terms.

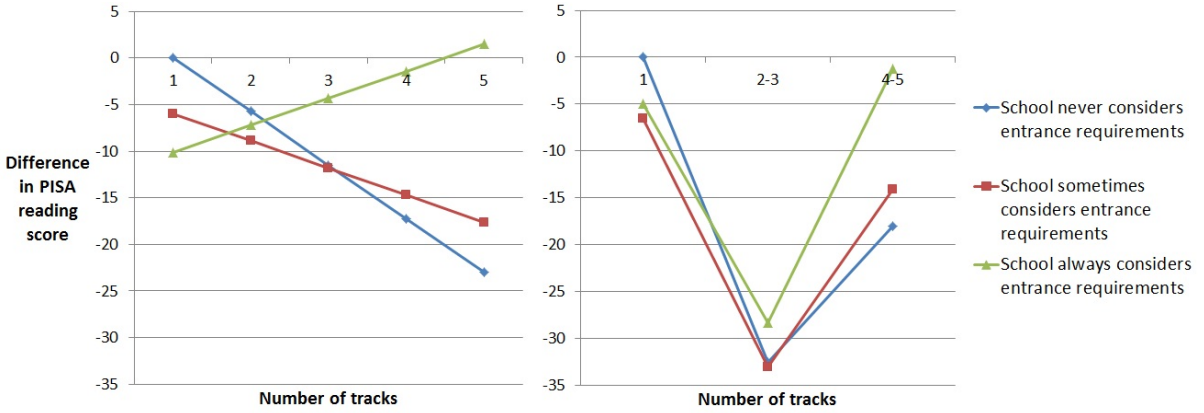
Table 9: Robustness checks, with the reading test score as the dependent variable

	Performance: FE		Equality: FE		Equality 2: FE		Late selection		Low FB		High FB	
	Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE
School considers entrance requirements												
Never												
Sometimes	-3.07**	1.53	0.35	1.11	0.06	1.12	-6.86***	2.20	-2.27	2.04	-0.25	1.74
Always	-3.11	2.22	9.67***	1.32	9.32***	1.33	-7.61	4.78	8.17**	3.24	7.77**	2.98
Number of tracks (0-4)	-	-	-	-	-	-	-7.85**	3.71	-1.89	2.93	-0.71	2.72
Sometimes* Number of tracks	2.36***	0.81					3.32**	1.62				
Always* Number of tracks	6.55***	0.97					6.31**	2.65				
Family background* Number of tracks			-3.46***	0.23	-2.36***	0.35			-1.62	1.21	-3.73***	1.29
Family background* Number of tracks* Sometimes					-0.67*	0.39						
Family background* Number of tracks* Always					-1.84***	0.35						
N	187768		187768		187768		131423		95406		92362	
(p)ll	-43727		-43718		-43717		-28138		-43358		-43877	

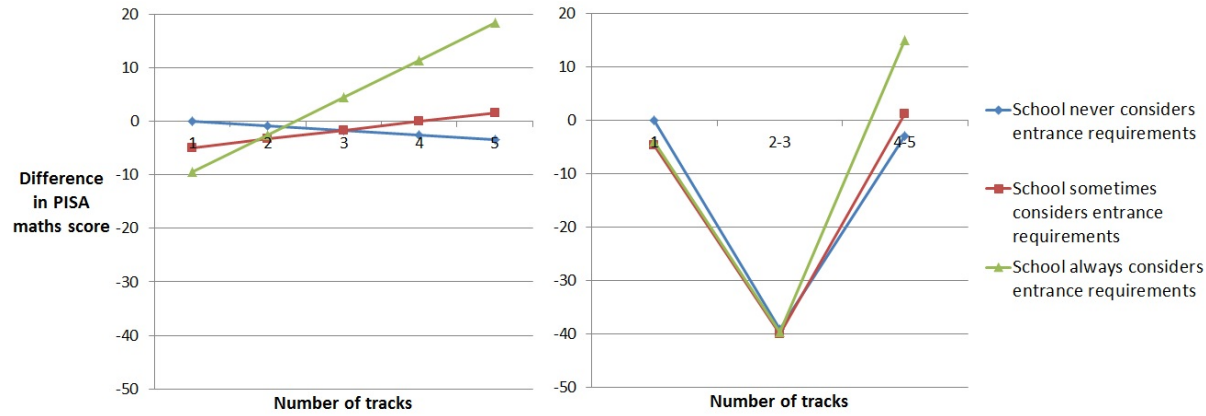
Notes: The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The first column is the same as in Table 4 but now with country fixed effects (FE) and no country error. The second and third columns display the first and fourth columns of Table 5, but now with country fixed effects and no country error. In the fourth column only countries that select students after the age of 14 are considered. In the fifth and sixth columns the sample is divided into low and high FB students, redoing the first column of Table 5. The control variables are equal to those in Table 3, except in the first three columns, where the GDP per capita is replaced by the country dummies. All models include imputation dummies and imputation variable interaction terms.

