

Tracking students in secondary education : consequences for student performance and inequality

Citation for published version (APA):

Korthals, R. A. (2015). *Tracking students in secondary education : consequences for student performance and inequality*. [Doctoral Thesis, Maastricht University]. ROA. <https://doi.org/10.26481/dis.20150618rk>

Document status and date:

Published: 01/01/2015

DOI:

[10.26481/dis.20150618rk](https://doi.org/10.26481/dis.20150618rk)

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.

Tracking Students in Secondary Education

© Roxanne Korthals, Maastricht 2015

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form, or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior permission in writing, from the author.

This research project has been financed by the Netherlands Organization for Scientific Research (NWO).

Published by ROA
P.O. Box 616
6200 MD Maastricht, The Netherlands

ISBN: 978-90-5321-536-4

Photo cover by Matthew Saum
Printed in the Netherlands by Canon

Tracking Students in Secondary Education

Consequences for Student Performance and Inequality

Dissertation

to obtain the degree of Doctor at
Maastricht University
on the authority of the Rector Magnificus,
Prof. dr. L.L.G. Soete
in accordance with the decision of the Board of Deans,
to be defended in public on
Thursday June 18, 2015, at 12:00 hours

by

Roxanne Amanda Korthals

Supervisors

Prof. dr. Jaap Dronkers

Prof. dr. Lex Borghans

Assessment Committee

Prof. dr. Rolf van der Velden (chairman)

Prof. dr. Eric Hanushek (Stanford University)

Prof. dr. Bas ter Weel

Prof. dr. Herman van de Werfhorst (University of Amsterdam)

Acknowledgements

Many people have helped me during the writing of this thesis, either by directly helping me with my work or by making the time working on my thesis so much fun. I will not thank everybody here, however a number of people deserve special mention.

First, I would like to thank my supervisor Jaap Dronkers, who has given me his fullest confidence from the start. Jaap, working together with you on our many projects always felt very naturally and I would like to thank you for letting me run free when I wanted to and pushing me when I needed a push. My second supervisor Lex Borghans has been invaluable as well. Lex, I would like to thank you for asking such sharp questions and helping me when I was stuck. I cannot believe how much I have learned from my supervisors and how much I have grown as a person under their care.

Long before I had thesis supervisors, I had someone who believed in me and provided me with the confidence that I could succeed. At some points in life you need someone who guides you around, learns you about new paths and is willing to help you get there. I am greatly indebted to Albert Jolink, who supported me to strive for a PhD. Thank you.

I would like to thank Eric Hanushek for giving me the opportunity to stay for five months as a visiting scholar at Stanford University during my PhD. Rick, I thank you for your hospitality and taking the time to discuss my work with me.

Nicolas Salamanca, Olivier Marie, and Eric Bonsang, thank you for answering my questions on numerous occasions.

I would also like to thank my colleagues (and friends!) from ROA and the economics of education group at Maastricht University, those who are there now but also those who have already left. I feel very lucky that I wrote my thesis in a place which has so many people interested in education. I always had someone around who could listen to my ideas, was interested in the topic and provided me with feedback. And our many seminars, lunches, dinners and movies were part of what made my time as a PhD so much fun. Lex and Trudie, you made me feel very welcome in the education group, although officially I was not a part of the group from the start. Thank you. I am also very happy I am not leaving the education group so that I can continue to enjoy its great company and be part of a cooperation that improves all our work.

Mark, I still do not know what I am going to do without you next to me. I could always count on you for a small chat, for endless discussions about work, life and society (preferably all three at the same time), or for exchanging tips for books, movies, series, art, etc. I hope we will succeed to continue as before, even though we no longer share an office.

I would also like to express great thanks to all my others friends and my family for shaping me and having fun together. Matt, thank you for the (un)welcome distractions while finishing my thesis. And least but not last, I would like to thank Phoebe, Pom, Joris, Saar, Ares, Cleo, Jackie, and Minos for providing entertainment, comfort, and joy.

Roxanne Korthals

April 2015

Table of Contents

1. Introduction	1
1.1 Aim of this thesis	2
1.2 Outline and results	3
1.3 Background on tracking	6
2. The Effect of Napoleon Bonaparte on Education Systems and its Consequences: The Education Revolution on Horseback	11
2.1 Introduction	12
2.2 Napoleon and education systems	14
2.2.1 Political pressure from the Napoleonic wars	14
2.3 Empirical support for the influence of Napoleon on education systems	18
2.3.1 Data	18
2.3.2 Results	23
2.4 Can the Napoleonic Wars explain student performance?	27
2.4.1 Data	28
2.4.2 Estimation Method	30
2.4.3 Results	31
2.5 Threats to the identification	34
2.5.2 Changes in tracking since Napoleon	35
2.6 Conclusion	36
3. Selection on Performance and Tracking	39
3.1 Introduction	40
3.2 Literature	42
3.3 Tracking	44
3.3.1 Tracking and schools	45
3.3.2 Tracking and equality of opportunity	46
3.4 Data	46
3.4.1 Tracking and selection by schools	47
3.4.2 Control variables	50
3.5 Empirical strategy	51
3.6 Results	53
3.6.1 Direct relation between tracking and student performance	53
3.6.2 Tracking and performance	54
3.6.3 Tracking and inequality	59
3.6.4 Robustness	62
3.7 Conclusion	63

4. The Effect of Track Placement on Cognitive and Non-Cognitive Skills	67
4.1 Introduction	68
4.2 Data and graphical analysis	70
4.3 Analyses	80
4.4 Conclusion	86
5. The Pre-Tracking Effects of Parental Background	89
5.1 Introduction	90
5.2 The Dutch education system	93
5.3 Data and methodology	95
5.4 Results for the elementary school exit test score	98
5.5 Results for the track recommendation	102
5.6 Conclusion	106
6. Does Early Educational Tracking Increase Inequality? Short and Long Term International Evidence	107
6.1 Introduction	108
6.2 Tracking and relative age	110
6.2.1 Relative age effects	110
6.2.2 Early tracking	111
6.2.3 Relative age effects and tracking	112
6.3 Empirical strategy	112
6.3.1 Conceptual framework	113
6.3.2 Differences in relative age effects between early and late tracking countries	113
6.3.3 Exploiting changes in age of tracking within countries to obtain long term outcomes	114
6.4 Data: PISA and PIAAC	115
6.4.1 PISA	115
6.4.2 PIAAC	118
6.5 Results	119
6.5.1 Short run results	119
6.5.2 Mechanism	122
6.5.3 Long run effects	124
6.6 Conclusions	127
7. Conclusion	129
7.1 Summary of the main findings	130
7.2 Conclusions	131
7.3 Limitations	133
7.4 Future research	133

8. Addendum: Valorization	135
References	139
Appendices	147
A: Summary of the historical case studies of Ramirez and Boli (1987) (Chapter 2)	147
B: The country borders (Chapter 2)	148
C: Historical categorization of the French influence by Grab (p. 18, 2003) (Chapter 2)	150
D: Standardization (Chapter 2)	151
E: Full second stage model of Table 9 (Chapter 2)	154
F: The exclusion restriction (Chapter 2)	155
G: Analyses using data from 1970 (Chapter 2)	158
H: Nonlinear models (Chapter 3)	162
I: Comparing different models for reading and science (Chapter 3)	164
J: Subsample of early tracking countries (Chapter 3)	166
K: Outcome variables (Chapter 4)	167
L: Other model specifications (Chapter 4)	170
M: Including controls for gender and parental background (Chapter 4)	173
N: Representative sample (Chapter 5)	176
O: Results without weights (Chapter 5)	177
P: First stage results from Table 2 and 3 (Chapter 5)	179
Q: Boys versus girls (Chapter 5)	180
Biography	181
ROA Dissertation Series	183

Introduction

No child is alike. Each child has a distinct personality and has its own talents and skills. Teachers in school will therefore try to approach each child in a different way to ensure the child will be able to learn optimally and develops the necessary non-cognitive skills. In all education systems and in all school types some form of grouping students is done, to the very least by grouping students together based on age and grade. When within grades students are formally separated based on ability level, this is called tracking, or formal differentiation. Tracked systems often consist of students of different tracks to be placed into different building or even different schools. The different tracks in which students are placed have different curricula tailored to the ability levels of the student population. This adjusted curriculum, combined with a more ability homogenous class which should facilitate teaching, is thought to enhance learning outcomes.

Notwithstanding the mentioned theoretical benefits, the use of tracking in secondary education systems is a contested practice in western society and academia. Questions which are often raised are (1) Does tracking increase the average cognitive development? and following from that (2) Does tracking increase inequality among students? (OECD, 2010). Opponents argue that allocating the more able students to the high track, will leave the less able students without the more able peers to learn from. If this is the case, any gain achieved by the high ability students could be at the expense of the low ability students, increasing inequality in outcomes and perhaps even lowering average performance (Hanushek and Woessmann, 2006). Besides lower ability peers, students in the low track could also potentially lose out since they receive a less challenging curriculum, although it is designed for their ability level, and the lower ability school might receive less resources (in terms of facilities, or lower educated teachers). The second question also arises from the fact that specific groups of students, for instance students of lower parental background or the relatively young students in the class, are more likely to be allocated to the low track (e.g. Dustmann, 2004, Muehlenweg and Puhani, 2010). When track choice has long run consequences this might further disadvantage these students.

1.1 Aim of this thesis

This thesis provides evidence for the two questions asked above: Whether tracking has a positive effect on student learning and whether tracking increases inequality within education systems and societies. It thereby aims to provide the reader with a deeper understanding of tracking and its effects on student performance and inequality. To start, Chapter 2 provides direct evidence on the first question looking at the effect of tracking on student performance. The next two chapters look closer at the performance differences between tracks and tracking regimes. The aim of chapter 3 is to consider

how implementation of tracking by schools can influence the outcomes of tracking. It does so by looking at whether using prior performance to decide on track placement is related to better outcomes, both in student performance and inequality. Chapter 4 looks at whether the student on the margin of going to the low or high track is better off in terms of cognitive and non-cognitive outcomes in the high track. Finally, the second question is addressed in Chapters 5 and 6. These chapters look at the effects of tracking on two types of inequalities: those caused by parental background and those caused by relative age within the class.

1.2 Outline and results

Chapter 2 starts off by providing a deeper understanding of tracking by presenting evidence on whether tracking increases the average cognitive development of students. Previous studies have looked at this question before but this has led to conflicting results (e.g. Hanushek and Woessmann, 2006, Ariga and Brunello, 2007, Jakubowski, 2009, and Van Elk *et al.*, 2011). Analyzing the effect of tracking on student performance is problematic since all the variation is between countries and much less variation exists within countries. Less within country variation exists since tracking is often a national or statewide policy leaving no room for specific schools or regions to deviate from this policy.¹ If only cross country variation exists then any country differences will confound the analyses since it is impossible to disentangle the effects due to tracking from effects due to other country characteristics. To obtain unbiased estimates of the effect of tracking on educational outcomes, this endogeneity has to be taken into account either by longitudinal individual level data or by statistical methods to remove it. In Chapter 2 of this thesis one of the statistical methods to remove endogeneity is used. It uses an instrumental variable approach to be able to estimate the effect of tracking on student performance. The employed instrument is the political pressure by Napoleon in the 19th century. This instrument has a relation with the extent of tracking in different European countries, and the results show that tracking has a positive effect on student performance at age 15. However, previous research has successfully used related instruments for other institutions which cast doubts on the exclusion restriction and the causal interpretation of the results.

Chapter 3 of this thesis examines whether the implementation of tracking influences student performance and inequality. In a tracked system students need to be allocated over the different tracks. Since the main differences between tracks are on the content and level of the curriculum, it seems obvious that students are also allocated to the

1. Exceptions are variation between states/*Länder* or possibilities to postpone tracking in some tracked system. In Germany the *Gesamtschule* provides such an option and also in the Netherlands students can postpone strict tracking by attending bridge classes in the first or first two years of secondary education.

tracks based on ability. However in some countries school principals select students based on religious or artistic criteria, or parents have a large say in the track placement of their children (e.g. Dustmann, 2004, Barg, 2013). Chapter 3 of this thesis makes use of a question answered by school principals on whether they consider prior performance on accepting the student to the school. From the results it follows that in a system with four or five tracks, tracking is positively related to student performance if students are indeed selected based on performance. When school principals do not consider prior performance when accepting the student, students do not benefit from tracking. A further finding is that the influence of parents on student performance is lower when prior performance is always considered. This logically follows from the use of prior performance for track placement which often replaces parental preferences as the leading criterion for track placement.

One possible downside of tracking is that it could aid the students in the high track at the expense of the students in the lower tracks. Before looking into inequalities for specific groups, it is useful to study the underlying assumption on which this inequality is based: Individual students in the low track would perform better if they would be in the high track. Chapter 4 of this thesis studies this question by making use of the marginal student, the student who is on the border of being allowed or not to go to the high track. Looking at the marginal student allows us to look at the effect of being in the higher track without having the confounding factor of a changing ability composition, which is the case if a large number of students would move from the low to the high track. In the Netherlands, at the time of this research, secondary schools decide on track placement based on an elementary school exit test and a track recommendation of the elementary school teacher. Using these two assignment variables we are able to use the threshold for when students are allowed to go to the high track as a discontinuity in the assignment. For the marginal student, being placed in the high track improves reading and IQ scores and the perceived probability to finish the degree. These improved outcomes are reached solely from being in the high track, without changing the ability composition of the higher track. The characteristics of the high track, consisting of better able peers, a different curriculum, and more, cause the marginal student in the high track to do better than if the student would have been in the low track. Being in the high track does not improve mathematics scores or personality traits, as the Big Five or motivation. Although the methodology only allows us to look at the marginal student, for this student and the student's reading and IQ scores and the perceived probability to finish the degree it does matter whether the student is placed in the low or high track.

The last two chapters finish the analyses of tracking and its consequences for inequality by looking at the effects of parental background (Chapter 5) and at whether

tracking influences the effects of relative age (Chapter 6). If track placement is done on prior performance students in the low track will be students with lower ability. However, often students in the low track are also students of lower parental background and the relatively younger students. Students with lower parental background perform worse in school due to, among other things, worse health, nutrition, parental support, and social capital (see for an overview Jackson, 2013). As a consequence of this lower performance they are more often placed into the low track. But students of lower parental background are also less likely to go the high track since less is expected from them by teachers or teacher respond differently to students with higher educated parents (e.g. Lavy and Sand, 2015, Barg, 2013, Jussim and Harber, 2005). That students with lower parental background have lower educational outcomes before and after tracking has taken place, and are more likely to go to the low track is well documented. The innovation of Chapter 5 is that it looks at whether there is an additional effect of parental background on track placement over and above the influence of parental background on performance in general. If this is the case, then higher educated parents are able to enlarge the advantage their children have in a tracked system as compared to in a comprehensive system. The research question is answered by investigating the effect of parental background on the two assignment variables for track placement in the Netherlands. These two assignment variables are the same as the ones used in Chapter 4: an elementary school exit test and a track recommendation of the elementary school teacher. The results show that parental background indeed has an effect on both the assignment variables and that it is mostly the high educated parents who are able (consciously or not) to generate better outcomes for their child on the elementary school exit test and the teacher recommendation.

Chapter 6 considers the effects of tracking on a different type of inequality, the inequality caused by month of birth. Just like children of low parental background, since relatively young students perform worse than relatively older students when the track placement decision is made, relatively younger students are more likely to go to the low track (e.g. Jorges and Schneider, 2007; Muhlenweg and Puhani, 2010; Dustmann *et al.*, 2014). Within each class students of different ages are grouped together: The youngest and oldest differ up to 11 months in age. At early ages this difference has been shown to largely explain differences in non-cognitive development (see for an overview Stipek, 2000). The effects on outcomes are caused by absolute age differences at start of schooling and relative age difference within the class. Two potential channels for the effect of relative age are that due to less maturity at school start, the relatively young students develop an early performance lag which might take time to catch up. This early performance lag might also affect the school motivation of the relatively young causing more long term disadvantages. As a consequence of this

relative age effect, relatively young students are more likely to be deemed lower ability students and thus are more likely placed in the low track. At later ages the relative age effect is smaller since the age difference of 11 months is relatively smaller for 15 year olds than for 10 year olds (Crawford *et al.*, 2010). Chapter 6 makes use of cross country variation in the age of selection into tracks and shows that early tracking indeed leads to a stronger relative age effect. However, it also leads to different ability distributions in the low and high track between countries that track early and those that track late. Countries that track late are better able to separate low and high ability students in the low and high track than countries that track early. Therefore, in countries that track early the low track is more heterogeneous in ability. Over time, relatively young students in the low track in countries that track early tend to outperform the relatively older students because at later ages the relative age effect has worn off and their true ability level is revealed. This advantage for the relatively young students is a disadvantage for the relatively old student: They earn less in adulthood and even are more likely to belong to the bottom ten percent earners. Due to the longer lasting relative age effects, countries that have early tracking have higher levels of inequality based on month of birth at later ages.

Chapters 2 to 6 each contribute to answering the two questions this thesis aims to answer: (1) Does tracking increase the average cognitive development? and following from that (2) Does tracking increase inequality among students? Chapter 7 concludes and provides directions for further research. Each of the Chapters 2 to 6 can be read independently of the others and each provides background information on the relevant aspects of tracking students in secondary education. For those new to the topic, the next section provides a basic background on tracking.

1.3 Background on tracking

Tracking is a system wide education characteristic, in the sense that all schools in a system (often comprising a country or state) adhere to it and how tracking is conducted is decided upon at the national or state level. In current day western education systems, tracking only takes place in (post-)secondary and tertiary education (OECD, 2010). No western country tracks students in elementary school. Although all tracked systems formally group students into different educational programs, other differences between tracked systems exist (OECD, 2010). In some countries children are separated into tracks at the age of 10 (Germany and Austria), while in other countries children are tracked later, for instance at age 16 (Sweden and the US).² In a tracked education

2. In the twentieth century a number of countries changed the tracking regimes in their education systems: they either postponed the age of tracking or they lowered the age at which tracking takes place. Examples of such changes are Sweden, and England who moved to tracking at age 16 in the sixties, while in the eighties in

system, two or more different tracks are available for students with the number of available tracks being dependent on the system.³ The curriculum differs between tracks, with some tracks offering solely vocational education; while at the other extreme tracks can consist of solely high ability level general education in so called grammar schools or gymnasiums (OECD, 2010). Since the number of tracks differs between countries and also the curriculum content is different among tracks, tracks cannot be easily compared across countries.

In some countries students of all ability levels are kept together in one class, in other words no tracking takes place. Education systems that do not track students are called comprehensive systems. In these systems schools might choose to still group students based on ability using ability grouping or streaming (OECD, 2010). Ability grouping is when students within schools are grouped on ability, for instance by allowing some students to participate in advanced placement classes or university preparatory classes. Streaming is when students are grouped within classes based on ability, for instance, by providing extra challenging assignments for those students who finish their course work early or by letting lower ability students use textbooks from previous years. Some examples of studies which look at ability grouping are Betts and Shkolnik (2000), Figlio and Page (2002), and Duflo *et al.* (2011). Ability grouping and streaming are not considered in the thesis. Comprehensive education systems are studied in this thesis insofar as they act as the counterpart of education systems that do have tracking.

Tracking students in secondary school creates a different environment for students than if students are kept together as they are in a comprehensive system. Three main differences exist between a tracked education system and a comprehensive education system: A tracked education system will have more homogenous classes, different curricula between the tracks, and it might have unequally allocated resources between the tracks. When classes are more homogenous in ability, teachers might find it easier to teach the class. To ensure most students fully comprehend their teacher, teachers logically adapt their teaching to the average student in their class. In teaching practices where a lot of centralized teaching time is required, teaching a homogenous class will allow more students to learn since students are closer to this average ability. In teaching practices where only a limited amount of classroom teaching is required, a homogenous class might still be positive for the teacher since the teacher will have to adapt less to each student that needs an explanation.

Flanders Belgium a policy shift to an earlier tracking age was made, essentially increasing the amount of differentiation (Braga, Checci, Meschi, 2013). Changes in the tracking regimes often followed political and societal debates around similar questions as analysed in this thesis.

3. The use of special needs education is often not considered a form of tracking and is therefore also excluded in the analyses in the thesis.

Homogenous classes will ensure that the student's peers are more similar in ability. Besides more homogenous peer groups, if tracking is done on ability, it will also lead to the removal of the very bright peers for the students in the lower tracks and the removal of the not so bright students in the higher tracks. The average peer ability will therefore be higher in the higher tracks than in the lower tracks. Since ability and parental background are highly correlated, tracking on ability also means to some extent tracking on parental background. So besides a more homogenous peer group in ability, tracking will possibly lead to classes being more homogenous on parental background. This peer composition can have effects on student outcomes, and is one of the channels through which the influence of tracking can be seen. Sacerdote (2011) provides an overview of all the different peer effects: From the linear peer effects which averages out across tracks, to the "shining light"/"bad apple" which point to potential beneficial/detrimental effects of very good/very bad peers respectively. Which peer effect dominates is an empirical question not addressed by this thesis. However, it should be noted that also in comprehensive systems homogenous peer groups can form. This can either be due to ability grouping or streaming, caused by religious criteria or by spatial segregation.

Two other differences between a tracked system and a comprehensive system are the differences in curriculum caused by different ability levels and the potential differences in resources. In a tracked education system each track will have a different curriculum, often ranging from more vocationally orientated courses to more general courses. A different curriculum could also have the form of a slower pace at which the same curriculum is given to students. Provided the curriculum is designed with the average ability level of the students of the track in mind, the different curricula should enhance learning in all tracks.

Differences in resources could arise when more (governmental or private) budget is allocated to specific tracks. It could either be that more resources are allocated to the low track since the students in the low track need more assistance, or it could be that more resources are allocated to the high track since it is more prestigious. However a difference in resources is not necessarily solely based on monetary resources, but could also be due to regulations, for instance on teacher credentials. Related to the previous point, if the high track is accompanied by a more demanding curriculum as is often the case, education authorities might require the teachers in that track to have a higher degree than teachers in the other track. This will lead to higher educated teachers in the high track as compared to the low track. It will also raise the costs of teacher salaries in the high track which will lead to a net decrease in resources for other inputs. It remains an empirical question whether more resources are indeed allocated to specific tracks. However, some papers indicate that resources have little to

no effect on educational outcomes (Hanushek, 1986; 1997) in which case this channel of tracking would not be expected to be very important.

The Effect of Napoleon Bonaparte on Education Systems and its Consequences: The Education Revolution on Horseback⁴

Abstract: Sociological literature argues that the threat for war with and invasion by the French around the 1800s induced European countries to introduce mass public education systems. On the basis of this theory, the aim of this chapter is twofold: First, it empirically establishes whether the proposed mechanism had an influence on the levels of tracking and standardization in European education systems. Second, it studies whether tracking has an effect on student performance by instrumenting tracking by the political pressure caused by the Napoleonic wars. Especially for tracking, a relation between education systems and the pressure from Napoleon is found. For standardization, no or a much weaker relation is found. Using Napoleon as an instrument for tracking leads to a strong first stage. The resulting analyses show a consistent positive effect between the level of tracking and student performance.

4. I would like to thank Jaap Dronkers and Sjoerd Karsten for bringing the paper of Ramirez and Boli (1987) under my attention, and the Forschungsdatenzentrum am Institut zur Qualitätsentwicklung im Bildungswesen for providing me with the extended PISA data for Germany. I further would like to thank Francisco Ramirez, David Figlio, Bas ter Weel, and participants of the International Workshop on Applied Economics of Education 2013, the Workshop on Comparative systems of educational and political systems at Stanford University, and ESPE 2014 for valuable comments.

“Our view is that European states became engaged in authorizing, funding and managing mass schooling as a part of the endeavor to construct a unified national policy. [...] External challenges [...] were important stimuli to state action in education [...]” (Ramirez and Boli, 1987, p. 3)

“We show that state educational systems did not originate in the most dominant countries, contrary to the widespread assumption that educational innovations inevitably start there [...] Rather, the most dominant powers were able to resist the system wide pressures favoring mass education [...].” (Ramirez and Boli, 1987, p. 4)

2.1 Introduction

Looking at the education systems in Europe (see Figure 1 in Section 2.2), patterns can be distinguished. For instance, it is easily seen that the Scandinavian countries and the UK have very similar comprehensive education systems with only one track available to students and late selection. Contrary to that, Germany, the Netherlands, Belgium, Switzerland, and Austria have highly differentiated systems with a high number of tracks and, in Germany and Austria, selection as early as 10 years old. The geographical spread of standardization of inputs (administrative standardization and content related standardization) and outputs (central exit exams) is less clear. Ramirez and Boli (1987) propose a theory that points to an explanation for the geographical spread of differentiation found in European education systems: the threat of war with and annexation by the French in the 1800s induced European countries to create mass public education systems. This chapter investigates whether the proposed mechanism had an effect on levels of tracking and standardization in European education systems and what the implementations of this are.

The aim of this chapter is twofold: First, it takes the theory by Ramirez and Boli (1987) and empirically tests whether the proposed mechanism also had effect on levels of tracking and standardization in European education systems. In Section 2.2 it will be shown that the pressure from France influenced European educational systems and that this influence can still be seen today, although not for all aspects of education systems. The political pressure from France, as theorized of Ramirez and Boli (1987), is related to the level of differentiation, both seen by the number of tracks available to fifteen-year-old students and by the age of first selection. The mechanism seems to have had less influence on content related standardization, and none on central exams and administrative standardization.

The second aim of this chapter, discussed in Section 2.3, is to explain current student performance by building on the theory by Ramirez and Boli (1987) and using the

historical variation in education systems. The pressure from France is used as an instrument to explain education system characteristics to alleviate the endogeneity between education systems and student performance. Naturally, the French influence in the 1800s influenced many aspects of European societies and thus these analyses should not be regarded as pure causal analyses, but as an alternative to existing studies to describe education systems and its consequences. Section 2.5 provides more information on possible other influences of Napoleon and other threats to the identification. Given the considerations there, tracking has a strong first stage and positive effects are found for the number of tracks in a country. Negative, but insignificant, effects are found for the age of first selection.

In the 18th and 19th century most European countries passed their first widespread compulsory schooling laws and centralized education systems were set up. Before this period, schooling was not available to all and the educational institutions were in non-state control, controlled especially by the church(es) and local authorities. The Enlightenment in the 18th century paved the way for mass public education, since schooling would ensure advancement of knowledge so sought after by this movement. However, it was other, more political, events which induced the actual introduction of the educational systems in Europe. The late 18th century started, partly as a consequence of the political success of the Enlightenment in France, a volatile period in European history, best seen by the French Revolution of 1789 and its impact on the continent. After Napoleon claimed power in France in 1799, the Napoleonic wars dominated the European continent for almost two decades. The French army conquered large areas of Europe until Napoleon was weakened by the failed Russian invasion in 1812 and finally defeated in 1815 in the battle of Waterloo. In the period in between, almost all European powers were invaded or threatened to be invaded by Napoleon.

As laid out by Ramirez and Boli (1987), the threat of war with and annexation by the French induced European countries to create a national identity among their population. An important part of this nation building was the introduction of mass public educational systems by the national state, intended mainly to educate the future generations in the cultural and political traditions of the nation. Creating a national identity in war times is important for states since it increases the public's (physical and mental) willingness to pay and fight for the survival of their own nation in order to remain independent.

This chapter is structured as follows: The next section gives an introduction to the history of French influence in the 1800s and elaborates on the theory by Ramirez and Boli (1987). Section 2.3 presents evidence on the influence of Napoleon on a number

of education system characteristics. Section 2.4 uses the proposed theory by Ramirez and Boli (1987) to employ an instrumental variable method to alleviate the endogeneity between education system characteristics and student performance. Section 2.5 gives attention to the exclusion restriction. Section 2.6 summarizes and concludes.

2.2 Napoleon and education systems

This section first gives a short historical overview of Napoleon and the Napoleonic Wars. Then it provides a discussion of the theory proposed by Ramirez and Boli (1987). This theory relates external pressure to the introduction of mass public education systems. The different stages of this process are discussed, and examples are given. Finally, using data on education systems, the relation is shown between political pressure from Napoleon and education systems.

2.2.1 *Political pressure from the Napoleonic wars*⁵

The external challenges, referred to by Ramirez and Boli (1987), were in most cases (political) pressure arising from the French army, and later and foremost from Napoleon. After the turmoil of the French Revolution and its aftermath, Napoleon Bonaparte, a general who just secured victories for France in Italy, Austria, and Egypt⁶, was called in to support a coup d'état by the National Convention, consisting of the bourgeois who wanted to overthrow the everlasting power of the clergy and the nobility. However, Napoleon took power for himself in November 1799 and declared himself First Consul in 1800.

Almost as soon as Napoleon took power, he started invading neighboring countries and annexed Piedmont in Italy, the Rhineland in Germany and Belgium, appointed himself president of the Republic of Italy and Mediator of the Swiss Confederation. In 1802 the peace treaty of Amiens was signed by France, England, Holland and Spain stopping combat and reaffirming the dependence of the earlier conquered states to France. But the peace did not last long as Napoleon kept intervening in neighboring countries and setting up strategic positions against England, the most dominant (naval) power at the time. Soon after his coronation to Emperor of France in 1804, Napoleon expanded his scope and also annexed the Ligurian Republic (1805, Italy, south of Piedmont). In response, in 1805 the Russian Empire and England formed an alliance

5. This paragraph is based on Santon and MacKay (2010) and the first and Chapter 13 of Grab (2003), *The formation of the Napoleonic Empire and The collapse of the Napoleonic Empire*.

6. Napoleons journey to Egypt did not end as well as it started. After Napoleon conquered Egypt, and with it was able to frustrate the trade routes of England, he lost most of its fleet in the Battle of Aboukir and was unable to return home. Napoleon ruled Egypt for over about a year before leaving behind his army and returning to France in 1799.

against France and its expansionary urges, which Austria (the Habsburger Empire), Sweden, and Naples later also joined (see Figure 2 for a division of Europe in the different empires). Napoleon formed alliances with Spain and Bavaria in return. Despite the alliance with Russia and England, Austria was conquered and was given harsh settlement terms and had to give up a number of territories. France became the dominant power on the continent.⁷

France was also able to force Prussia into submission: Prussia had to end trade with England but received the conquered region of Hanover from France in return. However, Prussia sided with Russia and sought combat with France but was defeated in 1806. Napoleon conquered Berlin and later also invaded Prussia's Poland and established the Duchy of Warsaw. On his way east, Napoleon met the combined forces from Prussia and Russia in 1807 but overcame. Russia had to acknowledge Napoleon's brothers as rulers of Naples, Holland, and Westphalia and in return Napoleon would leave the Baltics states and Turkey alone. Russia took its chance and occupied Finland from Sweden in 1808. Prussia had to give up territories to the Kingdom of Westphalia and the Duchy of Warsaw, which were under Napoleons influence, pay damages, and accept limitations on its army. Now also Germany and the east were under Napoleons control and Russia had given up its resistance to the French.

In 1808 while the French army was mostly tied up in Spain, Austria started to make preparations to go to war against France. The war was fought in southern Germany and Austria and in Italy and Poland. Both sides won and lost battles, and even though England came to Austria's defense in the Netherlands, in 1809 Austria had to concede to France once again. France received the Illyrian Provinces ("Croatia"), while also Bavaria and the Duchy of Warsaw received lands from Austria. Austria had to pay damages and had to obey limitations on its army. Later Napoleon married the daughter of Francis I, the emperor of Austria, and thus formed an alliance between both houses. Pope Pius VII opposed the marriage, as he had not approved Napoleon's divorce from his former wife Josephine, and also refused to prevent England ships from using its ports, and thus Napoleon marched south to Rome. In 1809 Napoleon had taken over control of Rome and the remaining Papal States and exiled Pope Pius.

Under Napoleons reign the French Empire grew substantially: At the height of the French Empire in 1810 it covered an area of 750,000 square kilometers (293,000 square miles) as compared to 543,965 square kilometer (210,026 square miles) for Metropolitan France today. But not all territories were under French control. There

7. Although Napoleon had less luck overseas, which can be illustrated by his defeat over England at the (sea) Battle of Trafalgar in 1805.

were “allied countries whose territory was expanded by Napoleon in some cases, and continued to be governed by their native rulers”, like Bavaria and Sweden. Grab (2003) calls these the *pays allies*. The conquered or annex countries were either “countries annexed to France and directly ruled by Napoleon” (*pays réunis*), like Belgium and Piedmont, or “satellite states that were entrusted to French rulers” (*pays conquis*), like Spain and Switzerland. Figure 3 shows how Europe was divided under Napoleons reign.

Napoleon became less popular in France over the year with the continued fighting, the exile of Pope Pius VII in 1809, and when due to the blockade of England and other countries opposed to Napoleon a shortage of materials arose. Also the alliance with Russia proved to be weak. When Russia refused to keep up the blockade towards England, both Russia and France made preparations for war. Alliances were sought and soldiers were readied. Napoleon advanced into Russia but lost many men along the way due to heat, hunger, and disease. Napoleon took Moscow in 1812 but in doing so the supply lines to the front line became too long and supplies grew thin while the cold was setting in. Napoleon retreated but lost even more men along the way back, this time due to cold and Russian troops.

Napoleon returned to Paris to raise another army. This time however, Russia, Prussia, England, Sweden and Austria united against France, and also France’s long-time ally Bavaria reluctantly joined this coalition later on. In Spain and Italy, Napoleon also faced problems and he finally abdicated when Paris was taken in April 1814. Napoleon was sent to Elba but he came back to Paris in 1815 for the Hundred Days until he was finally defeated in the battle of Waterloo. Napoleon was exiled to St. Helena and died there in 1821.

2.2.2 *A theory on the origins of education systems*⁸

In their 1987 paper *The Political Construction of Mass Schooling: European Origins and Worldwide Institutionalization* Ramirez and Boli present their theory on how external pressure led to similar events in multiple countries which eventually led to the introduction of mass public education systems in the nineteenth century. Their theory can best be described by the two quotes at the beginning of the chapter of which the first is: “Our view is that European states became engaged in authorizing, funding, and managing mass schooling as a part of the endeavor to construct a unified national policy. [...] External challenges [...] were important stimuli to state action in education [...]” (Ramirez and Boli, 1987, p. 3). Napoleon can be seen as one of the most pronounced external challenges of this era.

8. This section is based upon Ramirez and Boli (1987).

Ramirez and Boli (1987) illustrate their theory by describing this process in seven European regions: Prussia, Denmark, Austria, Sweden, Italy, France, and England. A summary of these case studies is provided in Appendix A. Sweden, France, and England experienced a general call from within for mandatory, universal and free education later than other countries. Prussia and Denmark already made early moves towards education for all. Italy is an interesting case since the South acted like France and England, while the North made, like Prussia and Denmark, earlier steps towards a mass public education system. This supports the theory by Ramirez and Boli (1987) that regions that were under external pressure (by Napoleon in most cases) were induced to introduce mass public education systems while “the most dominant powers were able to resist the system wide pressures favoring mass education [...]” (Ramirez and Boli, 1987, p. 4).

Although national events and characteristics naturally played a large role in the origins of education systems, Ramirez and Boli (1987) focus only on “transnational similarities in the institutional character of state educational systems” (p. 2). Each of the regions discussed by Ramirez and Boli showed the same pattern in response to external challenges. These patterns follow 4 steps (p. 9). First there was a “declaration of a national interest in mass education”. In Germany this can be seen by an address by the German philosopher Johann Gottlieb Fichte in 1807 which stated that “universal state-directed, compulsory education would teach all Germans to be good Germans and would prepare them to play whatever role [...] fell to them in helping the state reassert Prussian power” (p. 5). In Denmark this can be seen by the claims between 1807 and 1814 by N.F.S. Grundvig, a clergyman, that education “[is] a means for Denmark to regain its spiritual and national strength” (p. 6). Also in Italy: “Education was seen as a means of increasing Italian power and prestige” (p. 7).

The second step on the route towards mass public education was “legislation to make schooling compulsory”. Although most states had some legislation of education in place in the 18th century, often it was neither mandatory, nor universal, nor free. Secondly, in some cases legislation was set up but the necessary funds were not delivered and thus the legislation was not put into practice, as was for instance the case in Prussia in 1717 and in Denmark in 1739. Prussia was one of the first regions which provided state mass education and also established a tax instrument to finance it, but only after 1817.

With the legislation in place, the “creation of a state educational ministry or department” and the “establishment of state authority over existing and new schools” were possible. Prussia established a Bureau of Education in 1806 and after 1817 the state provided certification of teachers, while in England only in 1944 a national

ministry of education was formed to act as the central education authority. Churches often had influence on education until well in the 19th century. In Austria in 1855 full control of education was given back from the state to the church and even in Prussia, the government announced in 1844 that “religious instruction is more important than pedagogical theory”.

Before the state took control of the (existing and new) schools, schools with different curricula and structures could exist side by side. This provides an explanation to the current existence of differences in differentiation and the different levels of standardization between countries. When different school types exist next to each other and they have to merge into one education system, it might be chosen to keep the different school systems as complements, i.e. as different tracks. And it might also be that when these school merge into one system, the schools are allowed to keep some aspects of their former system (for instance their pedagogical views and related textbooks and courses) and therefore standardization might be low. How different countries came to different solutions for merging multiple school types into one is not the purpose of this chapter. However, in the sociological literature a number of theories exist to explain this. For instance, Archer (1979) or Boli *et al.* (1985).

2.3 Empirical support for the influence of Napoleon on education systems

This section builds upon the previous section in the sense that it will supplement the theory by Ramirez and Boli (1987) by empirical evidence that the political pressure from Napoleon influenced specific characteristics of European education systems and that this influence is still seen today.

2.3.1 Data

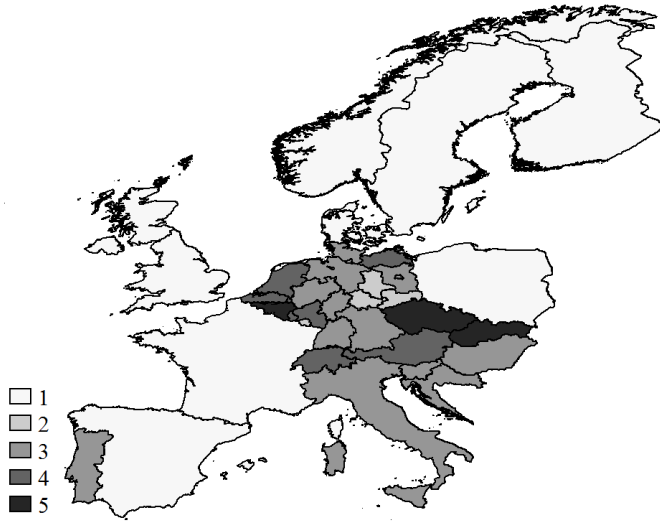
The data used in this section consists of data on input standardization from the Program for International Student Assessment (PISA) 2006 organized by the Organization for Economic Co-operation and Development (OECD), data on tracking and output standardization and data on the political pressure from Napoleon. The last two are obtained from various sources.

Figure 1 shows how education systems differ on two aspects: tracking and standardization. Tracking, or differentiation, can be shown by the number of tracks available to pupils and the age of first selection in the system. Data on the number of tracks is from the OECD (2007, Table 5.2) and adjusted for Flemish and French Belgium (Eurydice, 2013) and for the Germany states (Woessmann, 2007). Data on the age of first selection is from the OECD (2007, Table 5.2) and adjusted for Flemish and French Belgium (Eurydice, 2013) and for the Germany states

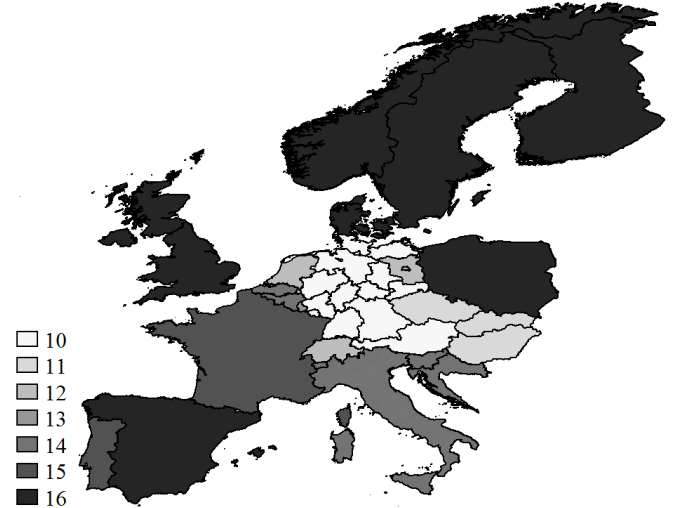
(Kultusminister Konferenz, 2013). Table 1 and Figure 1 show the levels of tracking for 29 countries. Eight countries have one track available to students, while five tracks are available in 3 countries. The earliest selection is at age 10, while the latest is at age 16.

Figure 1: Education systems in Europe

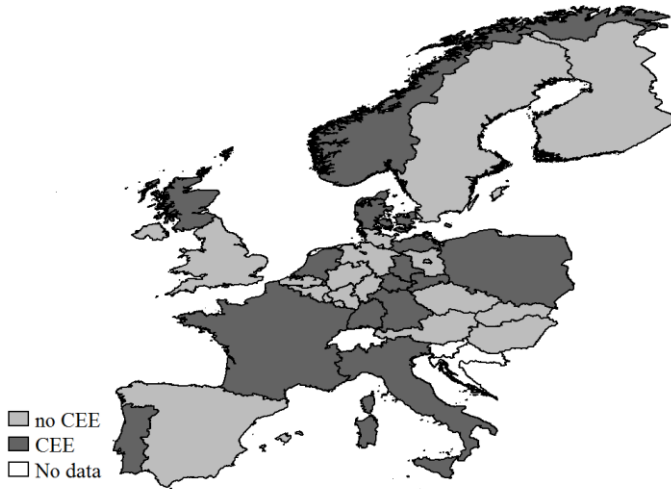
a. Number of tracks



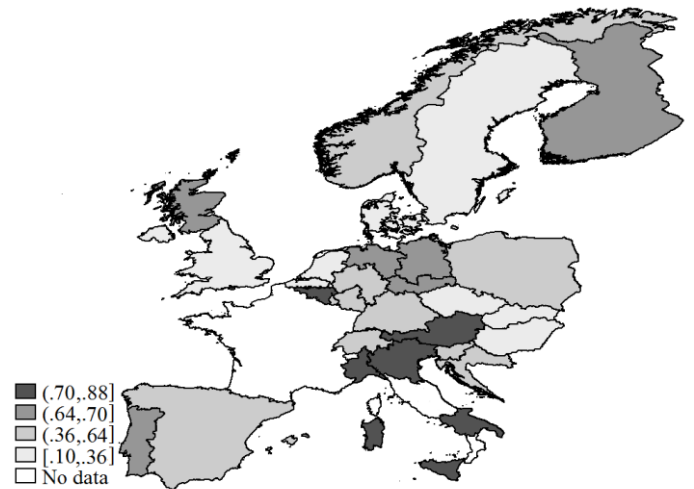
b. Age of first selection



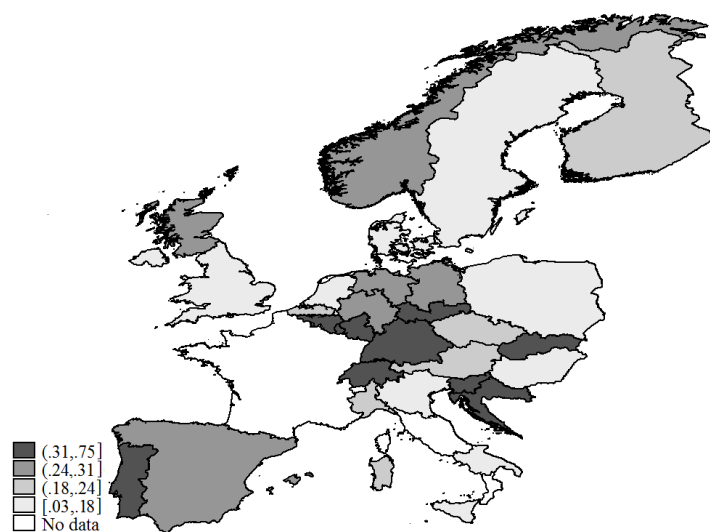
c. Central exit exams



d. Administrative administration



e. Content administration



Notes: The borders are drawn at the lowest possible level at which the data are available. In the case of the number of tracks, the age of first selection and the central exit exams this means at the country level, or at the region level for French Belgium and Flemish Belgium and for England and Scotland, or at the state level for Germany. The borders for standardization deviate from this, in the sense that the lowest possible level for Germany is the state cluster level as explained in Appendix B.

Table 1: Descriptive statistics education systems in Europe⁹

	Obs	Mean	Std. Dev.	Min	Max
Number of tracks	29	2.83	1.28	1.00	5.00
Age of first selection	29	13.07	2.28	10.00	16.00
Central Exit Exams	26	0.42	0.48	0.00	1.00
Administrative stand.	29	0.52	0.23	0.10	0.88
Content related stand.	29	0.28	0.16	0.03	0.75
Distance in km	29	940.69	505.55	264.00	2304.00
Empire	29	0.45	0.51	0.00	1.00
Pays réunis	29	0.38	0.49	0.00	1.00
Pays conquis	29	0.24	0.44	0.00	1.00
Pays allies	29	0.21	0.41	0.00	1.00

Notes: The sources of the variables are described in the text.