Figure 1: Differences in student performance for students in different education systems, that attend schools that have different entrance requirements policies. (Relative to students in a comprehensive system in schools that never consider entrance requirements.)

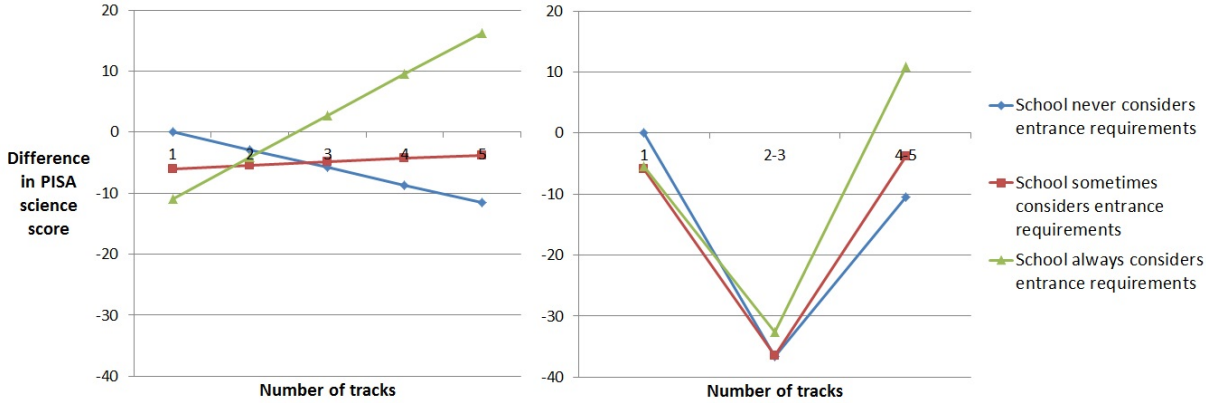
(a) Reading: Linear vs categorical



(b) Mathematics: Linear vs categorical



(c) Science: Linear vs categorical



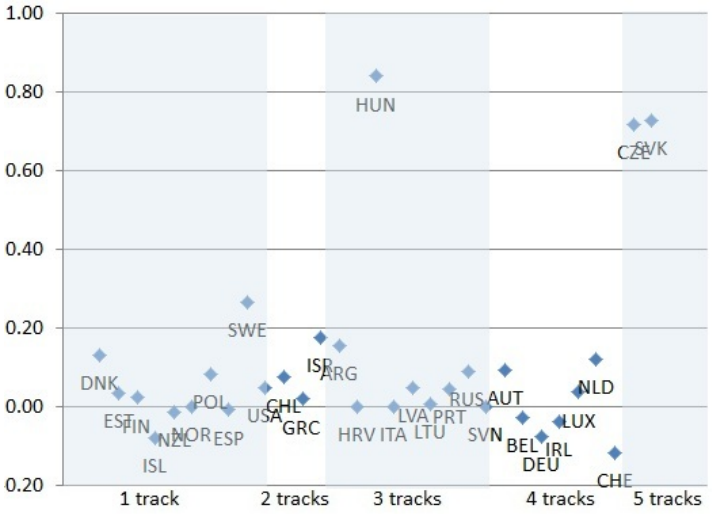
A Entrance requirements and other school characteristics

In each country, the schools that consider entrance requirements are other type of schools. In the figures below, the correlation between entrance requirements and some school characteristics are given per country. In these figures it is clear that countries differ a lot in which type of schools use entry requirements.[29] The school characteristics displayed here are the characteristics for which the highest country or overall correlations exists. Similar figures for the other school characteristics are available from the author upon request.

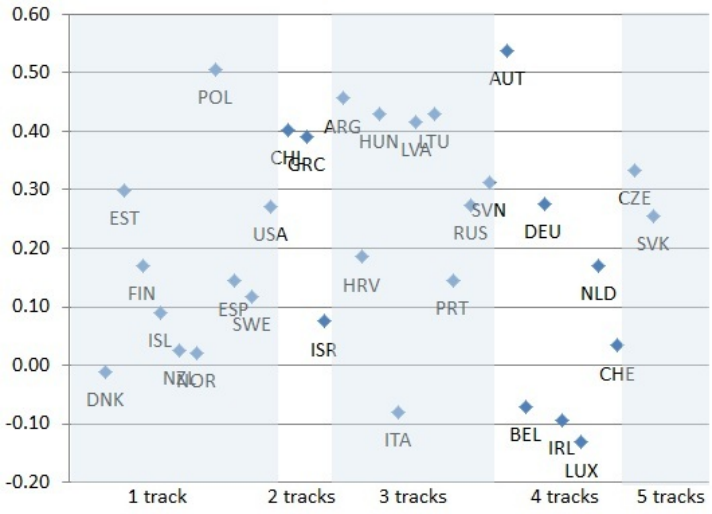
[29] From the top right figure it can be seen that especially schools that have entry requirements in the Slovak Republic and the Czech Republic are upper secondary schools. However, excluding Slovak Republic and Czech Republic from the sample does not alter the results.

Figure 2: Correlations per country between entrance requirements and specified school characteristics.

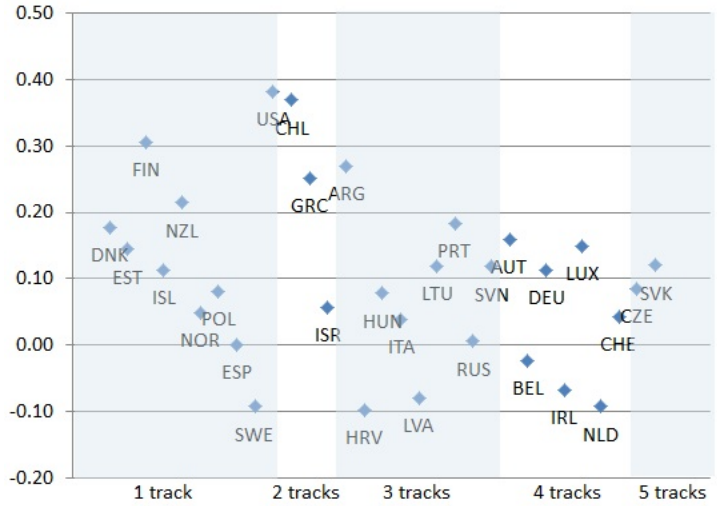
(a) ISCED level. Overall correlation: 0.26



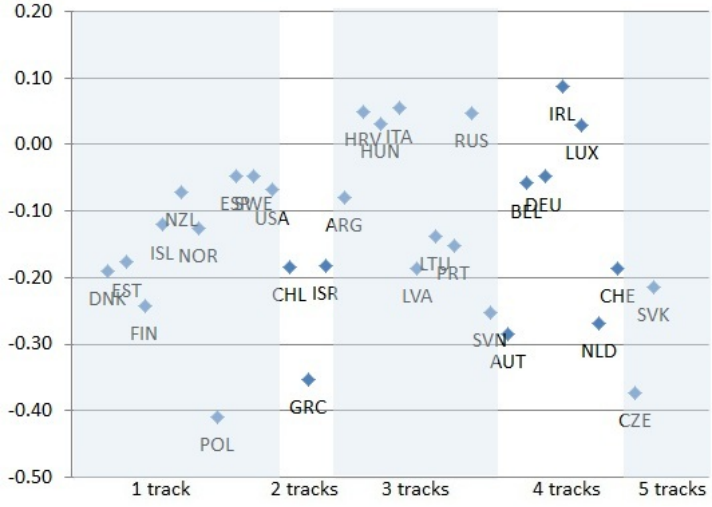
(b) School mean FB. Overall correlation: 0.10