Standardization here is divided into two parts: standardization on output, shown by the existence of a central exit examination, and standardization on inputs, which is divided into standardization on administrative aspects and standardization on content related issues. Data on central exit examinations is from OECD (2011, Table D5.1a) and

9. The countries/regions used in this Table are Austria, Belgium (split into Flemish Belgium and French Belgium), Switzerland, Czech Republic, Germany (split into Bavaria, Hanover, Prussia, Rhineland, Saxony, Westphalia), Denmark, Spain, Finland, United Kingdom (split into England-Wales and Scotland), Croatia, Hungary, Italy (split into the north west, the Republic of Italy, and the Kingdom of Naples), Luxembourg, Netherlands, Norway, Poland, Portugal, Slovak Republic, Slovenia, and Sweden. See Appendix B for a more extensive description of the regions.

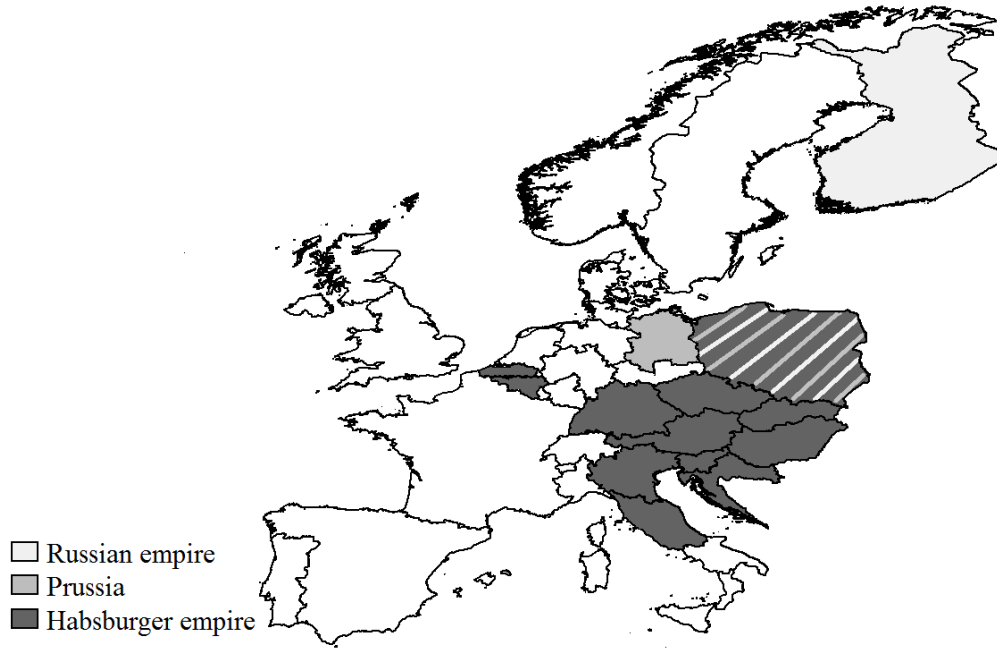
adjusted for the German states using Juerges *et al.* (2005). Data on input standardization is from the Program for International Student Assessment (PISA) 2006, which included answers given by school principals on who has authority for the textbooks, course content, teacher appointments, teacher employment conditions, and budget. These last five variables are transformed into two country averages, one on the administrative aspects (who is responsible for teacher hiring, salary increases and formulating the budget) and standardization on content related issues (who is responsible for textbook choice and course content). Figure 1 shows that countries in the periphery of Europe have less differentiation (both measured by the age of first selection and by the number of tracks) than countries in the inner region, but for standardization the picture is less clear.

The geographical spread in Figure 1 shows a (weak) relation between France and education systems. This geographical spread is best captured by modeling the political threat from Napoleon by distance of one's own capital to Paris. The air distance between Paris and the capital around 1800 in Table 1 is given in kilometers, 500 kilometers is added for a sea crossing. Belgium's Brussels is closest to Paris with a distance of 264 km, while Finland's capital of Turku is furthest away with 2304 km. Distance to Paris is a measure of the ex-ante political threat by Napoleon, since it uses the theoretical threat.

Whether or not countries belonged to an empire in the 1800s controls for any protection or incitement this might have ensured. Belonging to an empire would have either protected the country from outside threat due to the larger protective force of an empire compared to a (smaller) country and thus lowered the threat from Napoleon, or it would have increased the threat when there were struggles between Napoleon and the empire in question. The empires in Europe in the 1800s were the Russian Empire, the Habsburg Empire, and Prussia, as shown in Figure 2.¹⁰ Together with the distance to Paris, this provides a model to measure the ex-ante political threat of Napoleon.

10. Prussia was a Kingdom and not an empire but it will be referred to one in this chapter for simplicity. A fourth empire in Europe was the Ottoman Empire. However, only one country (Croatia) that once belonged to the Ottoman Empire is included in this chapter.

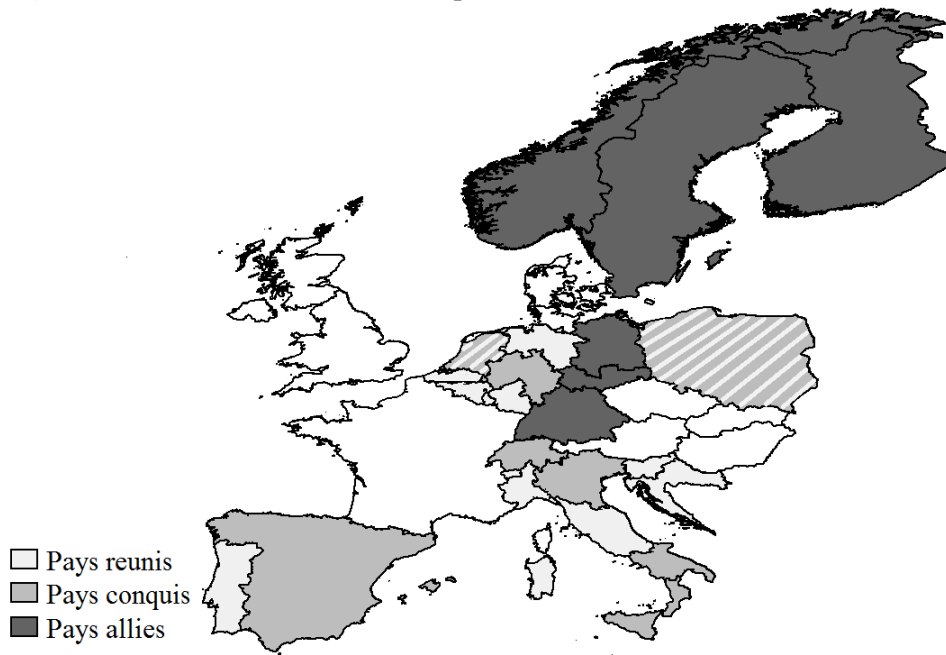
Figure 2: Empires in Europe around 1800s



Notes: These maps are based on the Centennia Historical Atlas. Appendix B provides a justification of the borders used in this chapter. Regions of Poland belonged to either Prussia (the north west), the Habsburger Empire (the south), and the Russian Empire (the east).

A measure of the ex-post threat used in this chapter is the categorization by Grab (2003), as shown before. This is ex-post since it relates to which countries were actually annexed and not which countries were under threat. The categorization from Grab (2003) divides the European countries into “countries annexed to France and directly ruled by Napoleon (*pays réunis*)”, conquered countries which were “satellite states that were entrusted to French rulers (*pays conquis*)”, and “allied countries whose territory was expanded by Napoleon in some cases, and continued to be governed by their native rulers (*pays allies*)” and is shown in Figure 3. Descriptive statistics are shown in Table 1. Appendix C provides more information on the categorization.

Figure 3: The reach of the French Empire (based on Grab, 2003)



Notes: *Pays réunis* are “countries annexed to France and directly ruled by Napoleon”. *Pays conquis* are conquered countries, or “satellite states that were entrusted to French rulers”. *Pays allies* are “allied countries whose territory was expanded by Napoleon in some cases, and continued to be governed by their native rulers”. The Netherlands and Poland were both *pays réunis* and *pays conquis*. Appendix B provides a justification of the borders used in this chapter. Appendix C provides the full list of the categorization of regions and countries by Grab (2003).

2.3.2 Results

Table 2 shows different models relating the distance to Paris, whether or not countries belonged to an empire, and the categorization by Grab (2003) to the number of tracks available in a country. The first model shows only the relation between distance and the number of tracks available and shows a strong negative correlation, meaning that the further away one’s own capital is from Paris the fewer tracks its education systems will have. Model 2 uses both the distance and the empire dummy, and is therefore the preferred model since it relates most closely to the ex-ante political threat by Napoleon. This model has a lot of power, explaining forty-five percent of the variation in differentiation. Countries belonging to an empire in 1800 have more tracks available to students than countries not belonging to an empire in 1800. The coefficient for the distance to Paris is similar to the previous model.

Model 3 and 4 use the historical categorization of Grab, both combined with distance and alone. The fourth model with only the three Grab dummies seems less able to explain the variation found in the number of tracks in Europe. None of the dummies are statistically significantly from zero, but an F test shows that jointly they are. Model 5 includes all independent variables and also these are jointly significant. A possible reason that the variables in model 3, 4 and 5 are jointly significance but that the Grab

dummies are not significant is the high correlation between the three: Only seven countries are not *pays réunis*, nor *pays conquis*, nor *pays allies*.¹¹ All other countries are *pays réunis*, or *pays conquis*, or *pays allies*, except for Poland and the Netherlands which are both *pays réunis* and *pays conquis* since their status changed over time. The seven countries which are neither *pays réunis*, nor *pays conquis*, nor *pays allies* are therefore the reference category for the categorization by Grab (2003).

The five models do remarkably well: They explain fifteen to forty-five percent of the variation and the variables in all models are jointly significant. The estimates say that countries that have their capital further away from Paris have fewer tracks available to pupils nowadays. Countries that were part of an empire have more tracks. Annexed countries ruled by Napoleon (*pays réunis*) have more tracks, while other annexed or allied countries have fewer tracks. However, the estimates for the Grab dummies might be difficult to interpret due to the correlation between the three.

Table 2: The effect of political pressure from Napoleon on the number of tracks

	(1)	(2)	(3)	(4)	(5)
Distance	-0.002*** (0.000)	-0.002*** (0.000)	-0.001** (0.001)		-0.001** (0.000)
Empire		1.031** (0.375)			1.094*** (0.387)
Pays réunis			0.486 (0.532)	0.992* (0.540)	0.726 (0.476)
Pays conquis			-0.397 (0.537)	-0.261 (0.586)	-0.097 (0.484)
Pays allies			-0.151 (0.649)	-0.692 (0.672)	-0.051 (0.573)
Constant	4.344*** (0.444)	3.932*** (0.426)	4.053*** (0.677)	2.692*** (0.436)	3.326*** (0.649)
# of countries	29	29	29	29	29
Adjusted R ²	0.32	0.45	0.29	0.15	0.45
p F test	0.001	0.000	0.014	0.073	0.002

Notes: The table presents coefficients (standard errors in parenthesis) on the relation between political pressure from Napoleon and the number of tracks in a country. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. See table 1 for notes on the data sources. Model 2 is the preferred model.

Table 3 shows the same models but now for the age of first selection. Also here the first three models are capable of explaining part of the variation in the age of selection, but the fourth model is not, as seen by the F test and the adjusted R². The coefficients have the reversed signs as compared to Table 2, as is to be expected since the number

11. These are Austria, Czech Republic, Denmark, England, Hungary, Scotland, and Slovak Republic.

of tracks available in a country and the age of first selection are highly negatively correlated (-0.66 in this sample).

Table 3: The effect of political pressure from Napoleon on the age of first selection

	(1)	(2)	(3)	(4)	(5)
Distance	0.002** (0.001)	0.002** (0.001)	0.003*** (0.001)		0.003*** (0.001)
Empire		-0.919 (0.807)			-0.760 (0.828)
Pays réunis			0.788 (0.998)	-0.451 (1.084)	0.621 (1.018)
Pays conquis			0.574 (1.007)	0.241 (1.177)	0.365 (1.035)
Pays allies			-1.643 (1.218)	-0.316 (1.348)	-1.712 (1.224)
Constant	11.140*** (0.863)	11.510*** (0.917)	9.982*** (1.270)	13.320*** (0.875)	10.490*** (1.387)
# of countries	29	29	29	29	29
Adjusted R ²	0.17	0.18	0.20	-0.11	0.19
p F test	0.015	0.028	0.053	0.96	0.074

Notes: The table presents coefficients (standard errors in parenthesis) on the relation between political pressure from Napoleon and the age of first selection in a country. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. See Table 1 for notes on the data sources. Model 2 is the preferred model.

Table 4 shows the models using central examinations as dependent variable. What is immediately clear is that although the pressure of Napoleon has a relation to the two measures of differentiation, it has no such relation with central exit examinations. No model has jointly significant variables. Perhaps the existence of central exit examination is too recent to be affected by the political pressure from Napoleon, or the use of central exit examinations changed too much over the years.

Table 5 shows how that the pressure from Napoleon is also not related to the level of standardization on administrative issues. None of the coefficients is significant, also not jointly.

Table 4: The effect of political pressure from Napoleon on the existence of central examinations

	(1)	(2)	(3)	(4)	(5)
Distance	0.000 (0.000)	0.000 (0.000)	-0.000 (0.000)		-0.000 (0.000)
Empire		-0.170 (0.207)			-0.180 (0.215)
Pays réunis			0.031 (0.269)	0.077 (0.240)	-0.039 (0.284)
Pays conquis			0.401 (0.262)	0.407 (0.257)	0.364 (0.268)
Pays allies			0.463 (0.297)	0.419 (0.272)	0.463 (0.299)
Constant	0.414* (0.214)	0.484** (0.232)	0.350 (0.310)	0.247 (0.180)	0.489 (0.354)
# of countries	26	26	26	26	26
Adjusted R ²	-0.04	-0.05	0.00	0.03	-0.02
p F test	0.802	0.695	0.445	0.304	0.498

Notes: The table presents coefficients (standard errors in parenthesis) on the relation between political pressure from Napoleon and the existence of central exit exams in a country. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. See Table 1 for notes on the data sources. Model 2 is the preferred model.

Table 5: The effect of political pressure from Napoleon on the administrative standardization

	(1)	(2)	(3)	(4)	(5)
Distance	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)		0.000 (0.000)
Empire		0.005 (0.089)			0.021 (0.096)
Pays réunis			0.040 (0.114)	0.053 (0.104)	0.045 (0.119)
Pays conquis			0.081 (0.115)	0.084 (0.113)	0.087 (0.121)
Pays allies			0.085 (0.139)	0.072 (0.129)	0.087 (0.143)
Constant	0.547*** (0.0930)	0.545*** (0.101)	0.500*** (0.145)	0.466*** (0.0837)	0.486*** (0.162)
# of countries	29	29	29	29	29
Adjusted R ²	-0.03	-0.07	-0.13	-0.09	-0.18
p F test	0.758	0.953	0.941	0.871	0.976

Notes: The table presents coefficients (standard errors in parenthesis) on the relation between political pressure from Napoleon and administrative standardization in a country. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. See Table 1 for notes on the data sources. Model 2 is the preferred model.

The relation between the pressure from Napoleon and the second form of input standardization, content related standardization, is very small, but model 3 and 4 of Table 6 do show jointly significant correlations. Here especially, annexed countries rules by Napoleon (*pays réunis*) are seen to have more content related standardization.

Table 6: The effect of political pressure from Napoleon on the content related standardization

	(1)	(2)	(3)	(4)	(5)
Distance	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)		0.000 (0.000)
Empire		0.011 (0.060)			0.016 (0.058)
Pays réunis			0.140* (0.069)	0.164** (0.0636)	0.144* (0.072)
Pays conquis			-0.052 (0.070)	-0.0460 (0.0690)	-0.048 (0.073)
Pays allies			0.095 (0.084)	0.0697 (0.0791)	0.097 (0.086)
Constant	0.357*** (0.063)	0.353*** (0.069)	0.276*** (0.088)	0.213*** (0.051)	0.266** (0.098)
# of countries	29	29	29	29	29
Adjusted R ²	0.04	0.00	0.16	0.17	0.13
p F test	0.166	0.384	0.0865	0.0581	0.153
<i>Notes:</i> The table presents coefficients (standard errors in parenthesis) on the relation between political pressure from Napoleon and content administration in a country. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. See Table 1 for notes on the data sources. Model 2 is the preferred model.					

All in all, pressure from Napoleon does seem to have a relation with some aspects of education systems in Europe; most notably with differentiation, as seen by the number of tracks and age of selection, and content related standardization. For the number of tracks in a country, the proposed variables are able to explain thirty to forty-five percent of the cross country variation. For the age of selection the pressure from Napoleon is able to explain around twenty percent. These results therefore provide support for the thesis of Ramirez and Boli (1987).

2.4 Can the Napoleonic Wars explain student performance?

Section 2.3 shows that the political pressure from Napoleon has a relation with some current characteristics of education systems in Europe. Therefore, it is possible to investigate the effect of education systems characteristics on student performance using pressure from Napoleon. In general, studying the effects of education systems characteristics on student performance is hindered by the endogeneity of education

systems, in which factors that shaped the current characteristics also (direct or indirect) influence current student performance. Using Instrumental Variable (IV) analysis this endogeneity can be resolved, and in this chapter pressure from Napoleon is used as an instrument for education systems characteristics. Section 2.5 pays closer attention to the exclusion restriction.

This chapter looks at two aspects of education systems: tracking and standardization. From Section 2.3 it is to be expected that the IV will work best for tracking since there the model could explain the cross country variation in system characteristics. For standardization, especially central exams, less is to be expected. Therefore the following section focuses on tracking. Standardization is not pursued any further, but Appendix D does present the results for the interested reader.

2.4.1 Data

Data on three topics are used. Individual level data on student performance and background characteristics are taken from PISA 2006. Country level data on tracking and on the political threat from Napoleon are as discussed in Section 2.3.

Student performance is obtained from PISA 2006 which was conducted in 57 countries, of which 21 are in Europe and located on the battle ground of the Napoleonic Wars. France is excluded from both samples since the argument of political threat from France does not apply to France itself. Four countries are split into regions: Belgium (into Flemish and French Belgium), the United Kingdom (into England/Wales and Scotland), Italy (into the North West, the Republic of Italy and the Kingdom of Naples¹²), and Germany (into Bavaria, Hanover, Prussia, Rhineland, Saxony, Westphalia).¹³ This chapter therefore uses 29 regions, which will be called countries for simplicity, 6578 schools, and 157,200 students.¹⁴ Missing values are replaced by either school or country averages. Analyses are done with student weights and clustering of errors on country level. Using country level variables on differentiation and clustering on the country level implies that the sample effectively only contains 29 observations.

PISA 2006 contains a reading, mathematics, and science test score. This chapter uses the mathematics test score. However, the results are very similar when I use the science test score, which was the main focus of PISA 2006, or the reading test score, although the reading results are not significant. Next to the test scores, PISA also provides information on the student and the school. In this chapter the following

12. Some Italian regions are excluded since in PISA they are grouped together in a non-geographical way.

13. For more information on these regions, see Appendix B.

14. Excluded are students with no information on age, gender or social economic background, migrant students, and students in schools with less than 5 students in PISA.

control variables are used: the gender of the student, the number of books in the household (as a proxy for parental background), the grade of the student, whether the student is in vocational education, the school average of the number of books of the students, the school size, the location of the school, whether the school has ability grouping, and GDP per capita 2005.

Table 8: Correlations between age of first selection and test scores

Database: Subject:	PISA 2000 reading	PISA 2006 reading	PISA 2006 math	PISA 2006 science
<i>Uses the OECD (2010) classification based upon age of first selection</i>				
OECD (2010) sample (27 cnts)	0.59	0.34	0.23	0.26
OECD (2010) minus MEX	0.54	0.23	0.07	0.10
Including NLD, HRV, SVK, SVN	0.43 ^a	0.25	0.04	0.09
Sample used in this chapter (20 cnts)	0.37 ^b	0.21	-0.04	0.02
<i>Uses age of first selection from OECD (2007)</i>				
Sample used in this chapter (20 cnts)	0.19 ^b	0.19	-0.16	-0.11
Sample used in this chapter (29 regions)	-	-0.14	-0.23	-0.35
<i>Notes:</i> ^a In PISA 2000 only the Netherlands can be added. ^b This sample misses Croatia, Slovak Republic and Slovenia.				

The same two aspects of differentiation in education systems are used here as in Section 2.3: the number of tracks available to 15-year-old pupils and the age of first selection. The correlation between the mathematics test score and the age of first selection is -0.23, while the correlation with the number of tracks at the country level is 0.06. These correlations deviate widely from the consensus in the public debate as illustrated by the following correlations reported by the OECD (2010): the country level correlation between reading scores and a classification based on the age of first selection in PISA 2000 is 0.59 for the sample of OECD countries (OECD, 2010, p. 58). This difference in correlations between this chapter and OECD (2010) is unraveled in Table 8. First, the OECD correlation is based on PISA 2000, while in this chapter PISA 2006 is used (making the correlation drop from 0.59 to 0.34). Second, the OECD used the reading test score, while I use the mathematics test score (0.34 vs 0.23). Third, I add five countries (the Netherlands, Croatia, Norway, Slovak Republic, and Slovenia) and drop all non-European countries (Australia, Canada, Japan, Korea, Mexico, New Zealand, and the United States) and some European countries which were not in the sphere of influence of France (France itself and Greece, Iceland and Ireland) (0.26 vs -0.04). However, the country that most drives the high positive correlation between age of first selection and the reading score in PISA 2006 is Mexico, as can be seen in the second row where Mexico is excluded from the OECD (2010) sample. Fourth, I split Belgium, Germany, Italy, and the United Kingdom into regions and in the analyses I use the age of first selection variable as provided by the

OECD (2007), while the OECD (2010) uses a classification based upon the age of first selection.¹⁵ This results in a correlation between the age of first selection and student performance of -0.23.

2.4.2 Estimation Method

The models estimated in this paper are straightforward IV models shown by equations (1) and (2). In this first stage, I use the political pressure from Napoleon ($NAPOLEON_c$) to capture the variation in tracking ($TRACKING_c$) due to this pressure for each country c . In the second stage, I then use the fitted values to estimate the effect of tracking on student performance ($TEST SCORE_{i,c}$). I control for individual, school and country characteristics captured in $X_{i,c}$. These background characteristics are the gender of the student, the number of books in the household, the grade of the student compared to the modal grade in the country, whether the student is in vocational education, the school average of the number of books of the students, the school size, the location of the school, whether the school has ability grouping, and GDP per capita in 2005.¹⁶ The variable of interest is β which displays the effect of tracking on student performance.

$$TRACKING_c = \gamma + \delta NAPOLEON_c + X_{i,c}\theta + \epsilon_i \quad (1)$$

$$TEST SCORE_{i,c} = \alpha + \beta \widehat{TRACKING}_c + X_{i,c}\vartheta + \epsilon_i \quad (2)$$

For a valid IV two conditions must hold: First, the instrument should have an effect on the endogenous variable. Second, the instrument should not have an effect on anything else that might influence the outcome, in this case, student performance. The second condition (the exclusion restriction) cannot be tested and must be taken on good faith. Section 2.5 elaborates more on this. The first condition can be verified by a strong first stage. Sections 2.3.2 and 2.4.3 show that this first condition is indeed fulfilled.

In these analyses the level of tracking is captured by two variables: the number of tracks available to 15 year old students and the age of first selection. In general, using the same instrument for multiple endogenous variables is problematic since this will violate the exclusion restriction. However, both the number of tracks and the age of first selection are proxies of the underlying level of tracking in an education system justifying the use of a single theory behind the instruments for both. The consequence is that the results have to be interpreted either separately for both endogenous variables

15. Countries with no selection before or at 15 are group 1, countries with selection at age 14 or 15 are group 2 and countries with selection before 14 are group 3. Between the age of first selection used in OECD (2010) and the variable used in this chapter are two differences: Switzerland selects at age 14 (instead of age 12) and Poland selects at age 16 instead of 15.

16. Missing values are replaced by school or country averages and imputation dummies and interactions are included to ensure the imputation method does not affect the estimates.

by assuming no effect of the other aspect of tracking or jointly. A third option would be to use more instruments for both aspects simultaneously. In the discussion of the results, this is further addressed.

2.4.3 Results

The negative correlation between the age of first selection and student performance shown in Table 8 is most probably clouded by endogeneity. The current section tries to alleviate this endogeneity by instrumenting differentiation by the political pressure from Napoleon as discussed in Section 2.3. Table 9 shows the OLS and IV results of differentiation on student performance, using the same model specifications as before.¹⁷ The upper panel of Table 9 shows the first stage, while the lower panel shows the second stage. In Section 2.3, it was shown that it was most promising to use the political pressure from Napoleon as an instrument for differentiation and also here the first stages are quite strong and very similar to in Section 2.3. Using the first dimension of differentiation, the number of tracks, reveals a significant effect of the number of tracks on student performance as seen by mathematics test scores. The OLS model reveals that one more track available to students leads to an 8.8 point (**) increase in student math test scores. Model 2 and 4 show that this estimate is downwards biased since the IV models reveal a coefficient of 16.3-16.7*** of the number of tracks. The point estimates of the other models are smaller (11.8-13.0) and the standard errors are bigger, resulting in insignificant coefficients. The downward bias in the OLS estimates could be explained by the endogeneity in the data for which I correct using IV. It must be that some countries which had bad student performance moved towards more tracking after the influence of Napoleon stopped or countries that had good performance shifted away from tracking, which explains the lower difference in the average student performance found by OLS.

The preferred model 2 says that when a system has one more track, student performance as measured by a math test score will be one-sixth of a standard deviation higher (mean: 507; sd: 93.1). This seems small, but it is 1.5 times the amount of moving an individual student from a household with 0 to 10 books to a household with 11 to 25 books (10.91***) which is quite substantial.¹⁸ The full second stage model including all the control variables is depicted in Appendix E.

The second dimension of tracking, the age of first selection, has no significant effects on student performance, as shown by the right panel of Table 3, although all estimates are consistently negative. However, note that the two measures of tracking are correlated and thus the results for both measures must either be interpreted separately

17. Model 4 is excluded since in Section 2.2 it was the weakest model.

18. In 2006 the variable *number of books in the household* has six categories: 0-10 books, 11-25 books, 26-100, 101-200, 201-500, more than 500.

by assuming the insignificance of the other or jointly. A joint interpretation would be that the level of tracking has a positive effect on student performance, but that it is unclear whether this comes from a positive effect of the number of tracks or a negative effect of the age of first selection. IV models using more than two instruments for both endogenous variables simultaneously resulted into similar coefficients (not shown, but available on request). However, the standard errors were much larger leading to insignificant results. This indicates that with only 29 countries this simultaneous estimation is too demanding for the data.

It is interesting to relate these results to those found by others. The literature on the effect of tracking is not as extensive, due to the problem of endogeneity, and the findings are not very consistent. Hanushek and Woessmann (2006) provide a difference-in-difference analyses into the effects of selection into tracks at an early age and find a negative effect of early tracking (although not robust to sample changes), which is confirmed by Van Elk *et al.* (2011). Ariga and Brunello (2007) look at the tracking length and find a positive effect of the time spend in tracks. Fuchs and Woessmann (2007), Pekkarinen (2008), and Kerr *et al.* (2013) find insignificant effects. The OECD correlations reported earlier are therefore neither widely confirmed by the literature, nor by this study.

Table 9: The effect of differentiation on mathematics test scores

	Number of tracks					Age of first selection				
	(0)	(1)	(2)	(3)	(4)	(0)	(1)	(2)	(3)	(4)
1st stage										
Distance		-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)		0.002*** (0.000)	0.002*** (0.000)	0.003*** (0.001)	0.003*** (0.001)
Empire			0.836*** (0.286)		0.950*** (0.280)			-0.609 (0.740)		0.0330 (0.712)
Pays réunis				0.455 (0.411)	0.612 (0.382)				0.671 (0.813)	0.676 (0.812)
Pays conquis				-0.312 (0.432)	-0.071 (0.445)				0.536 (0.795)	0.545 (0.814)
Pays allies				0.0241 (0.456)	-0.225 (0.343)				-1.967** (0.861)	-1.976** (0.870)
2nd stage										
Number of tracks	8.831** (3.403)	11.810 (9.434)	16.330** (7.630)	12.960 (9.180)	16.730** (7.435)					
Age of first selection						-1.991 (1.875)	-7.754 (6.901)	-10.01 (6.858)	-6.066 (4.522)	-5.974 (4.506)
# of students	157,200	157,200	157,200	157,200	157,200	157,200	157,200	157,200	157,200	157,200
# of countries	29	29	29	29	29	29	29	29	29	29
R ² 2 nd stage	0.33	0.34	0.33	0.33	0.63	0.33	0.31	0.30	0.32	0.32
Partial R ² 1 st stage		0.326	0.423	0.376	0.483		0.210	0.224	0.338	0.338
F test 1 st stage		64.89	27.61	19.34	15.34		24.26	10.64	12.53	10.41
<i>Notes:</i> The table presents coefficients (standard errors in parenthesis) on the effect of tracking on student performance, instrumenting in models (2) to (5) tracking with the political pressure from Napoleon. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Model 2 is the preferred IV model. The models include imputation dummies and imputation variable interaction terms, and clustering is done on the country level. The control variables are as explained in the text. Differences in the first stage as compared to Section 2 are due to the inclusion of the student level.										

2.5 Threats to the identification

2.5.1 Exclusion restriction

It seems quite reasonable that the pressure from Napoleon influenced more than just tracking in education systems (e.g. Acemoglu *et al.*, 2011) and also other historical events, likely to be correlated to Napoleon, influenced parts of the human capital development of nations (e.g. Becker and Woessmann, 2009; Dittmar, 2011; Cantoni and Yuchtman, 2014). These correlated events could potentially bias the IV results. The first stage analyses in Section 2.4.3 are therefore also performed by using the GINI coefficient, the number of police officer per 100,000 inhabitants, the expenditure on schooling per capita and the number of women seats in parliament.¹⁹ Table 10 shows the effect of the political pressure from Napoleon on these outcomes, using model 2 of Table 9 (the preferred model). What can be seen is that in general political pressure from Napoleon does not really affect these current day societal factors: The individual instruments have in two cases a significant effect, but in most instances they do not. The F test on joint significance shows the model on educational spending is significant (at the 10 percent level), but for the other three models this is not the case.

Table 10: The effect of Napoleon on other societal characteristics

	GINI	Police	Exp. Edu.	Women seats
Distance	-0.003* (0.001)	-0.052 (0.034)	0.001** (0.000)	0.003 (0.004)
Empire	-0.042 (1.349)	27.040 (33.470)	-0.268 (0.378)	-0.813 (3.711)
Constant	33.560*** (1.460)	353.700*** (38.030)	4.466*** (0.429)	24.900*** (4.217)
# of countries	26	29	29	29
Adjusted R ²	0.0752	0.03	0.11	-0.05
p F test	0.156	0.254	0.087	0.729

Notes: The table presents coefficients (standard errors in parenthesis) on the relation between the political pressure from Napoleon and societal characteristics. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Appendix F also shows the other models used as first stages in Table 9: The instruments using the GINI coefficient are found to be jointly significant in four out of five models, for the other societal characteristics the joint significance is only in two or one models out of five. Regardless, this suggests that at the very least pressure from Napoleon influenced (determinants of) the GINI coefficient in countries. And the level of inequality in a society, as measured by the GINI coefficient, might influence

19. More information on this is given in Appendix D.

educational opportunities or other aspects of education which have an effect on student performance.

2.5.2 *Changes in tracking since Napoleon*

This chapter uses data from 2006 to see whether there is an effect of differentiation on student performance, using as an instrument political pressure from Napoleon in the 1800s. It could be that in-between the 1800s and current day, education systems changed, which could affect the results. It is difficult to find data on differentiation (or schooling in general) before the twentieth century, but most change in the European education systems occurred in the last decades, following after the formation of the United Nations in 1945 and the OECD in 1961. The emphasis of states on education, both for development and economic growth, increased and globalization of education and education systems started, best seen by the growing number of national reports on the state of education since that time (e.g. UNESCO, 1968a).

To take possible changes into account, Table 11 shows analyses from OLS and (the preferred) IV models using data from before most of the change. For these analyses I use the First International Science Study (FISS) from 1970 for ten European countries and construct a differentiation index using National Reports from UNESCO(1968a; 1968b; 1971a; 1971b; 1971c; 1971d; 1971e; 1973a; 1973b; 1975). These ten countries are Flemish Belgium, French Belgium, England, Finland, West-Germany, Hungary, Italy, the Netherlands, Scotland, and Sweden. More information on the data can be found in Appendix G.

The first stage results in Table 11 are very similar to before, but the second stage coefficients are much smaller in size. The left panel of Table 11 show even a negative effect of tracking in 1970, which lowers the confidence in the causal interpretation of the results of Section 2.4.3. However, these analyses use data for only ten countries, the F statistic of the first stage is below ten, and also the data quality on both student performance and tracking is questionable. For comparison reasons, the right panel shows results for a similar differentiation index for the same ten countries using PISA 2006 data. No significant effect is found using this set up and sample, although the sign is positive. Appendix G shows the results for the other model specifications used in this chapter.

Table 11: The effect of differentiation on student performance (science) in 1970 and 2006

Model:	1970		2006	
	OLS	IV	OLS	IV
1st stage				
Distance		-0.001** (0.000)		-0.001*** (0.000)
Empire		0.632 (0.425)		1.029 (0.659)
2nd stage				
Differentiation	-0.083* (0.045)	-0.251** (0.114)	0.01 (0.062)	0.078 (0.116)
# of students	25,409	25,409	73,171	73,171
# of countries	10	10	10	10
R ² 2 nd stage	0.23	0.211	0.353	0.348
Partial R ² 1 st stage		0.279		0.547
F test 1 st stage		5.219		11.22
<i>Notes:</i> The table presents coefficients (standard errors in parenthesis) on the relation between tracking and student performance, instrumenting differentiation with the political pressure from Napoleon. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.				

2.6 Conclusion

This chapter builds on the 1987 paper of Ramirez and Boli: *The Political Construction of Mass Schooling: European Origins and Worldwide Institutionalization*. In this paper, Ramirez and Boli argue that the introduction of European education systems in the nineteenth century was induced by external challenges on the continent. One of the most pronounced challenges was the political pressure from Napoleon, who dominated the continent for two decades. Almost all European countries were at one stage involved with Napoleon and many wars were fought on numerous battlefields. To unite the people and ensure a strong and willing workforce, countries introduced mass public education systems to educate the people as citizens of the nation willing to fight for independence. In this manner, political pressure from Napoleon had an influence on European education systems.

This chapter empirically tests whether the proposed mechanism of Ramirez and Boli (1987) had an effect on levels of differentiation and standardization in European education systems. Especially for differentiation, both seen by the number of tracks and age of selection, a relation between education systems and the political pressure from Napoleon in the nineteenth century is found, as theorized by Ramirez and Boli (1987). The proposed variables capturing the political pressure from Napoleon are able to explain thirty to forty-five of the cross country variation in the number of tracks

available to students and around twenty percent for the age of selection. For standardization relating to the content, school administration or the existence of central exit exams, no or a much weaker relation is found. Still the theory proposed by Ramirez and Boli (1987) is confirmed by empirical evidence on specific education system characteristics.

As a second step, the relation between political pressure from Napoleon and education systems is used in an instrumental variables analysis to analyze student performance. In general, the effects of education systems on student performance are obscured by the endogeneity of education systems, in which factors that influenced the characteristics also (direct or indirect) influence student performance. By using exogenous variation that influenced education systems but not student performance, this endogeneity can be alleviated. This chapter uses the political pressure from Napoleon as exogenous variation that can explain the level of differentiation in education systems.

Using Napoleon as an instrument for the level of differentiation in education systems leads to a strong first stage. When the number of tracks available to students is used as a differentiation measure the instruments also lead to a strong second stage. Jointly interpreting the results on the number of tracks and the age of first selection I find a positive effect of tracking on student performance. The analyses are unable to tell whether this originates from a positive effect of the number of tracks or a negative effect of the age of first selection.

One important caveat of this analysis is that political pressure from Napoleon influenced many facets of European countries. However, this chapter provides an additional approach which supplements the existing evidence on the effects of education system characteristics on student performance. From this chapter, it can be said that on some dimensions (mainly differentiation) there is a strong and robust relation between the political pressure from Napoleon and education system characteristics. Consequently, using the political pressure from Napoleon as an instrument for differentiation leads to a strong first stage. The resulting analyses show a consistent positive relation between differentiation and student performance, while it is only significant for the number of tracks available to students.

Selection on Performance and Tracking²⁰

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

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

3.1 Introduction

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

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

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

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

To take into account that school principals who consider prior performance possibly do so to be able to accept only the best students to their school, we compare the results of different models which look at between-school between-countries variation, between-schools variation, and between-countries variation.²¹ In the first model, we control for the best available internationally comparable track level of the individual students. In the second model, we use only the between-schools variation by using country fixed effects, which alleviates possible country heterogeneity. And thirdly, by using the national percentage of schools which consider prior performance, we look at between-countries variation only, to try to isolate the bias due to selective student acceptance. That tracking has a positive effect on student performance when the method of track placement is taken into account is seen in all three the models. The coefficients are the smallest in the within country model, while they are the largest, although insignificant, in the between countries model. If the relation was purely driven by a selection bias created by school principals, this would not be expected since then most of the variation would be between schools in a given country.

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

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

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

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

3.2 Literature

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

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

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

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

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

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

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

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

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

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

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

Schuetz *et al.* (2008) find that the effect of PB on performance is stronger when countries track early and that these countries do not perform better, on average. In

other words, there is no trade-off between efficiency and equality: Tracking affects both negatively.

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

3.3 Tracking

Tracking, as the formal practice of separating students into distinct educational programs is called, is a widely used education systems characteristics in secondary school systems around the world. Between the countries that track, the implementation of tracking can differ on the number of tracks available to 15 year old students (most frequently ranging from two to five) and on the age of selection into tracks (from 10 to 16) or alternatively on the length of the school system that is tracked. In the countries that have formalized tracking, tracks are institutionalized in different school types and often located in different buildings and administrative units, while in countries without formal tracking non-institutionalized tracking can occur within schools, either by ability grouping (different classes within schools) or seating (different curricula within classes).²²

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

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

tracking can have nationwide effects. The theoretical models of non-linear peer effects support either positive or negative effects of tracking (See Sacerdote, 2011). For instance, when especially high-ability students benefit from high-ability peers, tracking has a positive effect; when especially low-ability students benefit from high-ability peers, tracking has a negative effect.

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

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

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

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

3.3.1 Tracking and schools

Schools selecting students into tracks without an awareness of their observed abilities or basing their selection on non-academic criteria can severely limit the expected

positive effects of tracking. Basing track placement on non-academic grounds will lower the envisioned class homogeneity of tracking. This problem can be mitigated by ensuring that the selection of students into tracks is always based (primarily) on ability. A proxy for ability can be prior performance, although it still is an imperfect measure of student ability. Schools that have information on prior performance when they place students into tracks, may be better able to ensure homogeneous classes than schools that do not have information on prior performance.

3.3.2 *Tracking and equality of opportunity*

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

3.4 Data

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

The first wave of PISA was presented in 2000 and, since then, every three years a representative sample of students from all participating countries is subjected to tests on reading, mathematics, and science. The test results are standardized to a mean of 500 and a standard deviation of 100 on the PISA reading test in 2000 for the OECD countries.²³ In addition to the tests, the students and school principals are surveyed. A total of 75 countries participated in PISA 2009. Since these countries are very diverse in their economic development, this chapter uses a selection of comparable western countries to limit country heterogeneity. All 31 countries in this analysis have a gross domestic product (GDP) per capita above the minimum of the OECD and available data on national tracking policies.²⁴ These limitations on the sample are imposed to ensure that country differences do not drive the results, although also country fixed effects models are used to further take country difference into account.

23. The OECD provided five plausible values estimated using item response theory for the test scores since students do not receive all questions. In this chapter only one plausible value is used.

24. Countries such as Australia, Canada, France, and the United Kingdom, are not included since these have missing country-level data or too much missing school-level data. Mexico is excluded since many of its student and school variables are outliers.

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 2009.²⁵ This amounts to 187,768 students in 7,489 schools in 31 countries.

3.4.1 Tracking and selection by schools

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

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

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

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

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

Table 1: Tracking and selection by schools

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

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

3.4.2 Control variables

All student variables are collected through student surveys. This study controls for gender, age, PB, a dummy for (pre-) vocational education as opposed to general education, and a dummy for upper secondary school as opposed to lower secondary education. This division is based on the International Standard Classification of Education (ISCED) level, which provides internationally comparable standards for comparing education levels. PB is captured by a widely used index composed by the OECD that describes the student's economic, social, and cultural status.

Table 2: Descriptive statistics

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

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

The school-level variables are collected through a survey completed by the school principals. School composition is captured by the school average and standard deviation of the PB of all the students per school and by the percentage of them who

speak a language other than the test language at home.²⁶ School inputs are captured by the student-teacher ratio, an index of possible teacher shortages, dummy variables indicating whether the school is hindered by a shortage in instruction material, and an index indicating whether the school is responsible for the curriculum and assessment. Other school characteristics indicate whether school achievement is tracked by an educational authority; whether the school is a public, private government-dependent, or private government-independent school; whether the school has to compete with other schools for students; the school location; school size; and whether the school uses ability grouping²⁷. We also control for the GDP per capita, which is for 2008 from the World Bank (2012). Table 2 provides descriptive statistics for all variables.

3.5 Empirical strategy

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

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

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

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

In these equations, $Test_{isc}$ is the individual PISA test score in reading, mathematics, or science of student i in school s in country c . $Student_{isc}$ is a matrix of student variables, $School_{sc}$ is a matrix of school variables, while $EntrReq_{sc}$ is a matrix containing the dummies on whether schools consider prior performance. $\#ofTracks_c$ is a vector containing the number of tracks available to students in each country. As compared to the main model in equation (1), model (2) adds country fixed effects, C_c , while in model (3) we use the national percentage of school whose

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

27. Excluding the control for ability grouping leaves the results unchanged. These results are not shown, but are available from the author upon request.

principals sometimes or always consider prior performance on accepting the student to the school, $Nat \% EntrReq_c$, as opposed to the school level variable. $GDPpc_c$ is a vector containing GDP per capita. We use random effect models, which are estimated using maximum likelihood, to take into account error terms for countries, schools, and individuals. Separate error terms for countries, schools, and individuals are necessary, since students are nested within schools within countries. If we would ignore the nested data structure, we would implicitly assume that all observations are independent from each other. This would lead to a downwards bias of the standard errors. Weighting is used to ensure representative samples.²⁸ The control variables are like discussed in Section 3.4.2.

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

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

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

Almost all student- and school-level variables have some missing observations. Although most variables have below 3 percent missing values, deleting all observations with missing variables would lead to a drop in observations from around 185,000 to around 130,000. To delete observations with missing values would be to implicitly assume that the missing values are missing at random, which is a

28. The sum of the weights per country are equalised over all countries such that each country has an equal weight in the estimation. All statistics and estimations are executed using weights, unless otherwise specified.

questionable assumption. Another reason for not deleting all observations is that it leads to distorted weighting. Therefore, the missing values in the sample are replaced by group averages.²⁹ To control for possible bias introduced by the method for replacing missing values, imputation dummies and imputation interactions are used in all models.³⁰

3.6 Results

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

3.6.1 *Direct relation between tracking and student performance*

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

All the control variables, which are excluded from the table, have the expected sign. PB has a strong and positive relation with student performance. Females do better in reading and males are better in mathematics and science. Older children do somewhat better in reading and science but not in mathematics. Students in vocational education perform worse than students in general education and students in upper secondary education perform better than those in lower secondary school. The school average in PB has a positive relation with student performance. The negative relation between the percentage of students with a different first language and performance only becomes significant when more than 20 percent of the students speak a different language at home (for mathematics and science) or more than 40 percent do so (for reading).

29. The student variables are replaced by the average value of the students in the same school, the school variables are replaced by the country average. Country variables are never missing.

30. The results are robust to the exclusion of the imputation variable interaction terms and to the exclusion of the imputation dummies.

Shortages in teachers and materials have a negative relation with student performance and students in private schools that do not rely on the government for financial support perform considerably worse. Finally, school size, location, and whether it uses ability grouping are also important factors. Students in larger schools or in smaller agglomerations perform better and students in schools that use ability grouping perform worse. The results of the full models are available upon request.

Table 3: The effect of tracking on performance

Dependent variable:	(1) Reading	(2) Mathematics	(3) Science
School considers prior performance			
<i>Sometimes</i>	-2.02 (1.95)	-1.37 (1.89)	-1.13 (2.06)
<i>Always</i>	6.34* (3.27)	5.51* (3.07)	7.45** (3.79)
Number of tracks (0-4)	-2.56 (2.55)	1.94 (3.67)	0.75 (3.30)
(pseudo-)LL	-43,761	-43,739	-44,094
# of students	187,768	187,768	187,768
# of countries	31	31	31
<i>Notes:</i> The table presents coefficients from random effects models (standard errors in parenthesis) on the relation between student performance and the number of tracks in a country, controlling for whether or not schools consider prior performance when selecting students. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The control variables are as described in the text. The models include imputation dummies and imputation variable interaction terms.			