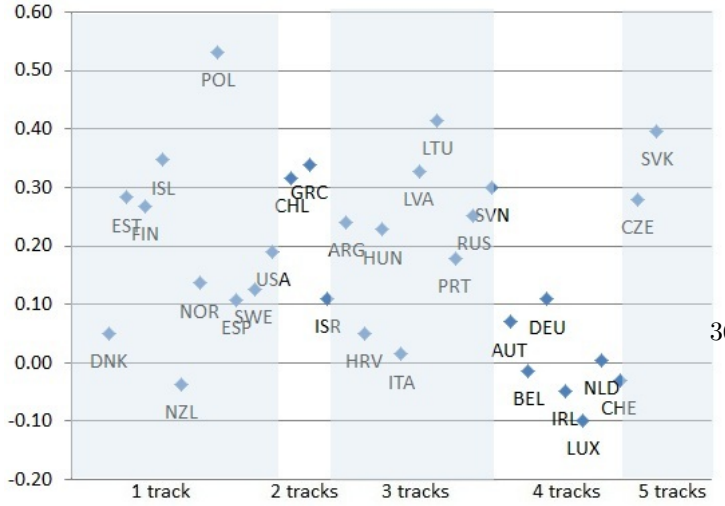
(c) School type. Overall correlation: 0.09



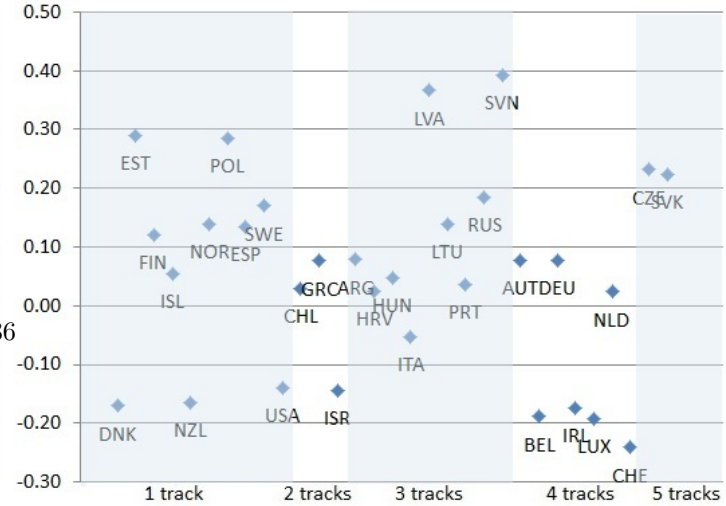
(d) School competition. Overall correlation: -0.17



(e) School location. Overall correlation: 0.21



(f) School size. Overall correlation: 0.11



B Full models without interaction

Dependent variable	Reading		Mathematics		Science	
	Coef	*SE*	*Coef*	*SE*	*Coef*	*SE*
Constant	478.53 ∗ ∗∗	11.12	492.81 ∗ ∗∗	15.88	508.19 ∗ ∗∗	15.32
Family background	18.12 ∗ ∗∗	1.68	19.24 ∗ ∗∗	1.74	18.89 ∗ ∗∗	1.83
Female	35.71 ∗ ∗∗	1.93	−14.76 ∗ ∗∗	1.99	−4.49 ∗ ∗∗	1.74
Age	7.23 ∗ ∗	3.16	4.55	3.37	7.07 ∗ ∗	3.20
Student in (pre-)vocational education	−41.87 ∗ ∗∗	6.57	−39.47 ∗ ∗∗	7.27	−39.72 ∗ ∗∗	8.72
Student in upper secondary school	42.83 ∗ ∗∗	5.21	47.87 ∗ ∗∗	6.44	43.60 ∗ ∗∗	5.70
School average family background	51.71 ∗ ∗∗	4.99	51.62 ∗ ∗∗	5.08	50.60 ∗ ∗∗	5.43
School std. dev. family background	−9.30	7.51	−12.90∗	6.76	−7.63	7.24
Percentage of the students in the school who speak a different language at home						
0% (ref)						
0-10%	1.24	2.57	1.14	2.70	0.83	2.70
10-20%	−0.73	3.54	−0.57	4.39	−2.05	3.90
20-40%	−2.02	4.23	−2.62	3.87	−6.51	4.70
40-60%	−6.95	5.72	−6.55	5.75	−11.21	6.85
≻ 60%	−10.73 ∗ ∗	4.46	−6.45∗	3.61	−11.21 ∗ ∗∗	3.78
Student teacher ratio	−0.34∗	0.19	−0.26	0.20	−0.18	0.17
Teacher shortage index	−2.74 ∗ ∗∗	1.02	−2.92 ∗ ∗∗	0.96	−2.46 ∗ ∗	1.13
The extent of hindrance due to shortage in instruction material						
Not at all (ref)						
Very little	−0.61	1.63	−0.85	1.77	−1.43	1.78
To some extent	−4.00 ∗ ∗	1.72	−4.48 ∗ ∗	1.86	−4.68 ∗ ∗	2.05
A lot	−17.30 ∗ ∗∗	3.46	−16.46 ∗ ∗∗	3.56	−16.68 ∗ ∗∗	3.19
Achievement tracked by authority	0.48	1.11	−0.03	1.11	0.50	1.18
School responsibility for curriculum and assessment	0.07	1.58	−1.41	1.59	−1.82	1.76
School type						
Public school (ref)						
Government dependent school	3.11	4.37	−0.38	4.16	1.35	4.07
Government independent school	−20.30 ∗ ∗∗	6.69	−26.26 ∗ ∗∗	7.13	−19.60 ∗ ∗∗	5.59
Number of schools competing for students						
No other (ref)						
One other	1.24	3.15	−1.94	3.29	−0.11	4.11
More than one other	1.88	3.15	−0.01	2.97	2.14	3.65
School location						
Village (ref)						
Town	−8.24 ∗ ∗∗	2.98	−8.03 ∗ ∗∗	2.80	−9.60 ∗ ∗∗	3.40
Small city	−10.50 ∗ ∗∗	3.59	−12.82 ∗ ∗∗	2.78	−14.77 ∗ ∗∗	3.81

Continues on next page...

Dependent variable	Reading		Mathematics		Science	
	Coef	*SE*	*Coef*	*SE*	*Coef*	*SE*
City	−13.53 ∗ ∗∗	4.44	−18.25 ∗ ∗∗	3.75	−19.92 ∗ ∗∗	5.01
Large city	−18.49 ∗ ∗∗	4.61	−25.15 ∗ ∗∗	4.09	−25.03 ∗ ∗∗	4.74
School size	0.01 ∗ ∗∗	0.00	0.01 ∗ ∗∗	0.00	0.02 ∗ ∗∗	0.00
School uses ability grouping						
For no subjects (ref)						
For some subjects	−7.08 ∗ ∗∗	2.15	−6.39 ∗ ∗∗	2.46	−5.77 ∗ ∗	2.53
For all subjects	−7.21 ∗ ∗	3.04	−8.74 ∗ ∗	4.10	−7.75 ∗ ∗	3.10
School considers entrance requirements						
Never (ref)						
Sometimes	−2.02	1.95	−1.37	1.89	−1.13	2.06
Always	6.34∗	3.37	5.51∗	3.07	7.45 ∗ ∗	3.76
Number of tracks (0-4)	−2.56	2.55	1.94	3.67	0.75	3.30
GDP pc 2008	0.00	0.00	0.00	0.00	0.00	0.00
N	187768		187768		187768	
(pseudo-)LL	-43761		-43739		-44094	

Notes. Standard errors are displayed in brackets. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. All models also include imputation dummies and imputation variable interaction terms. The control variables age, FB, school FB, the standard deviation of school FB, the student–teacher ratio, the teacher shortage index, the index of the extent of responsibility of the school for the curriculum and assessment, and school size are centred to obtain a meaningful constant.