3.6.2 Tracking and performance

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

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

Dependent variable	(1) Reading	(2) Mathematics	(3) Science
School considers prior performance			
<i>Sometimes</i>	-5.97** (2.54)	-4.95** (2.40)	-5.99** (2.44)
<i>Always</i>	-10.13** (4.11)	-9.54** (4.25)	-10.90** (5.06)
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)
(pseudo)LL	-43,759	-43,738	-44,092
# of students	187,768	187,768	187,768
# of countries	31	31	31

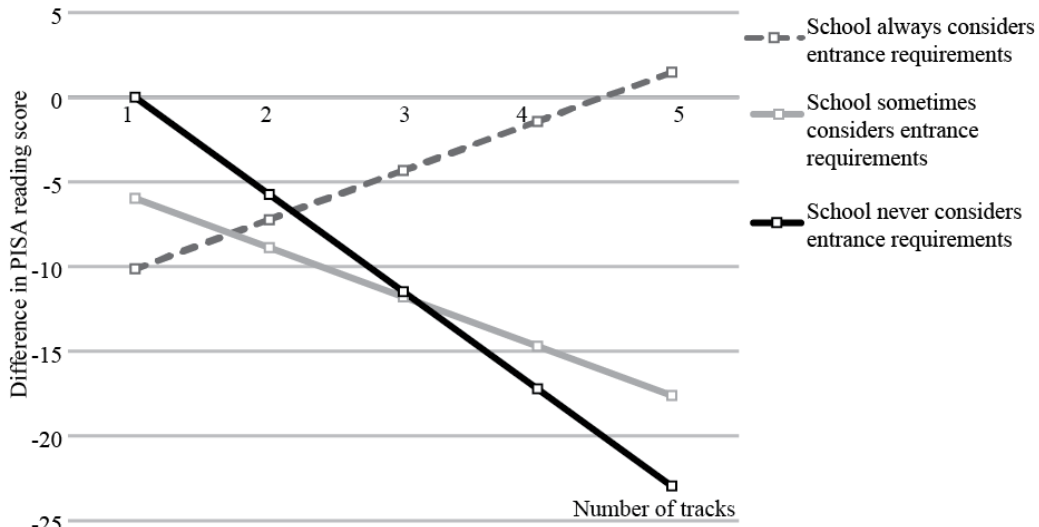
Notes: The table presents coefficients from random effects models (standard errors in parenthesis) on the relation between student performance and whether or not schools consider prior performance when selecting students and the number of tracks in a country. The superscripts *, **, and *** indicate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

To facilitate the interpretation of the interaction terms, Figure 1 shows the combined coefficients for the three models. Figure 1 show for each combination of number of tracks and whether schools consider prior performance what the relation between the two and student performance is, relative to students in a system with only one track in schools that never consider prior performance. Looking at the figures, one sees the same trend for all three subjects (reading, mathematics, and science): Schools in multiple track systems do better when they consider prior performance more often, while schools in comprehensive systems perform better when they do not consider prior performance. When only the significant differences in the graphs are considered, it becomes clear that for two or more tracks whether schools consider prior performance only changes the results when schools always consider prior performance. The coefficients for Never and Sometimes are not significantly different from each other when the number of tracks is two or more.³¹

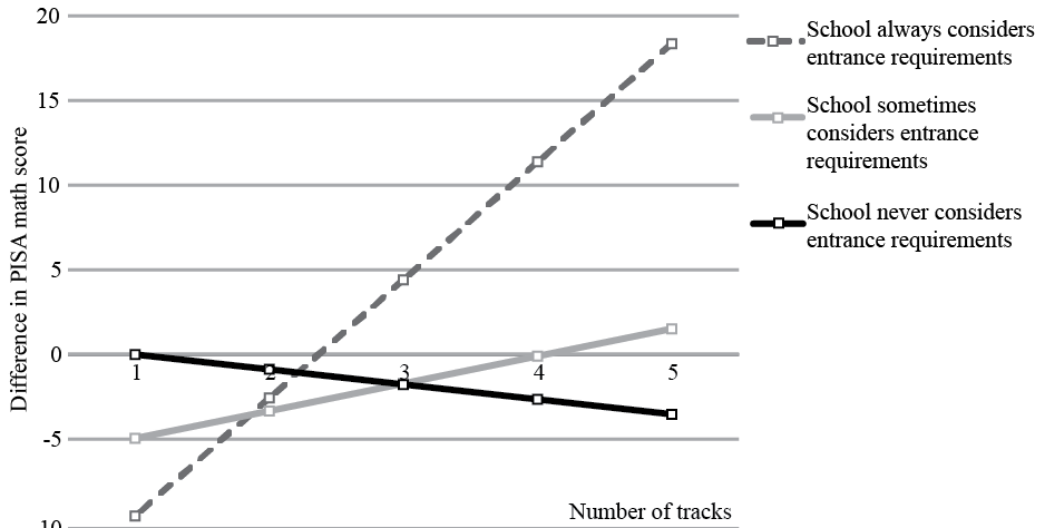
31. Except for science, where the combined coefficients for the Never and Sometimes are significantly different in a system with five tracks.

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

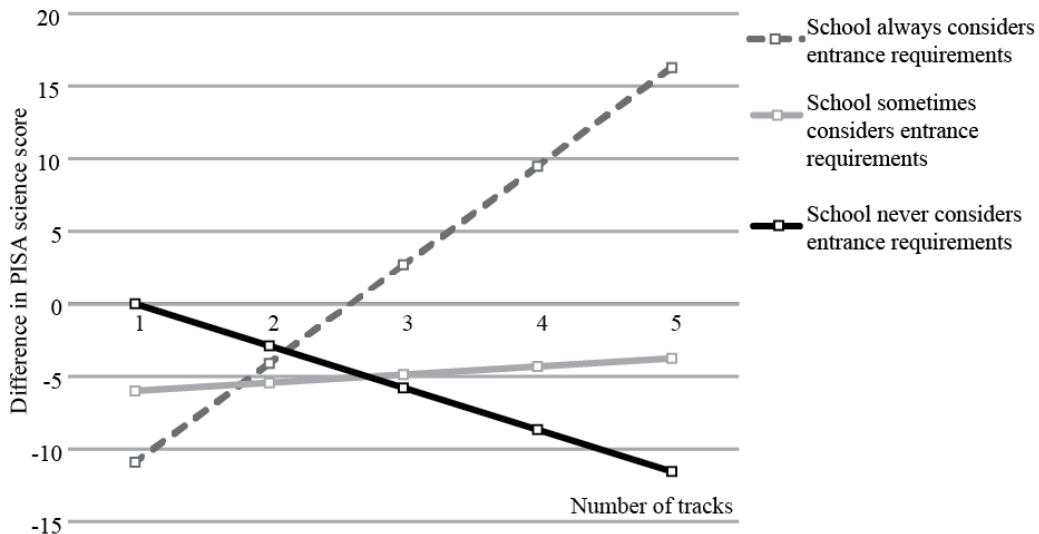
a. Reading



b. Mathematics



c. Science



Notes: The figures depict the differences in PISA test scores relative to students in a comprehensive system in schools that never consider prior performance.

The models in Table 4 use the number of tracks in a country as a continuous variable ranging from zero to four. Appendix H contains non-linear models which include the number of tracks in a country as three dummy variables: no tracks (only one track available to 15-year-olds), a few tracks (two or three tracks available), and a large number of tracks (four or five tracks). The results are qualitatively similar.

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

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

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

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

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

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

mathematics, and science is best if there are five tracks. For a country with a very high percentage of schools that always consider prior performance, students perform best in a five-track system, regardless of the subject.³³ For countries with high numbers of schools that never or sometimes consider prior performance, students perform best with two tracks, also regardless of the subject.³⁴ These results seem to indicate that, although sorting into schools could be a problem, it is unlikely that it alone drives our results.

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

	(1) Between countries, between schools	(2) Within country	(3) Between countries
School considers prior performance			
<i>Sometimes</i>	-4.95** (2.40)	-3.82* (2.03)	99.29* (54.46)
<i>Always</i>	-9.54** (4.25)	-2.03 (3.99)	-134.23*** (40.72)
Number of tracks (0-4)	-0.88 (3.75)	-	9.60 (17.72)
Sometimes*Number of tracks	2.50** (1.16)	2.28** (1.01)	-58.77 (37.00)
Always*Number of tracks	7.85*** (1.90)	5.11** (2.01)	49.89** (22.7)
Country FE	-	√	-
(pseudo)LL	-43,738	-43,695	-43,733
# of students	187,768	187,768	187,768
# of countries	31	31	31
<i>Notes:</i> The table presents coefficients from random effects models (standard errors in parenthesis) on the relation between student performance and whether or not schools consider prior performance when selecting students and the number of tracks in a country using three specifications. Column (1) shows the main model as depicted in column (2) of Table 4. Column (2) shows the same models but with country fixed effects included. Column (3) measures the school variables “school principal consider prior performance” on a national level and thus depicts the proportion of schools (between 0 and 1) in the country with school which say they always or sometimes consider prior performance. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The control variables are as described in the text. The models include imputation dummies and imputation variable interaction terms.			

3.6.3 Tracking and inequality

The results for whether using performance to select students into tracks has an effect on the relation between tracking and educational opportunities are displayed in Table 6, and described in equations (4) and (5). As expected and consistent with the

33. A country with a very high percentage of students in schools that always consider prior performance is defined as a country with 16 percent of its students in schools that sometimes consider prior performance (the mean minus one standard deviation) and 58 percent in schools that always consider prior performance for student acceptance (the mean plus one standard deviation).

34. The definition of such a country is one where 42 percent of students attend schools that sometimes consider prior performance (the mean plus one standard deviation) and 6 percent attend schools that always consider prior performance (the mean minus one standard deviation).

literature, PB has a positive and substantial relation with test scores. This effect is similar over the three PISA test subjects, reading, mathematics, and science and comprise about a quarter of a standard deviation in the test scores. If we look at the interaction between PB and tracking in columns (1), (2), and (3), it can be seen that tracking mitigates the association with PB and therefore reduces inequality of opportunity: The interaction of the number of tracks and PB is negative and highly significant. The first three columns show that in a system with five tracks the association of PB is lowered by 17.3 points in reading, 16.2 in mathematics, and 18.2 in science.

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

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

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

Finding that tracking reduces inequality is not fully consistent with the literature, which most often finds that tracking increases inequality or has no effect. To further investigate the drivers of this positive effect of tracking, we show in columns (4), (5), and (6) in Table 6 the results of the same models, but now including also interactions

between PB, the number of tracks, and whether schools consider prior performance on deciding to accept the student to the school. This allows us to check if for schools that consider prior performance, and are assumed to use this information for track placement, the effect of PB is lower. The last three columns show that the interaction between the number of tracks and PB is no longer significant; however, the triple interactions (number of tracks, PB, and whether schools consider prior performance) are. The table indicates that it is primarily the schools that always consider prior performance in systems with multiple tracks that mitigate the relation of PB with performance. Thus it is not tracking itself that diminishes the association of PB and performance but, rather, tracking combined with whether schools consider prior performance. This is consistent with our expectation that when schools consider prior performance, parents have less influence on their child's track choice and subsequent performance.

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

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

to students with high PB, already have an efficient match between their ability and track placement.

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

Model:	(1) FE	(2) FE	(3) Low PB	(4) High PB
School considers prior performance				
<i>Sometimes</i>	-0.56 (1.78)	-1.14 (1.88)	-2.86 (2.12)	-0.70 (1.76)
<i>Always</i>	7.65*** (2.70)	6.95*** (2.70)	3.65 (3.16)	7.64** (3.37)
Number of tracks (0-4)			2.84 (4.09)	4.04 (3.72)
Parental background* Number of tracks	-3.24*** (1.13)	-1.90 (1.29)	-0.50 (1.78)	-3.08** (1.31)
Parental background* Number of tracks* <i>Sometimes</i>		-1.05* (0.55)	-1.17 (0.74)	-0.36 (1.13)
Parental background* Number of tracks* <i>Always</i>		-2.04*** (0.75)	-2.54** (1.19)	-0.86 (0.93)
Country FE	√	√	-	-
(Pseudo) LL	-43,677	-43,687	-43,311	-43,918
# of students	187,768	187,768	95,406	92,362
# of countries	31	31	31	31

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

3.6.4 Robustness

Other possible distorting factors are our measure of tracking, the sample of countries and the control variables. We report all robustness checks with the PISA mathematics score, unless otherwise stated the results for reading and science are robust to the various checks.³⁵ The results which are not shown in this chapter are available on request.

Unlike some other papers on this topic, we use the number of tracks available to students at the age of 15 to characterize a country's tracking regime. Since the PISA

35. Most of the changes are changes in the coefficient for the dummy indicating that the school always considers prior performance. The main result, that schools that consider prior performance for student acceptance mitigate the relation between strong tracking and performance, holds in any specification.

test is conducted when the students are between 15 and 16, it is possible that although multiple tracks are available to students at the age of 15, students have not yet been tracked for a substantial amount of time. If this is the case, the association with tracking in late selection countries may be too weak to be picked up. To check for this, we redo the analysis using only late trackers.³⁶ The results are presented in Appendix J. The results for the late-tracking countries have the same sign and are similar in coefficient size, although the interactions are insignificant for mathematics and science, possible due to the large drop in observations.

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

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

3.7 Conclusion

The variation in tracking in education systems throughout the Western world is quite large: Many countries have no formal tracking in secondary school (although most have some form of non-institutionalized tracking), while some countries distinguish up to five tracks for students. Also the manner in which tracking is implemented on the school level differs widely. Some countries put more emphasis on the use of performance to place students into the available tracks (e.g. the Netherlands), while in other countries parents have more influence on the track their child will go (e.g. Germany). In this chapter it is argued that the inconsistencies between the theory of

36. Late-tracking countries, which select at age 14 or later, are Croatia, Denmark, Estonia, Finland, Greece, Iceland, Ireland, Israel, Italy, Latvia, Lithuania, New Zealand, Norway, Poland, Portugal, the Russian Federation, Slovenia, Spain, Sweden, and the United States.

tracking and some of the empirical results in the literature could be explained by country differences in track placement. When track placement is done not based on an ability measure but mainly on PB, or, related, residential areas, the theoretical benefits of tracking might not arise. These theoretical benefits rely heavily on the idea that tracking leads to more homogenous ability classes which the teacher is better able to teach and where students benefit from a curriculum tailored to their needs and abilities. When students are placed into tracks based not on prior performance but on PB, classes will be more heterogeneous and students might not be taught a fitting curriculum. Furthermore, when PB is used to decide on track placement, educational opportunities are affected and inequality is likely to grow.

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

The association between PB, tracking, and student performance shows that equality of opportunity is best provided for in a system with a high number of tracks combined with schools always consider prior performance on accepting the student to the school. It turns out that for high PB students in these systems, tracking weakens the positive relation between PB and performance, whereas for low PB students the (for them negative) relation between tracking and performance is lowered primarily when they attend schools that always consider prior performance. Thus it seems that high PB students might be harmed by tracking when schools consider prior performance.

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

whether schools consider prior performance in accepting the student, tracking benefits both student performance and educational opportunities.

The Effect of Track Placement on Cognitive and Non-Cognitive Skills³⁷

Abstract: Tracking in education is used to tailor education to the capabilities and the needs of each child. If every child is assigned to the track that fits his needs best, one would expect that children at the margin would be indifferent between the two tracks at stake. The aim of this chapter is to investigate the effect of being in the higher track for students at the margin for a wide set of outcomes, including both cognitive and non-cognitive outcomes. For the analysis, we use a longitudinal dataset on cognitive and non-cognitive skill development in both elementary and secondary education in a Dutch region. We apply a fuzzy regression discontinuity design using the discontinuity in a test score and a teacher recommendation in the assignment to tracks. Our main finding is that track placement influences IQ, the reading skills development and the self-perceived probability to obtain the degree for the marginal student but has no effect on personality traits, other non-cognitive skills and mathematics. Track mobility does not counteract the initial track placement.

37. This chapter is based on joint work with Lex Borghans and Trudie Schils. We would like to thank Bas ter Weel, participants of the Economics of Education group in Maastricht University, of the International Workshop on Applied Economics of Education 2014, of the CEPA PhD workshop (Stanford University), and of the AMCIS conference on Educational Systems in 2014 for useful comments.

4.1 Introduction

Tracking in education is used to tailor education to the capabilities and the needs of each child. If every child is assigned to the program that fits his needs best, one would expect that children at the margin would be indifferent between the two tracks at stake. In practice however, parents and children tend to put in a lot of effort in getting into higher tracks. This suggests that, at least from their perspective, the high track is more attractive than the middle track for a larger group of students. Getting into this higher track would then qualify as a tournament.

The aim of this chapter is to investigate the effect of being in the high track for students at the margin for a wide set of outcomes, including both cognitive and non-cognitive outcomes. We apply a fuzzy regression discontinuity design (RDD; Imbens and Lemieux, 2007) using a discontinuity in a test score and the teacher recommendation in the assignment to tracks. Our main finding is that track placement influences IQ, the reading skills development and the self-perceived probability to obtain the degree for the marginal student but has no effect on personality traits, other non-cognitive skills and mathematics. Track mobility does not counteract the initial track placement.

For the analysis we use a longitudinal dataset on cognitive and non-cognitive skill development in both primary and secondary education in a Dutch region. In the Netherlands students are placed into tracks between 6th (elementary school) and 7th grade (secondary school). This dataset contains the two main sources of information Dutch secondary schools receive from the elementary school to decide on track placement: the score of a uniform elementary school exit test and the elementary school teacher recommendation. We exploit these two signals to look at the marginal student. However, secondary schools differ somewhat in the assignment procedures they adhere to, which does not allow for a sharp RDD. Each school is free in its student acceptance policies, although all are required by law to use the two sources of information received from the elementary schools. To check for bias due to remaining endogeneity in the tracking decision, for a number of outcomes variables we additionally use available panel information. For several outcome variables we have similar measures in both before and after track placement.

This chapter contributes to the literature on tracking, but is also closely related to issues on ability grouping and selective schools.³⁸ The literature on the effects of tracking, streaming, and ability grouping is very extensive and can be divided into

38. Selective schools can be considered as the higher track, for instance when they prepare students for university entrance exams (ie. the so-called preparatory schools). These so-called preparatory schools are quite common in France, but exist also in the United States, United Kingdom and Canada.

papers which look at the effects of a substantial increase in the number of students entering the higher track or those looking at the marginal student who moves track.³⁹ The papers which look at a substantial inflow of lower ability students into the high track show, besides the tracking effect, also the effects of a changing composition of the high track since more lower ability peers are allowed into the higher track. Guyon, Maurin, and McNally (2012) and Van Elk *et al.* (2011) look at such an increased inflow of students into the high track in Northern Ireland and the Netherlands and find positive effects on outcomes of these students. Duflo, Dupas, and Kremer (2011) find, using an experiment in Kenya in which groups of students were assigned to a school with and without ability grouping, that ability grouping has positive overall effects on cognitive outcomes.

This chapter does not look at the effects of a substantial increase in the number of students going to the high track, but focusses on the marginal student who does or does not go to this high track. Consequently, this study and related studies are able to isolate the treatment effect of being in the high track on the individual student since the composition of the higher track does not change when the marginal student enters the high track. An example of a similar paper is Borghans *et al.* (2011). They show that the threshold in the Netherlands for the highest track is too high: Students below the threshold would benefit from being in the high track both in test scores and in later earnings. Dustmann *et al.* (2014) use month of birth as an instrument for track placement and show, using a reduced form, that month of birth has no effects on labor market outcomes. Pop-Echeles and Urquioia (2011) and Jackson (2010) use formal assignment rules in Romania and Trinidad and Tobago to instrument attendance of better achieving, or more selective, schools.⁴⁰ Both find that pupils in better schools have higher test scores at the end of secondary school. Jackson (2010) also finds that students in better schools pass more exams and more often earn a certificate that gives access to university, while Pop-Echeles and Urquiola (2011) also look at behavior aspects and find that better teachers sort into better schools, parents at those schools are more involved, children do more homework, and child's self-perception is more positive.

39. Some studies on tracking, like Hanushek and Woessmann (2006) and Ariga and Brunello (2007), focus on the overall effects of tracking which compare different tracking policies across countries. A number of other papers on tracking make use of different tracking policies within one country, often due to policy changes, to look at the effect of tracking. E.g. Pekkarinen (2008), Pekkarinen *et al.* (2013), and Hall (2012) who all find little effect. For ability grouping, Betts and Shkolnik (2000) find that only the grouped classes with average ability suffer from grouping, while there is no effect for the lower ability groups and a small positive effect for the high ability grouped classes. Figlio and Page (2002) find no negative effect of ability grouping for low-ability students and find some evidence they might even benefit from ability grouping.

40. See Hoekstra (2009) for similar analyses for entry into selective colleges.

There is a growing literature which analyzes the relation between non-cognitive skills, for instance the big 5 personality traits or motivation, and student performance (e.g. Heckman and Rubinstein, 2001; Heller *et al.*, 2012). These non-cognitive skills are also shown to influence later outcomes (e.g. Heckman *et al.*, 2012; Heckman and Rubinstein 2001). However, notwithstanding this growing awareness of the importance of non-cognitive skills, little is known about the effects of education on non-cognitive skills. The contribution of this chapter is that we look at the marginal student who is just able to go to the high track and we look at a wide set of both cognitive and non-cognitive outcomes, while we also use information on the same outcome variables before tracking has taken place.

The structure of this chapter is as follows: Section 4.2 will elaborate on the dataset and the graphical analysis. The model and results are provided in Section 4.3. Section 4.4 concludes.

4.2 Data and graphical analysis

The data used in this chapter are the result of a cooperative project with schools, schools boards and municipalities in which almost all elementary and secondary schools in Zuid-Limburg, a region in the South of the Netherlands, participate. The data comprise the cohort of students that were in the 6th grade in 2009 (last grade of elementary school) and in the 9th grade in 2012 (first grade of secondary school). Students enter the tracked system in the 7th grade which comprises three main tracks, with some further subdivisions in mainly the lowest track.⁴¹ In 2011, a little more than fifty percent of the students aged 15 attended the lowest track; another 20 percent the middle track and twenty-five percent of students was in the highest track (CBS, 2012, Figure 1.2.4). In this chapter, we focus on the two upper tracks in which a total of 45 percent of students are enrolled. For the students in the sample, the data include extensive information, including non-cognitive skills, reading, math, and IQ test scores in both 6th and 9th grade.⁴² The data also include the information on the elementary

41. The three tracks are VMBO, HAVO, and VWO. VMBO is preparatory middle-level vocational education which lasts 4 years, and consists of the sub-tracks pure practical education (pro), VMBO-basic profession-oriented, VMBO-middle management-oriented, VMBO-mixed, and VMBO theoretical. HAVO is higher general continued education and lasts 5 years. VWO is preparatory scholarly education and lasts 6 years. VWO is split into the sub-tracks athenaeum and gymnasium which are essentially the same, except that gymnasium students also have the courses Latin and/or Greek. Secondary schools with only students of a single track and schools with multiple tracks exist alongside each other, although the tracks could be separated across different school buildings. This is especially the case for the bottom track. In the first year of secondary school, or sometimes in the first two years, so-called bridge classes exist in which students of multiple tracks are grouped together, but these classes only rarely consist of more than two tracks.

42. Not all children received the complete student questionnaire, resulting in a smaller sample for civic engagement and school well-being questions. Also, not all children took all tests or all test questions. Using IRT test scores are put on the same scale for all children who saw 13 or more test questions on each of the tests. We use the expected posterior estimates using a 2 parameter Bayesian Markov Chain Monte Carlo model.

school exit test and the elementary school teacher recommendation which is necessary for our identification strategy. Finally, information on the socio-economic background of the student's parents and information about the school is available.

The dataset contains 9,124 students in 9th grade of secondary school, and for 5910 we also know in which track they were in 7th grade (the first grade of secondary school). We focus here on the top two tracks which gives us 1,067 in the high track and 2,151 in the middle track.⁴³ Of these 3,218 students for 42 students we miss both their elementary school exit test score and their elementary school track recommendation, leaving us with 2,117 in the middle track and 1,059 in the highest track. We use the full sample, and do not restrict our sample to those within a small bandwidth around the cutoff, to obtain more precision (Lee and Lemieux, 2010).

Table 1 shows the descriptive statistics on some key variables, separated for students in the highest and the middle track.⁴⁴ Students in the two tracks differ in some respects: compared to students in the middle track students in the highest track not only have higher IQ and higher reading and math test scores, they also have higher perseverance, social skills, are more open, have a higher self-perceived probability to obtain a secondary school degree, are less positive about their labor market chances, and have higher educated parents. To see whether these differences occur due to selection or due to being in the high track is the goal of this chapter.

Acceptance and track placement of students in secondary school is guided by the Dutch government: Each elementary school is required to send to the preferred secondary school of the student the elementary school teacher recommendation for track placement, and a second independent and objective measure (Kingdom of the Netherlands, 1981). To obtain this independent and objective measure at the end of the last grade in elementary school almost all students take a centralized exit test (the so-called CITO test).⁴⁵ The elementary exit test score ranges from 500 to 550 and the guidelines for the highest track state that a score of 538 is needed to go to the highest track and a score of 533 to go to the second highest track (CITO Score, 2014). The mean test score in the highest track in our sample is 547 and for the middle track 540, with considerable variation.

43. The remaining 2692 students are in the bottom track or among the 167 students who are in the upper two tracks but who repeated the 7th grade. Of this last group we have no elementary school data (and for 11 students we have also no elementary school exit score) and since these students entered the school the year before, the threshold which they faced was different from the threshold of the other students. For these reasons we dropped them.

44. Appendix K provides the items on which the variables are based.

45. It is not prescribed which independent and objective measure is needed and thus multiple elementary school exit tests are used. However, for eighty-five percent of schools this second objective measure is the CITO test score (CITO, 2014).

Table 1: Descriptive statistics of students in the 9th grade

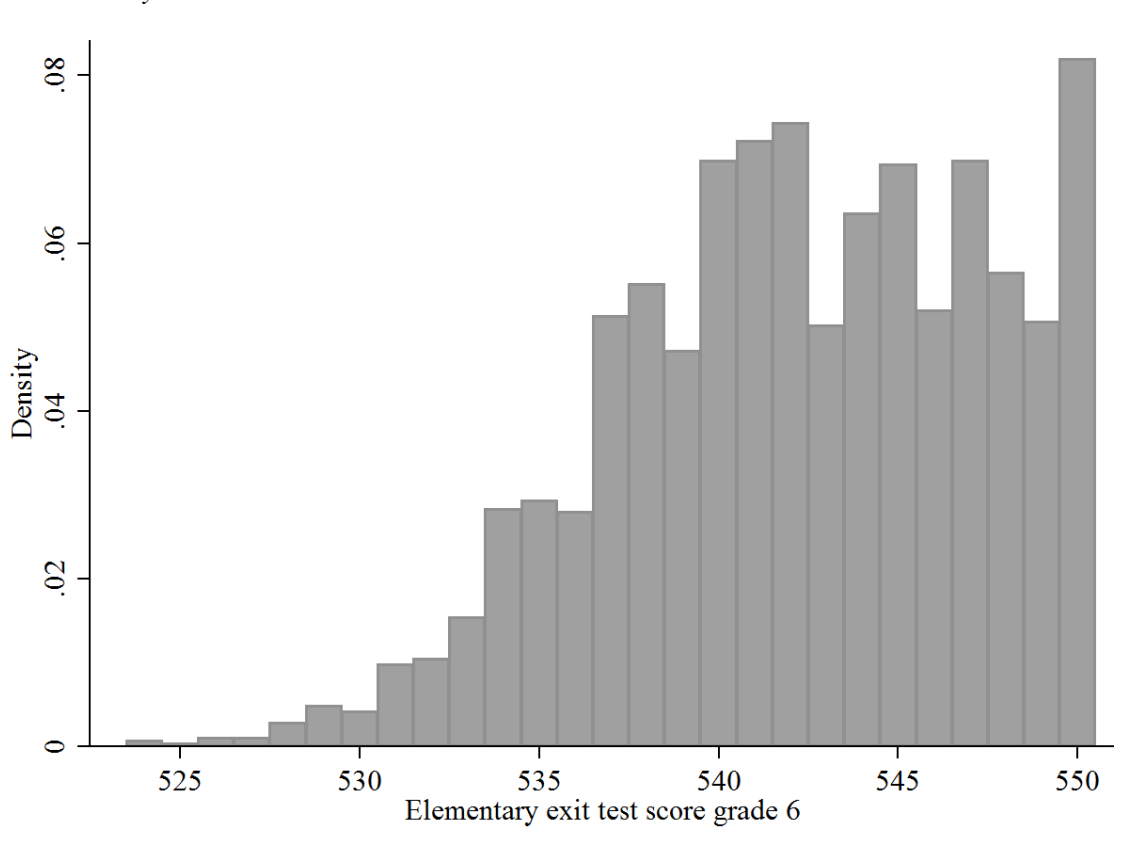
	Middle track (HAVO)					Highest track (VWO)					Total obs	Dif in means
	Obs	Mean	Std. Dev.	Min	Max	Obs	Mean	Std. Dev.	Min	Max		
<i>Age</i>	1618	181.28	5.03	164.00	206.00	895	180.36	4.98	160.00	200.00	2513	0.92***
<i>Gender</i>	1585	0.54	0.50	0.00	1.00	888	0.50	0.50	0.00	1.00	2473	0.04*
<i>Parental education</i>	1447	2.52	1.03	1.00	4.00	811	2.89	1.01	1.00	4.00	2258	-0.37***
<i>Elementary school exit test score</i>	1618	540.40	4.91	524.00	550.00	895	546.01	3.72	527.00	550.00	2513	-5.61***
<i>Recommendation</i>	1618	15.84	1.83	4.00	19.00	895	17.81	1.08	3.00	19.00	2513	-1.97***
<i>IQ</i>	644	-0.23	0.98	-3.70	1.97	531	0.30	0.90	-3.70	1.97	1175	-0.53***
<i>Math^a</i>	1197	0.26	0.79	-2.15	2.75	665	0.60	0.88	-2.01	3.35	1862	-0.34***
<i>Reading^a</i>	1244	0.11	0.82	-3.03	2.09	683	0.54	0.97	-3.06	2.24	1927	-0.43***
<i>Track grade 9</i>	1618	5.14	0.78	1.00	6.00	895	5.86	0.37	4.00	6.00	2513	-0.72***
<i>Openness</i>	515	0.59	0.15	0.13	0.94	439	0.64	0.16	0.13	1.00	954	-0.05***
<i>Conscientiousness</i>	515	0.53	0.17	0.00	1.00	439	0.52	0.18	0.00	0.95	954	0.01
<i>Extraversion</i>	515	0.58	0.18	0.06	1.00	437	0.57	0.20	0.00	1.00	952	0.01
<i>Agreeableness</i>	515	0.74	0.16	0.00	1.00	436	0.75	0.17	0.00	1.00	951	-0.01
<i>Neurotism</i>	515	0.40	0.21	0.00	1.00	438	0.38	0.21	0.00	1.00	953	0.01
<i>Perseverance</i>	513	0.65	0.18	0.13	1.00	436	0.67	0.17	0.06	1.00	949	-0.02**
<i>Competitive spirit</i>	513	0.88	0.15	0.25	1.00	432	0.87	0.16	0.00	1.00	945	0.01
<i>Social skills</i>	499	0.62	0.09	0.00	0.90	433	0.63	0.08	0.25	0.90	932	-0.01***
<i>Soc. sk.: Social</i>	498	0.60	0.09	0.00	0.90	433	0.62	0.09	0.24	0.90	931	-0.02***
<i>Soc. sk.: Action</i>	499	0.64	0.11	0.00	0.90	431	0.64	0.10	0.14	0.90	930	0.00
<i>Self- p prob of obt degree</i>	496	4.22	0.91	1.00	5.00	424	4.34	0.83	1.00	5.00	920	-0.12**
<i>Prob of job 1</i>	111	3.35	0.96	1.00	5.00	54	3.02	0.84	1.00	4.00	165	0.33
<i>Prob of job 2</i>	111	4.09	0.71	2.00	5.00	54	4.11	0.60	2.00	5.00	165	-0.02
<i>Sch satisfaction</i>	1367	6.49	1.36	1.00	10.00	723	6.61	1.38	1.00	10.00	2090	-0.11*
<i>School motivation</i>	1402	0.68	0.13	0.13	1.00	754	0.72	0.13	0.13	1.00	2156	-0.04***

Notes: How the outcome variables are defined is discussed in Appendix K. ^a Students made a math and a reading test, but not all students had the same questions. To ensure all students receive a test score on the same scale we used IRT to rescale the test scores. In italics, the variables for which there is a significant difference between children in the two tracks. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Figure 1 shows the density of the test score and the elementary school recommendation for students entering in the top two tracks. From Figure 1a it is clear that a ceiling effect occurs: The density of the test score is negatively skewed and many students get scores in the top range of the scale. The same can be seen in Figure 1b for the elementary school teacher recommendation, although to a lesser extent. Figure 2 shows that both test score and the elementary school recommendation clearly influence track placement, as the probability to enter in the highest track increases with both measures. Since there is no predefined cut off point, we apply a fuzzy RDD which assumes that, although the probability to enter in the highest track does not jump to 1 after the cut off, the probability increases for larger values of the forcing variable (Imbens and Lemieux, 2007).⁴⁶ Using a fuzzy RDD we essentially instrument track placement by passing the threshold of the forcing variables: the elementary school exit test score and the elementary school teacher recommendation.

Figure 1: Density of forcing variables

1a: Elementary school exit test score



46. Given our fuzzy RD and the lack of a (predetermined) cutoff we do not have to worry about students trying to manipulate their score to be above the cutoff. It is in all students' best interest to have the highest possible exit test score and teacher recommendation.

1b: Elementary school teacher recommendation

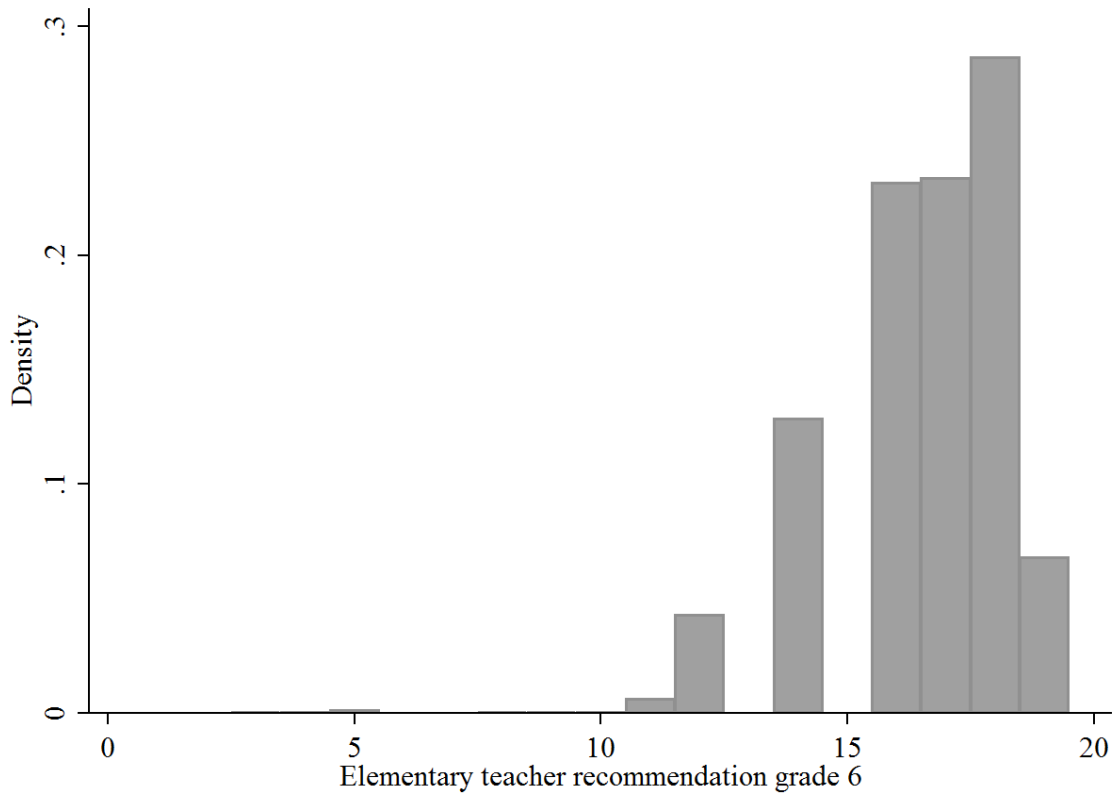
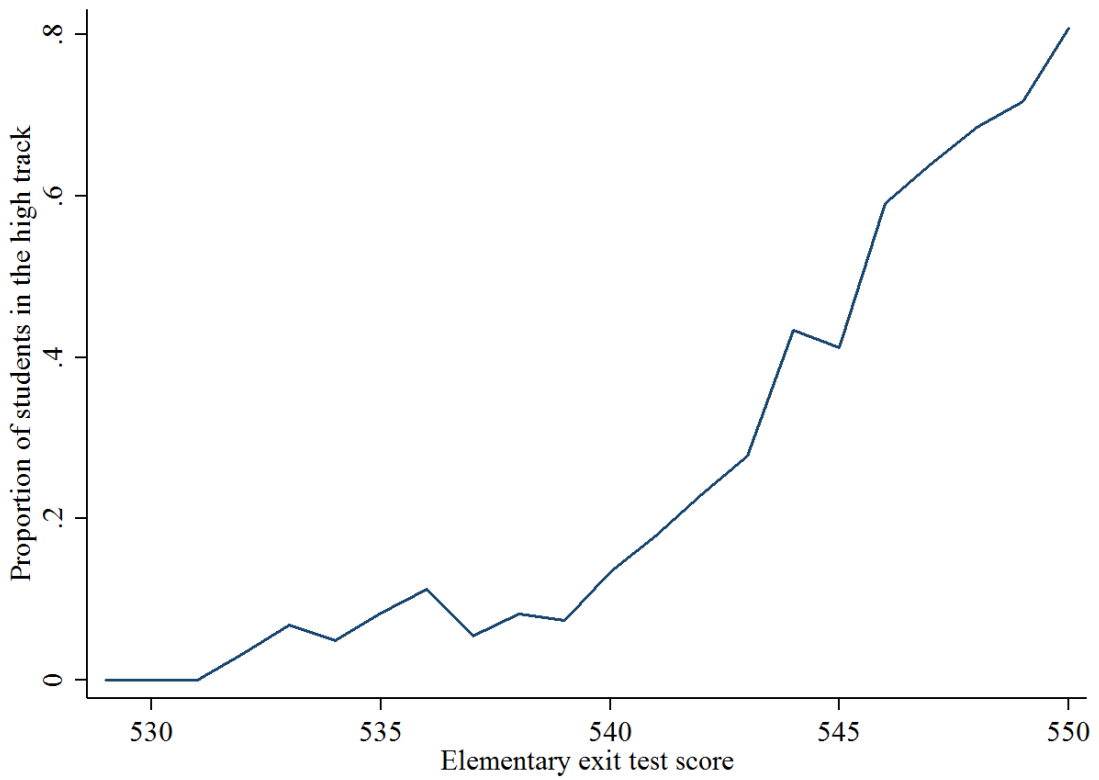
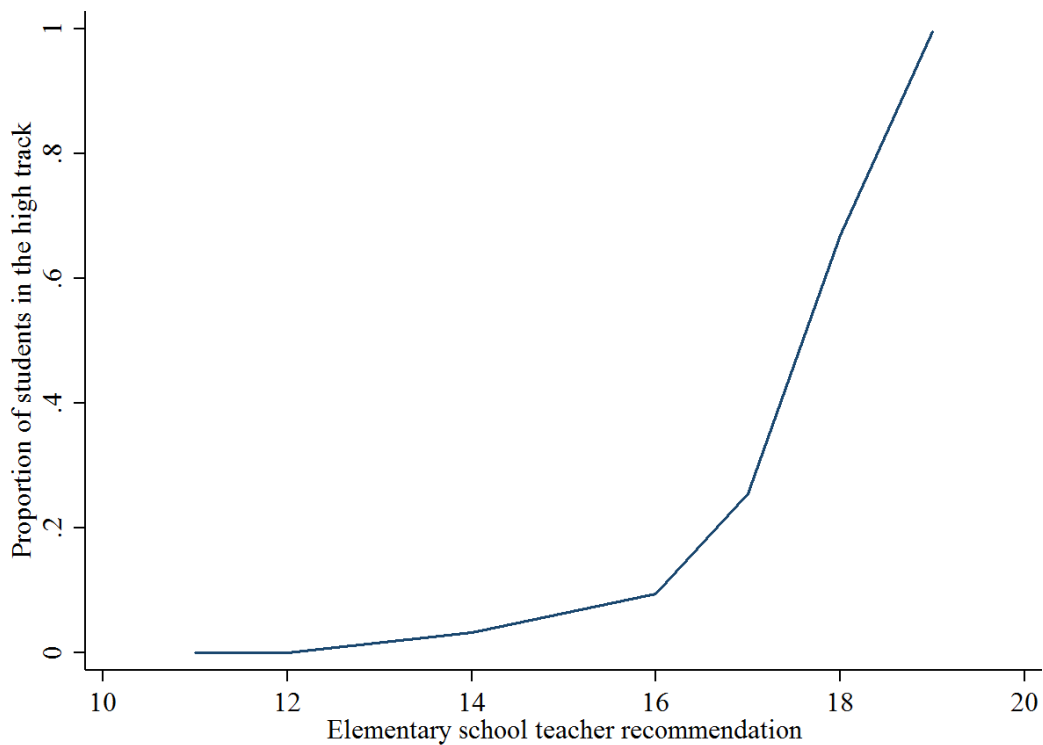


Figure 2: Fraction of being in the highest track over the forcing variables

1a: Elementary school exit test score



1b: Elementary school teacher recommendation

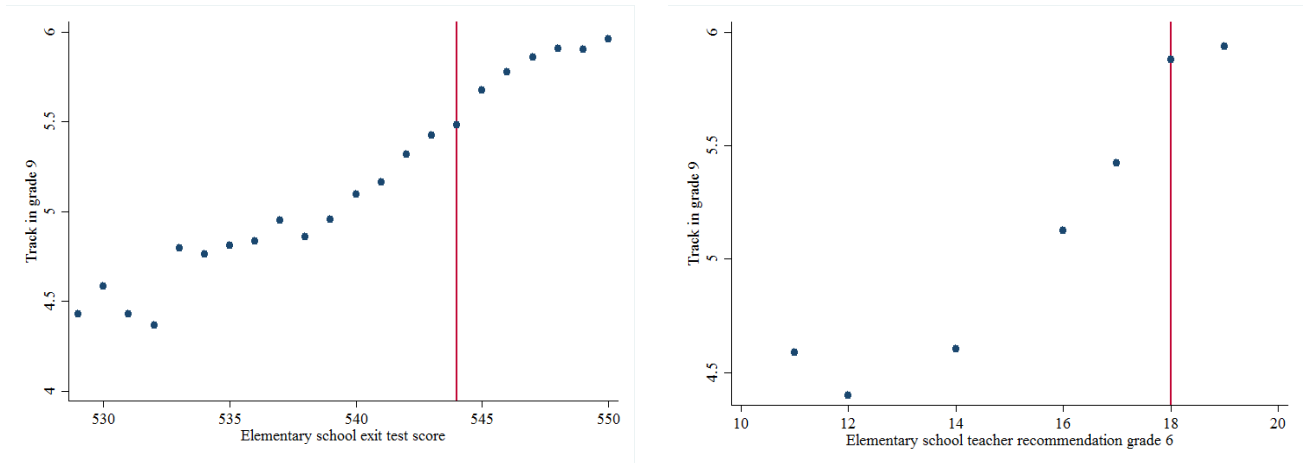


Notes: Only students with at least a reading or math score are included. Primary exit test scores which less than 10 students scored are not included.

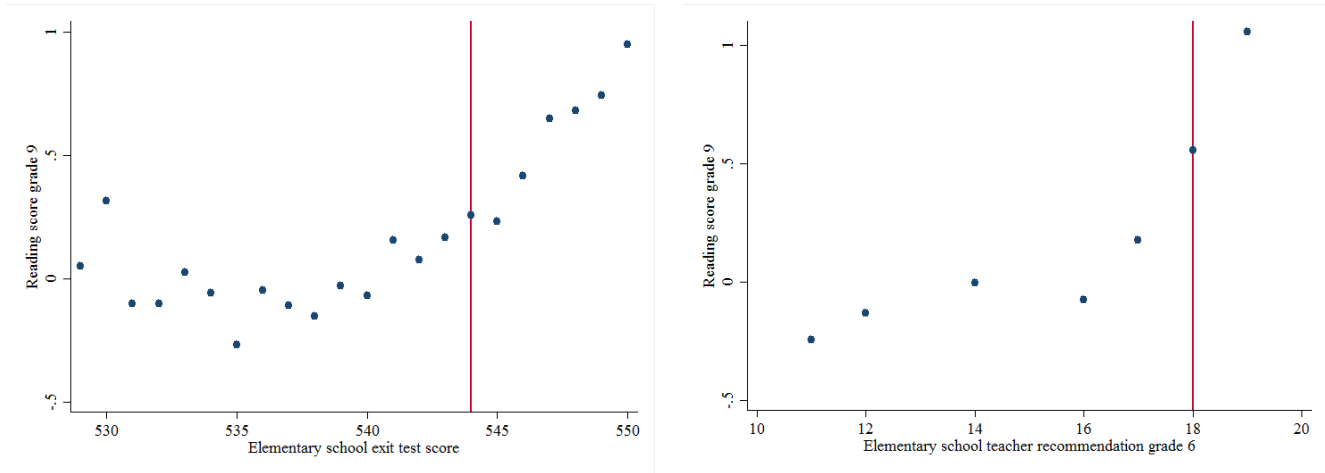
Figure 3 shows for a number of outcome variables the average value over our two forcing variables: the exit test score and the elementary school teacher track recommendation. In these figures we see a clear semi-linear trend for the track in grade 9 (Figure 3a) and the reading test scores (Figure 3b) over the exit test score and the elementary school teacher recommendation. This trend is less clear for extraversion and social skills (Figure 3c and 3d). Since we use a fuzzy RDD, these figures are not very informative about the effects of being in the higher track for the marginal student. Using both the discontinuity in the test score and the elementary school recommendation, Section 4.3 shows that track placement has an effect on a number of outcome variables.

Figure 3: Average outcome values over the forcing variables

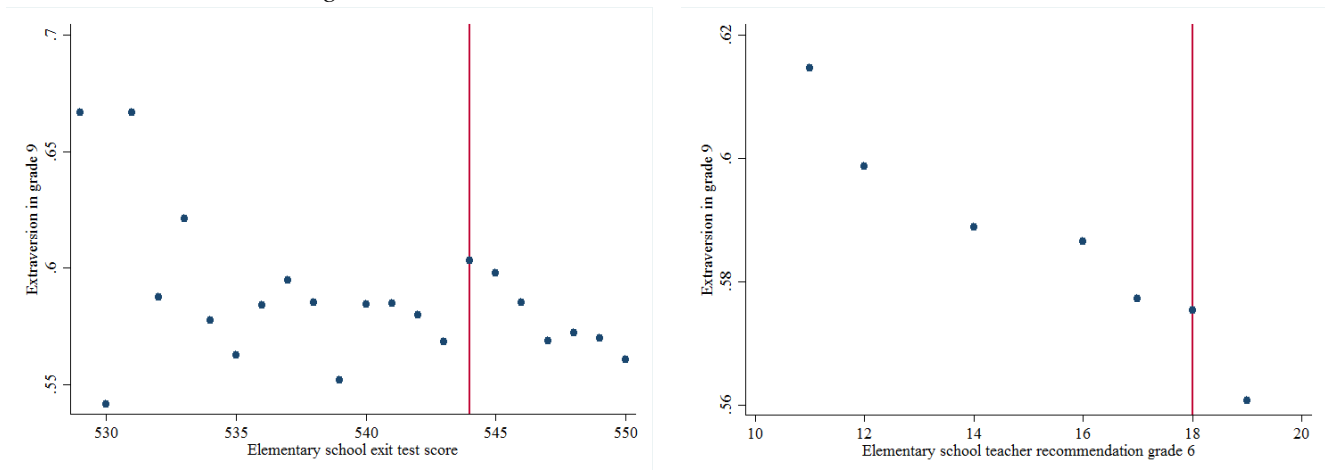
3a: Track in grade 9



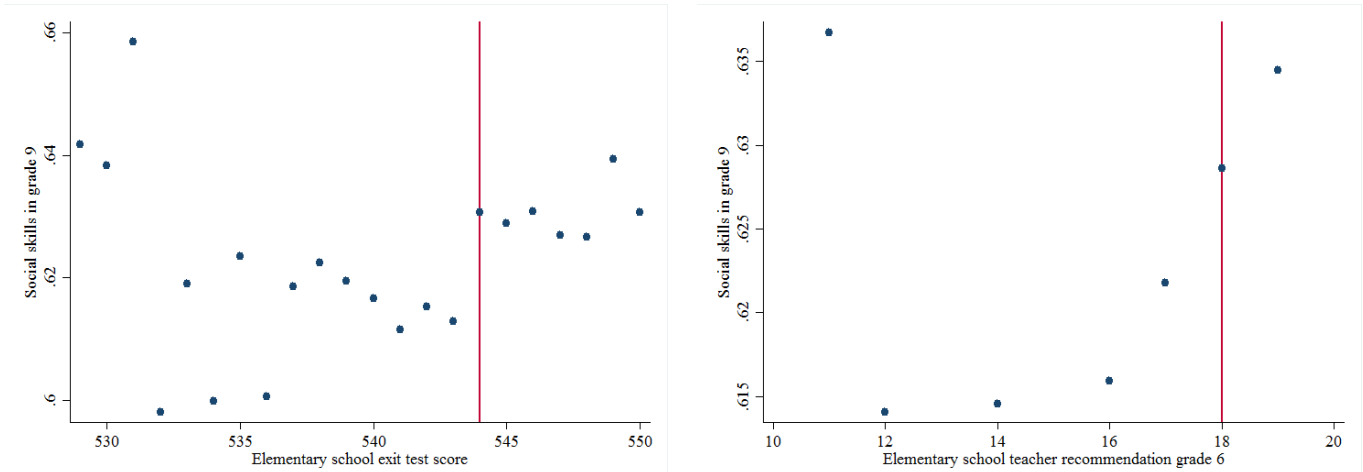
3b: Reading score in grade 9



3c: Extraversion in grade 9



3d: Social skills in grade 9



Notes: Red bar is the threshold. The figures are only for those elementary school exit test scores and elementary school teacher recommendations which more than 10 students received.

Figures similar to Figure 3 to illustrate that background characteristics of the students do not influence track placement, or in other words that the students are identical around the threshold, are equally uninformative. To tackle this we use our IV strategy to predict some background characteristics, and Table 2 shows that we fail to do so showing that students are identical around the threshold.

Table 2: Identical students around the threshold

Dep var:	Gender	Age	Parental edu	Work Father	Work mother	Tradition al family	IQ
High Track	0.09 (0.14)	-2.62 (2.11)	0.65 (0.48)	0.11 (0.20)	-0.12 (0.19)	-0.21 (0.18)	-0.09 (0.15)
Test score	-0.01** (0.01)	0.088 (0.07)	-0.01 (0.02)	-0.00 (0.01)	0.01 (0.01)	0.01 (0.01)	0.07*** (0.01)
Recommendation	0.02 (0.01)	-0.02 (0.29)	0.04 (0.04)	-0.04* (0.03)	0.01 (0.03)	0.02 (0.02)	0.03 (0.02)
Constant	5.86** (2.80)	135.10** * (39.34)	8.60 (10.23)	2.97 (3.98)	-2.43 (4.12)	-3.88 (3.98)	35.56*** (5.45)
# of students	1,175	1,175	923	911	817	970	1,175
# of schools	17	17	17	17	17	17	17
R ²	-	-	0.01	-	0.00	-	0.13
F 1 st stage	15.25	15.25	9.72	8.44	6.42	8.56	15.25

Notes: The table presents coefficients (robust standard errors in parentheses) from IV models using the IQ sample. The first stage results are not shown for these samples, but Table 4 shows the first stage for the IQ sample. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

For some outcome variables (all measured in 9th grade) we have similar information available from 6th grade, the year before the students are being tracked. The panel dimension of our data is illustrated in Table 3, which provides descriptive statistics of the same variables as in Table 1, but now measured in elementary school in 6th grade. Again these statistics are separated by track, but at this age the students were still grouped together and the division into middle track versus highest track here is therefore merely an illustrative division. In 6th grade, the students who later entered the highest track had higher IQ, were more open, more agreeable, and had more social skills. However, they were also less extraverted and felt less at home at school. Between 6th and 9th grade we see that personality and school related measures change for students. This personality change over time is found more often (e.g. Roberts *et al.*, 2006), but it is yet unclear whether it is due to age differences in personality or due to changing environments over time, for instance entering in a new school or school type. In the last column of Table 3 the significant difference for students in the two tracks between the difference in values between 6th and 9th grade are shown. For instance, openness for students in the middle track was 0.02 points higher in 9th grade compared to 6th grade, while for students in the highest track this difference is -0.01. Students in the highest track became less open between 6th and 9th grade compared to students in the middle track, and this difference is significant at a 2%-significance level.

Table 3: Descriptive statistics of students in the 6th grade

	Middle track (HAVO)					Highest track (VWO)					Total obs	Diff in means	P-value diff grade 6-9 by tracks
	Obs	Mean	Std. Dev.	Min	Max	Obs	Mean	Std. Dev.	Min	Max			
<i>IQ</i>	644	-0.18	0.98	-4.20	2.01	531	0.24	0.98	-4.52	2.01	1175	-0.43***	0.11
<i>Openness</i>	515	0.57	0.16	0.06	1.00	439	0.65	0.16	0.19	1.00	954	-0.08***	0.02
Conscientiousness	513	0.87	0.14	0.38	1.00	432	0.88	0.13	0.38	1.00	945	-0.01	0.06
<i>Extraversion</i>	515	0.52	0.16	0.06	1.00	439	0.54	0.17	0.05	1.00	954	-0.02	0.01
<i>Agreeableness</i>	515	0.48	0.13	0.00	0.81	437	0.45	0.13	0.06	0.94	952	0.02***	0.34
Neurotism	515	0.64	0.10	0.19	0.94	436	0.65	0.09	0.44	1.00	951	-0.01**	0.70
Perserverance	515	0.40	0.19	0.00	1.00	438	0.38	0.20	0.00	1.00	953	0.02	0.61
Competitive spirit	513	0.52	0.10	0.13	0.88	436	0.52	0.09	0.25	0.75	949	0.00	0.11
<i>Social skills</i>	499	0.49	0.10	0.21	0.75	433	0.52	0.09	0.27	0.75	932	-0.03***	0.09
<i>Soc. sk.: Social</i>	498	0.50	0.11	0.07	0.75	433	0.54	0.10	0.18	0.75	931	-0.04***	0.01
Soc. sk.: Action	499	0.49	0.12	0.15	0.75	431	0.49	0.12	0.15	0.75	930	0.00	0.90
Prob of obt degree	496	3.10	0.55	1.00	4.00	424	3.11	0.52	1.00	4.00	920	-0.01	0.08

Notes: In italics the variables for which there is a significant difference between children in the two tracks. In bold the variables for which there is a significant difference between the change between grade 6 and grade 9. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The last column shows the p-values of the difference between grade 9 and grade 6 for students in the middle and those in the highest track.

4.3 Analyses

Our model combines the advantages of a fuzzy RD design with a panel dimension. Using the cut off observed in the data for both the 6th grade test score and the elementary school teacher recommendation, we apply a fuzzy RD design in which we instrument track placement in 7th grade by passing the threshold for the elementary school teacher recommendation, or advice, and test score to study a number of outcomes using equation (1) and (2a). However, unlike Pop-Echeles and Urquoa (2011) and Jackson (2010), who use formal assignment rules to instrument selective school attendance, in the Netherlands no centralized cut off point is set. Schools are obliged to base their track placement decision on the elementary school exit test and the elementary school teacher recommendation, but each school is free to set its own cut off point with regards to its supply of students. We therefore instrument track placement in 7th grade by the two signals secondary schools receive to decide on track placement (Imbens and Lemieux, 2007). Some remaining endogeneity may still exist, for instance when schools deviate from the placement guideline for the test score and the elementary school teacher recommendation with reason. Therefore we also use the panel dimension of this data to limit our measurement error and remove any remaining selection: By controlling for the grade 6 outcome variable we only make use of the change in the outcomes variable due track placement (equation 1 and 2b).

Since there is no official elementary school exit test thresholds for which above the student automatically goes to the higher track, we use the test score for which we find the strongest link between track placement in 7th grade and an indicator function of having a test score above the cut off.⁴⁷ The analysis reveals that 544 is the unofficial cut off as seen in the data, and we subsequently use this cut off as if it was the cut off used by schools. For the elementary school teacher recommendation we use as cut off the recommendation that states that the child should go to the highest track.⁴⁸

We estimate the following model:

$$HighTrack_i = \gamma + \delta_1 I(TEST \geq 544)_i + \delta_2 I(RECOM \geq 18)_i + \mathbf{X}_i \theta + \epsilon_i \quad (1)$$

$$Y_{i,t=9} = \alpha + \beta_1 \widehat{HighTrack}_i + \beta_2 TEST_i + \beta_3 RECOM_i + \epsilon_i \quad (2a)$$

$$Y_{i,t=9} = \alpha + \beta_1 \widehat{HighTrack}_i + \beta_2 TEST_i + \beta_3 RECOM_i + \beta_4 Y_{i,t=6} + \epsilon_i \quad (2b)$$

47. The cut-off with the strongest link between track placement and the indicator function is the cut-off for which the F statistic reveals the strongest link.

48. There are actually two categories that related to a elementary school teacher recommendation of the highest track. Recommendation 18 refers to the VWO-atheneum, and recommendation 19 refers to the VWO-gymnasium, or bilingual education. A elementary school teacher recommendation for a bridge class of HAVO and VWO (the two upper tracks) is categorized as a recommendation for the middle track.

where $HighTrack_i$ is an indicator whether the student was placed into the high track in 7th grade and is estimated in equation (1) and the fitted values from equation (1) ($\widehat{HighTrack}_i$) are used as an explanatory variable in equation (2a) and (2b). $HighTrack_i$ is estimated using the two data thresholds: the test score and the elementary school teacher recommendation. The matrix X_i contains the included instruments, i.e. the two forcing variables when using equation (1a) and also the 6th grade outcome variables when using equation (1b). $Y_{i,t}$ is an outcome variable in grade $t=6$ or $t=9$ (for instance a reading test score or extraversion), and $TEST_i$, the individual test score, and $RECOM_i$, the elementary school teacher recommendation for the student, are our two running variables. In Appendix L, we supplement this main model with models where either the test score or the elementary school teacher recommendation is used as an instrument, or where an interaction term is included for students who have both a test score and an elementary school teacher recommendation above the threshold. These models show qualitatively similar results.

The first stage is depicted in the first column of Table 4 and shows that students with higher test scores and with higher elementary school teacher recommendations are more likely to be in the high track as is to be expected. Our two instruments (having a test score of greater or above 544 and an elementary school teacher recommendation greater or above 18) are both highly significant in predicting track placement in 7th grade. Judging by the amount of explained variation the test score has more predictive power than the elementary school teacher recommendation on its own, and together they have the most explanatory power with an F statistics of 15, well above the required F statistic of 10 as proposed by Staiger and Stock (1997) and later refined by Stock and Yogo (2005). Depending on the dependent variable in the second stage the sample will change, and subsequently also the corresponding F statistic of the first stage changes. For that reason, all tables will also include the F statistic of the excluded instruments. We also estimated models with either the test score (F statistic of 18) or the elementary school teacher recommendation (F statistic of 9) as instrument, or where an interaction term is included for students who have both a test score and an elementary school teacher recommendation above the threshold (F statistic of 10). See Appendix L for these results.

Table 4: First stage results following equation (1a).

	High track	High track
I(Test score \geq 544)	0.13*** (0.05)	0.13*** (0.05)
I(Recommendation \geq 18)	0.34** (0.15)	0.34** (0.15)
Test score	0.01** (0.00)	0.01* (0.00)
Recommendation	0.06** (0.02)	0.06** (0.02)
Grade 6 IQ		0.00 (0.01)
Constant	-4.45** (1.79)	-4.33** (1.76)
# of students	1,175	1,175
# of schools	17	17
R ²	0.44	0.44

Notes: The table presents first stage coefficients (robust standard errors in parentheses) from IV models using the IQ sample. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 5 and 6 shows the results for the cognitive outcomes in 9th grade using OLS and the RD approach. The OLS results suggest that track placement affects the 9th grade track position of students, while for IQ the coefficient is significant only at the 10% level. Using RDD, we also find no effect of track placement for math, but we do find that track placements affects reading, and similarly to OLS, although a much stronger effect, the track in 9th grade. The significance level for the effect of IQ is just above 5 percent ($p=0.053$). Table 7 and 8 presents results for the effects of track placement on non-cognitive skills. Using OLS we find that track placement does not lead to any differences in any of our non-cognitive outcomes for the marginal students, except for school motivation. This suggests that the placement procedures of schools are able to correctly place students into tracks. If we remove the endogeneity, we find an effect for the self-perceived probability of obtaining the degree.⁴⁹ The reason that the OLS and IV estimates differ is because the IV estimates capture the local average treatment effect for those students affected by our instrument while the OLS estimates depicts average differences, both controlling for the forcing variables (Lee and Lemieux, 2010). The downward bias of the OLS estimates could be explained by the role the forcing variables take: In the IV models we see that removing the endogeneity of being in the higher track shifts part of the effect of the forcing variables (as seen in the

49. The non-cognitive skills for which no effect with OLS or RD is found are extraversion, conscientiousness, openness, neurotism, agreeableness, competitive spirit, perseverance, social skills, school motivation, and school satisfaction.

OLS models) to the dummy for being in the higher track. Part of the being in the high track effect is in the OLS model wrongly attributed to the fact that on average these students also perform better in general. We correct for this using IV.

Table 5: OLS: The effects of being in the higher track on cognitive outcomes in grade 9

Dep var:	IQ	Math	Reading	Track grade 9
High track	0.20* (0.10)	0.03 (0.09)	0.10 (0.09)	0.25** (0.09)
Test score	0.04*** (0.01)	0.03*** (0.01)	0.04*** (0.01)	0.04*** (0.01)
Recommendation	0.05* (0.02)	0.07*** (0.02)	0.04** (0.02)	0.14*** (0.02)
Constant	-21.92*** (5.04)	-17.99*** (2.50)	-24.23*** (2.94)	-16.26*** (3.39)
# of students	1,175	1,862	1,927	2,866
# of schools	17	22	22	22
R ²	0.13	0.11	0.13	0.43

Notes: The table presents coefficients (robust standard errors in parentheses) from OLS models depicting the relation between cognitive outcomes and being in the higher track. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 6: IV: The effects of being in the higher track on cognitive outcomes in grade 9

Dep var:	IQ	Math	Reading	Track grade 9
High track	0.44* (0.23)	0.31 (0.23)	0.86** (0.35)	0.48*** (0.16)
Test score	0.03*** (0.01)	0.02*** (0.01)	0.02** (0.01)	0.03*** (0.01)
Recommendation	0.02 (0.03)	0.05* (0.03)	-0.02 (0.03)	0.13*** (0.03)
Constant	-17.79*** (5.48)	-13.78*** (2.83)	-12.05** (5.18)	-12.59*** (3.35)
# of students	1,175	1,862	1,927	2,866
# of schools	17	22	22	22
R ²	0.11	0.10	0.01	0.42
F excluded instruments	15.25	11.80	17.11	19.64

Notes: The table presents coefficients (robust standard errors in parentheses) from IV models depicting the effect of being in the higher track on cognitive outcomes. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The first stage results are presented for the IQ sample in Table 4.

Table 7: OLS: The effects of being in the higher track on non-cognitive outcomes in grade 9

Dep var:	Self-perceived probability for completing degree	Extraversion	Conscientiousness	School motivation
High track	-0.04 (0.07)	0.01 (0.01)	-0.01 (0.01)	0.02** (0.01)
Test score	0.01 (0.01)	-0.00 (0.00)	-0.00 (0.00)	0.00*** (0.00)
Recommendation	0.04 (0.03)	-0.00 (0.01)	0.01* (0.00)	0.00 (0.00)
Constant	-3.57 (4.28)	1.95** (0.89)	1.65 (1.02)	-0.76* (0.41)
# of students	994	1,026	1,028	2285
# of schools	16	16	16	22
R ²	0.02	0.00	0.00	0.04

Notes: The table presents coefficients (robust standard errors in parentheses) from OLS models depicting the relation between non-cognitive outcomes and being in the higher track. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 8: IV: The effects of being in the higher track on non-cognitive outcomes in grade 9

Dep var:	Self-perceived probability for completing degree	Extraversion	Conscientiousness	School motivation
High track	0.60** (0.25)	0.03 (0.03)	-0.02 (0.03)	0.03 (0.03)
Test score	-0.01 (0.01)	-0.00 (0.00)	-0.00 (0.00)	0.00** (0.00)
Recommendation	-0.03 (0.04)	-0.00 (0.01)	0.01** (0.00)	0.00 (0.00)
Constant	7.46 (5.32)	2.35* (1.23)	1.53 (1.16)	-0.53 (0.49)
# of students	994	1,026	1,028	2285
# of schools	16	16	16	22
R ²	-	0.00	0.00	0.11
F excluded instruments	11.42	10.64	10.89	12.42

Notes: The table presents coefficients (robust standard errors in parentheses) from IV models depicting the effect of being in the higher track on non-cognitive outcomes. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The first stage results are presented for the IQ sample in Table 4.

For part of our independent variables the same or similar variables are available for 6th grade, so we can include these as controls. These results are presented in Table 9 for IQ and the self-perceived probability of obtaining the degree. The other variables showed insignificant results. Unfortunately we do not have test scores for either math

or reading in 6th grade. Like before, the OLS results suggest an effect on IQ but not on the self-perceived probability to obtaining the degree. Table 9 shows that for IQ and for the self-perceived probability of obtaining the degree there is an effect of track placement. For IQ this effect is quite large and about twice as large as the OLS result: Being placed in the higher track leads to an increase in the IQ score of half a standard deviation. So only because a child is placed in the higher track in 7th grade, this child has a higher IQ score in 9th grade and think more positively of its chances of obtaining a secondary school diploma.

Table 9: The effects of being in the higher track on outcomes in grade 9, controlled for characteristics in grade 6.

Dep var:	IQ		Self-perceived probability for completing degree	
	OLS	IV	OLS	IV
High track	0.19** (0.09)	0.47** (0.22)	-0.04 (0.07)	0.57** (0.23)
Test score	0.02** (0.01)	0.02* (0.01)	0.01 (0.01)	-0.01 (0.01)
Recommendation	0.04* (0.02)	0.02 (0.02)	0.04 (0.03)	-0.02 (0.04)
Grade 6 variable	0.25*** (0.03)	0.25*** (0.03)	0.14*** (0.04)	0.15*** (0.04)
Constant	-13.46** (4.68)	-8.94* (4.93)	-3.80 (4.22)	6.55 (5.20)
# of students	1,175	1,175	994	994
# of schools	0.18	17	0.03	16
R ²	17	0.17	16	-
F excl. instr.	-	15.17	-	10.90

Notes: The table presents coefficients (robust standard errors in parentheses) from OLS and panel IV models depicting the effect of being in the higher track on IQ and the self-perceived probability for completing degree. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Heterogeneity in the results exists for gender and is shown in Table 10.⁵⁰ The effect of track placement on IQ seems to come entirely from the boys, since the girls' IQ scores are not affected by track placement. The reading score of both boys and girls is affected by being in the higher track, but for boys this effect is only significant at the 10 percent level.

50. No real heterogeneity exists for specific age groups. Young children seem to be more extraverted if they are placed into the highest track, but the F statistic is insufficient (5.8), perhaps due to the small sample size (N=498). The results of these split sample analyses are available on request.

Table 10: Heterogeneity effects of being in the higher track: boys vs girls

Dependent variable: Sample:	IQ Boys	IQ Girls	Reading Boys	Reading Girls
High track	0.92*** (0.28)	0.14 (0.29)	0.80* (0.46)	0.82*** (0.31)
Test score	0.01 (0.01)	0.05*** (0.01)	0.02 (0.02)	0.03*** (0.01)
Recommendation	0.02 (0.04)	0.02 (0.03)	-0.01 (0.04)	-0.02 (0.03)
Constant	-7.65 (5.30)	-25.79*** (6.36)	-10.04 (8.46)	-15.72*** (4.55)
# of students	552	623	892	1,004
# of schools	17	17	22	22
R ²	0.08	0.12	0.01	0.08
F excl. instr.	13.52	6.206	16.69	11.54

Notes: The table presents coefficients (robust standard errors in parentheses) from IV models separately for boys and girls. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

As said before, we have also used different model specifications as presented in Appendix L, where we used not both the elementary school exit test score and the elementary school teacher recommendation but either on. And we also included an interaction between the two indicator functions in the first stage or added an interaction between the two forcing variables in the second stage. The results are quite robust to these model specifications, although the F statistics varies over the models.

Since we have not restricted our sample using a bandwidth around the threshold to go to the higher track, and to provide an extra check to ensure students characteristic difference around the cutoff do not influence the results, we also estimated our models using controls for gender and parental background (Imbens and Lemieux, 2007). These results, which are very similar to those presented in the main text, can be found in Appendix M.

4.4 Conclusion

In this chapter we use a fuzzy RD design together with a panel dimension to separate the selection effects from the track placement effects for students in the two higher tracks in the Netherlands. We look at differences in student characteristics in 9th grade using the two main sources of information Dutch secondary schools receive from the elementary school to decide on track placement for students in 7th grade: the score of a uniform elementary school exit test and the elementary school teacher

recommendation. We find that most of the significant differences between students in the upper two tracks are due to selection and not due to track placement: The track placement procedures of schools in the Netherlands are able to correctly select students into tracks for almost all non-cognitive outcomes, but less so for cognitive outcomes. We find that IQ and reading test scores are enhanced simply by being in the highest track, and that track placement also increases the likelihood of being in the higher track in 9th grade. Maths scores, on the other hand, are not affected by track placement. Furthermore, the only non-cognitive skill that is affected by track placement is the self-perceived probability of obtaining the secondary school degree: Students in the high track are more positive about their chances of graduating. Other non-cognitive skills as personality or social skills are not affected.

Although some parents strive for the highest track for their child due to the positive learning outcomes, other parents are hesitant to do so since they believe it might hamper their child's non cognitive development. We show that the non-cognitive development of the marginal student who goes to the high track is not affected by track placement, but that the child's cognitive development benefits from the higher track placement. Given that we find no negative effects for the marginal student to go to the high track, irrespective of the large number of insignificant results, it might be better if schools were more accommodating to accept the marginal student to the higher track.

The results presented in this chapter do not imply that all students are better off in the high track. For the students who are not on the margin it might be beneficial for both cognitive and non-cognitive outcomes to remain in the low or middle track. Unfortunately, our set up does not allow us to investigate this.

The Pre-Tracking Effects of Parental Background⁵¹

Abstract. Tracking students in secondary school could increase the effect of parental background (PB) on student performance, especially if parents can influence the track choice. This influence can be either direct or indirect, and either purposefully or not. Little is known about these indirect effects of PB that could arise before tracking has taken place. In the Netherlands the track placement decision of individual students is made by secondary schools that base their decision on two performance signals that they receive from the elementary school of applying students: an elementary school exit test score and an elementary school teacher track recommendation. Using longitudinal data from the Netherlands, I find that high PB parents are able to increase their child's teacher recommendation (purposefully or not): The odds of having the highest track recommendation as compared to the other recommendations, for students whose parents have a tertiary education degree are 3.6 times greater than for students whose parents only have a primary education degree. For the math exit test score I find no effect, while for reading an effect is found but not robust.

51. This chapter is based on Korthals (2015), and an extended version of this chapter is Dronkers and Korthals (under review). I would like to thank Nicolás Salamanaca, Trudie Schils, Louise Elffers, the participants of the Economics of Education group, and the DUHR-AiO seminar in Maastricht University for valuable comments.

5.1 Introduction

To go to the high track in secondary school is often perceived as something positive for all students: The status of the track is higher, teachers and the curriculum in the high track are better, there are more post-secondary schooling options after successful completion of the high track, and later outcomes (educational attainment, wages, unemployment chances) might also be positively influenced by attendance of the high track. It is therefore no surprise that most parents want their child to attend the high track. In the Netherlands the track placement decision of individual students is made by the secondary school based on two performance measures: an elementary school exit test score and an elementary school teacher track recommendation. Parents therefore have no formal route to influence the track decision. However, parents will still have an influence on track placement either direct or indirect, and either purposefully or not. Parents that want to exert influence on the track decision might search for new areas in which they can exert influence when they do not have direct influence. When a high elementary school exit test score gives students access to the high track, parents will exert effort to ensure a good test score for their children, either by studying with their child for the test or by providing financial means to help the child study. And when the recommendation of the elementary school teacher is important, parents might try to persuade the teacher to give the highest recommendation. It might also be that parents have an influence because their characteristics (unconsciously) influence teachers who provide different recommendation for students with the same performance. Or specific groups of parents can have an influence by understanding the importance of the elementary school exit test score and the teacher recommendation, and consequently providing more support for their children.

The aim of this chapter is to examine whether a parental background (PB) effect exists on the factors that determine track placement, the two track determinants, above and beyond a PB effect on ability in general. A PB effect on ability is a common result, but whether at the moment of track placement an additional parental effect exists above this general effect has so far not been investigated. I make use of data from elementary school students in the Netherlands from before tracking has taken place. The Netherlands has a strict formal tracking regime in which students are tracked when they enter secondary school at the age of 12 (grade 7). The receiving secondary schools decide on the track placement of students and they base their decision on an elementary school exit test score and the track recommendation of the elementary school teacher, which elementary schools are required to deliver them by law

(Kingdom of the Netherlands, 1981).⁵² Since schools are free to accept students based on their own criteria, parents have no formal route to decide on the track choice of their child(ren). In practice, parents can help their child study for the elementary school exit test, exert pressure on the elementary school teacher to change the track recommendation, and due to free school choice, they can try to find a school which will accept their child to the desired track.

To be able to disentangle the general PB effect on ability and the PB effect on the track determinants, I control for previous test scores which are assumed to capture the general PB effect. Since these previous test scores might measure ability with error, an instrumental variable approach is used to limit the amount of measurement error in these previous tests. I then look at whether there is an additional relation between PB and the elementary school exit test score and between PB and the elementary school teacher recommendation besides the common relation of PB and ability. I find that an additional PB effect does exist: For instance, the odds for students whose parents have a tertiary education degree to get the highest teacher recommendation is around 3.6 times greater than for students whose parents only have a primary education degree. For the reading exit test score, high PB parents are able to increase their child's reading elementary school exit test score with one sixth of a standard deviation, but this effect is less robust. I find no effect on the math exit test score. To my knowledge this is the first study that explicitly looks at the additional PB effects that might arise in a tracked education system before the track decision has taken place, while taking into account measurement error on early ability.

PB, as proxied by parental education, parental income or parental employment status, and its effect on student performance is a widely studied phenomenon. The early contributions on this field are from Girard and Bastide (1963) and Boudon (1974), sociologists who distinguished between primary and secondary PB effects.^{53,54} This chapter uses the term *general PB effect* for what is in the sociological literature called the primary PB effect, and *additional PB effect* for the secondary PB effect on track placement. Primary PB effects are the direct effects of PB on ability or performance due to nature and nurture. Some of the mechanisms are that children from higher PB parents receive better nutrition and health care allowing them to reach their potential,

52. In the academic year 2014/2015 the use of the elementary school exit test in deciding the track choice is severely limited by the government and secondary schools are only allowed to use the elementary school teacher track recommendation in their track placement decision.

53. Jackson (2013b) provides an overview of the primary and secondary PB effects.

54. Recently Esser (under review) introduced the term tertiary PB effects for the differential effect that PB has on outsiders which influence the transition of the child. For instance, teachers might give higher track recommendations to children of high PB given their performance than to lower PB children. Teachers might give a high PB child the benefit of the doubt since the teacher thinks that (s)he will be better supported by her/his parents or unconsciously have different expectations on the children's ability.

these children are also read to more often which stimulates their language use and they receive a set of norms and values needed for successful educational careers. Secondary PB effects are the indirect effects of PB on performance based on different choices at important transition points. That these PB effects are not purely based on nature is shown by Black *et al.* (2005) who use exogenous changes in parental education and show that the education of the mother has a positive effect on the child's educational attainment. By controlling for the general PB effect (i.e. primary PB effect) with an early test score, this chapter only looks at the additional PB effect (i.e. secondary effect) on the two track determinants.

Tracking has often been found to increase the effect of PB on student performance (Ammermueller, 2005; Schuetz et al, 2008).⁵⁵ Part of this increase in the effect of PB could be explained by a PB effect on the tracking decision. For a number of countries, studies show this positive effect of PB on track choice (See Tieben *et al.* (2009) for the Netherlands, Dustmann (2004) and Schneider and Tieben (2011) for Germany, Sullivan *et al.* (2011) for United Kingdom, Ichou and Vallet (2011) for France, Horn (2013) for Hungary, and Panichella and Triventi (2014) for Italy). Korthals (2012) shows in a cross-country comparison that when principals always consider prior performance in accepting the student to the school in countries that track, the effect of PB on performance is lower. There it is argued that tracking is good for equal opportunities if track placement is done based on prior performance, since this lowers the influence parents have on their child's educational path. Dollman (2011) shows that the influence of PB on the track choice is indeed reduced when mandatory teacher recommendations play a role in track placement decisions. The effect of PB on the transition into a tracked system also runs through other mechanisms, for instance through the risk attitudes of the parents (Woelfel and Heineck, 2012) or on kindergarten attendance (Landvoigt *et al.*, 2007).

However, since a direct influence of parents on the track choice is very limited in the Netherlands, parents are left with using indirect influence. One way of indirect influence parents have is to help prepare their child for the elementary school exit test. Since the exit test score is one of two signals that secondary school use to base the track decision on, a higher elementary school exit test score will increase the chances of going to the higher track. To increase the test score of their child, parents can help their child study, pay for private tutoring and use other ways to help them perform better at the test. Another way to influence the track decision is to influence the elementary school teacher recommendation by convincing the teacher that a higher

55. However some studies find different results. Waldinger (2006) finds no effect of early tracking on the effect of PB once the PB effect in elementary school is taken into account. And also Brunello and Checchi (2007) do not find evidence for a reinforcing effect of tracking and they find tracking may even lower the effect of PB on performance after secondary school.

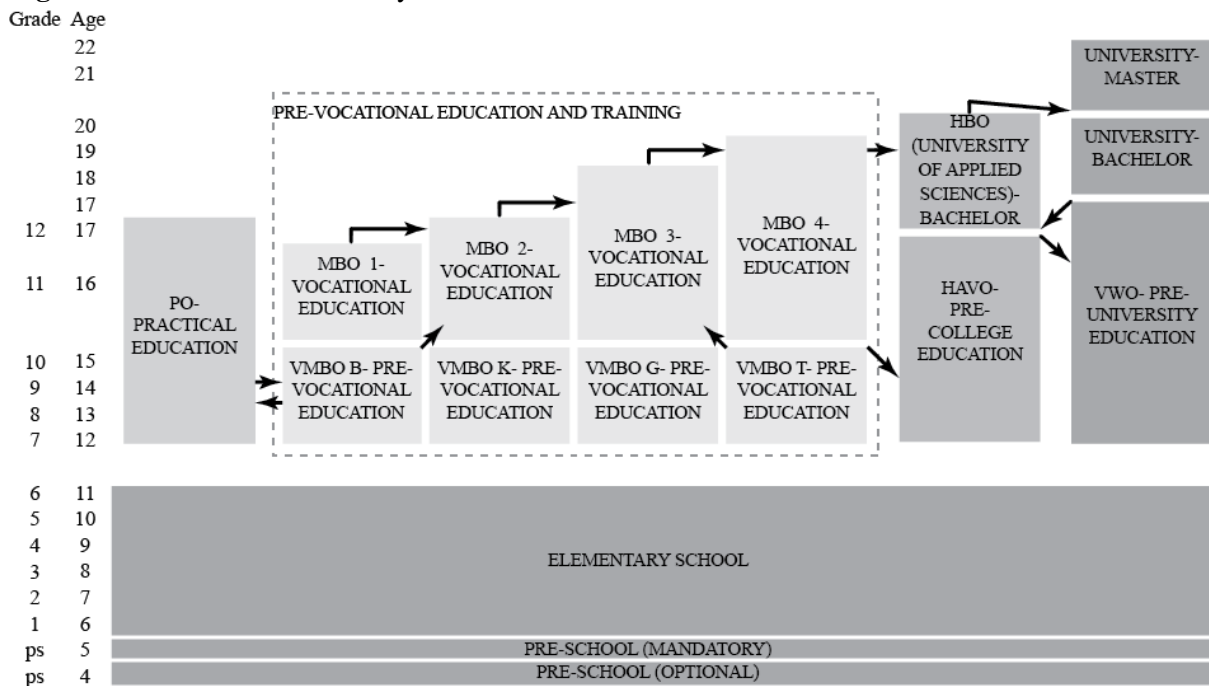
track choice is more suited for the child. When both these strategies did not lead to high enough signals for the secondary school to accept the child to the higher track, parents can try to convince the secondary school otherwise. Barg (2013) shows that when parents in France object to the upper secondary track recommendation of their child given by the teacher the school staff often reconsiders.

The structure of this chapter is as follows. First, I present some information on the Dutch education system. Then I elaborate on the data and the methodology. Subsequently, in Sections 5.4 and 5.5, the results are presented. Finally, I conclude.

5.2 The Dutch education system

The Dutch secondary education system is a highly differentiated system with early selection into tracks at age 12 with four main tracks available to students. Figure 1 shows the complete education system of the Netherlands. The move of students into the tracked system coincides with the transition from elementary to secondary school which is at age 12. The four tracks are practical education (PO), pre-vocational track (VMBO), and two general tracks: the track which directly leads to university of applied sciences (pre-college track, HAVO) and the track that directly leads to university (pre-university track, VWO).⁵⁶ The pre-vocational track is further subdivided into 4 tracks: VMBO b, VMBO k, VMBO g, VMBO t. VMBO b is the most practical track, while VMBO k, VMBO g, VMBO t have an increasing theoretical focus. When students first enter secondary school it is possible to enter “bridge classes” in which students from (most often) two adjoining tracks are grouped together, for instance HAVO-VWO classes or VMBO t-HAVO classes. Table 1 shows the distribution of students across track in all grades and it shows that in the seventh grade (first grade in secondary school) 80 percent of students are in such a bridge class, but most bridge classes only exist for one or two years. In grade 9 only 2 percent are still in a bridge class as can be seen in Table 1. Table 1 also shows that most students are in VMBO, HAVO and VWO. The practical education (PO) track contains less than 1 percent of students across all grades.

56. In VWO a further sub division exists for *atheneum* and *gymnasium*. Both sub divisions offer the same type of education except that at a gymnasium school students learn Greek and/or Latin.

Figure 1: The Dutch education system

Notes: Compulsory education is till the age of 16 if a start qualification is obtained. A start qualification is a degree in VWO, HAVO, or MBO level 2 and up. If no start qualification is obtained, education is compulsory till the age of 18. Special needs education is excluded.

Table 1: Student distribution across tracks in grade 7-9 (average from 2003-2010)

Track	Grade 7	Grade 8	Grade 9
PO	0.27	0.68	0.87
VMBO b	5.69	6.58	19.06
VMBO k	3.15	4.44	19.81
VMBO g	0.40	0.58	9.61
VMBO t	3.65	10.01	22.45
Bridge class	80.38	60.24	2.20
HAVO	1.29	7.84	13.52
VWO	5.16	9.64	12.48
Total	100.00	100.00	100.00

Source: Own calculations using DUO data.

The decision on which tracks student go to is solely made by the secondary schools. Secondary schools receive two signals on which they base their track placement decisions: an elementary school exit test score and a track recommendation of the elementary school teacher of the students. Parents can (try to) influence the track decision, but do not have direct influence on the track choice for their children. Parents who want to influence the track choice of their children can do so by helping their child perform well on the elementary school exit test, by exerting pressure on the elementary school teacher to give a high track recommendation or by exerting pressure on the secondary school. Since school choice is free, parents can also approach

multiple secondary schools for their child until one of these schools accepts their child in the desired track.

5.3 Data and methodology

To look at the effect of PB that takes place before tracking, I make use of a longitudinal data set representative for the Netherlands called COOL⁵⁻¹⁸. COOL⁵⁻¹⁸ started in the academic year 2007-2008 and every three years it collects administrative data from a representative sample of schools, including test scores from centralized tests in each grade, and survey data from the corresponding students, their parents and the schools. My analyses focus primarily on grade 6 and look at the additional effect of PB on the two performance signals secondary school use to decide on track placement, over and above the effect of PB on ability in general. To precisely isolate the additional effect of PB, the general PB effect must be completely controlled for. In the analyses this is done by including previous test scores which captures the child's ability, coming both from nature and nurture of parents. Since tests measure ability with error and thus will also control for the effect of PB with error, a second test score is necessary to be able to use one test score as an instrument for the other. Therefore in this chapter I use two waves of the data: Students are in grade 3 in 2007-2008 and in grade 6 in 2010-2011. I then use the grade 3 test score as an instrument for the grade 6 test score.

The data consists of 649 schools in 2007-2008 with a total of 15,473 students in grade 3 and 552 schools participated in 2010-2011 with 12,538 students in grade 6. Unfortunately in total only 62 percent of schools from the first wave participated in the second wave (Driessen *et al.*, 2012) which leads to a much smaller sample for the longitudinal data set. I exclude 26 students for whom the answers in 2007-2008 do not seem to match with the answers in 2010-2011 on gender or birth year. Furthermore for my analyses it is necessary to have enough test score data and to know parental education. Therefore I exclude another 3,811 students for whom I miss these vital data. My final sample consists of 2,621 students in 156 schools. Since the final sample deviates to some extent from the representative sample, I estimate the models using weights constructed to ensure representativeness for this smaller sample. However, no large qualitative differences arise when not using the weights. In Appendix N differences between the full grade 6 COOL⁵⁻¹⁸ sample and the sample used in this chapter are discussed and Appendix O reports the results without weights.

As a measure of PB of the children the highest obtained educational degree of both the parents is used. The highest obtained educational degree used in this chapter is provided by the school and is most often taken from the school administration

(Driessen *et al.*, 2012). The reason a school reported measure is used, instead of a parents reported measure, is that the number of missing observations on the school reported data is much smaller. The two measures have a correlation of 0.69. For those children for whom parental education was not available in grade 6, grade 3 data was used. Table 2 shows the descriptive statistics. There are four options in parental education: 8.0 percent of parents have at most a primary education degree, 20.0 percent at most a lower vocational education degree (similar to VMBO, pre-vocational education), 41.2 percent have at most a vocational education degree (MBO), and 30.9 percent of parents have a tertiary education degree (a degree from a university or a university of applied sciences).

In this chapter the aim is to study whether there is an additional PB effect on the two track decision signals, over and above the general PB effect on ability. To be able to do this I control for this general PB effect since otherwise the PB variable will capture the full PB effect. The PB effect on ability, which has arisen during childhood, is a common occurrence in all children around the world. Since ability itself is not observed, I will use a test score to approximate it. During the school year the students take multiple tests to allow their teachers to follow their progress, also in relation to children at other schools. All schools therefore have the same set of tests available to them provided by a test company in the Netherlands (CITO). Although schools have access to the same set of tests, they have discretion in the timing of testing and they grade the tests themselves. However, not all schools take the same set of tests from the available tests. To avoid too many missing observations on the test variables, I combine two tests on reading and two on mathematics.⁵⁷ For both subjects, the two tests are comparable in content, but the score distribution is not the same (Driessen *et al.*, 2012). Therefore I standardize both tests and use either one of the tests. I do this for both reading and mathematics. Table 2 shows that the final (composite) test scores used in this chapter are still fully standardized. And also the distributions of the different tests approach a normal distribution (not shown, but available on request).

The two outcome variables, the elementary school exit test score and the elementary school teacher recommendation, are as follows. In February of each year (most) students in grade 6 take an elementary school exit test. This test score is one of only two signals that secondary schools receive to decide on track placement. The test consists of a part on mathematics, reading and world views. The scores on the test range from 500 to 550, while the sub section on math ranges from 0 to 60 and reading from 0 to 100.⁵⁸ The elementary school exit test and the test in grade 6 are taken at

57. Including a dummy for which test was used does not change the results.

58. The distribution of the elementary school exit test scores for both reading and mathematics are slightly skewed to the right. However, the reading exit test score is truncated for only 1 student and the math

approximately the same time with only two months difference.⁵⁹ The elementary school teacher track recommendation is a measure with 15 recommendation options that teachers used. There are so many recommendation options since all forms of bridge classes are possible. For the sake of simplicity I collapsed this set of recommendation to eight: PO & VMBO b, VMBO k, VMBO g, VMBO t, VMBO t & HAVO, HAVO, HAVO & VWO, and VWO. These eight groups are of fairly equal size: 12.7 percent of students received a PO & VMBO b recommendation, 12.8 percent a VMBO k recommendation, 7.5 percent a VMBO g recommendation, 18.1 percent a VMBO t recommendation, 8.8 percent a VMBO t & HAVO recommendation, 17.8 percent a HAVO recommendation, 9.0 percent a HAVO & VWO recommendation, and 13.3 percent a VWO recommendation. The elementary school teacher recommendation is given during the spring of grade 6 and theoretically contains all the information on the child’s ability level and progress until that point.

Table 2: Descriptive statistics

Variable	# of students	# of schools	Mean	Std. Dev.	Min	Max
Math exit test score	2,621	156	43.27	10.54	8.00	60.00
Reading exit test score	2,621	156	75.64	11.50	6.00	100.00
Math score grade 3 (Z)	2,621	156	0.10	0.95	-4.19	3.67
Math score grade 6	2,621	156	0.10	0.96	-4.23	3.58
Reading score grade 3 (Z)	2,466	153	0.11	1.01	-1.93	5.18
Reading score grade 6	2,551	152	0.13	0.99	-3.82	5.04
Recommendation	2,520	156	4.87	2.21	1.00	8.00
Parental education	2,621	156	3.20	0.79	1.00	4.00
Gender	2,606	156	1.51	0.50	1.00	2.00
# of students in grade 6	2,621	156	15.20	12.18	1.00	67.00

Notes: Weighted data, COOL⁵⁻¹⁸.

Two potential confounding factors play a role. First, since the grade 6 test and the elementary school exit test are both taken within a short period of time, part of the additional PB effect could run through the grade 6 test score. This means that in the analyses of this chapter only the additional PB effect of the last two months is analyzed. This can be seen as a lower bound of the additional PB effect. Second, an ability measure in the form of a prior test score will only capture the PB effect adequately if the test score measures ability without error. Since this is unlikely to be the case, I employ an instrumental variable technique to alleviate the measurement error in the grade 6 test score. The instrument I use for this is a test score obtained in grade 3. To look at the effect of PB I will estimate the following model:

elementary exit test score for 16 students. This level of truncation does therefore not necessitate a tobit analysis and thus OLS is used.

59. All tests were taken between January and March. The CITO test is in February.

$$TestGrade6 = \varphi_0 + \varphi_1 * TestGrade3 + \varphi_3 * PB + u \quad (1-I)$$

$$ExitTest = \beta_0 + \beta_1 * \widehat{TestGrade6} + \beta_2 * PB + \varepsilon \quad (1-II)$$

Model (1-I) depicts the first stage in which measurement error is removed from the test score in grade 6, while models (1-II) depict the second stage for the elementary exit test. Both models are analyzed for the reading and mathematics elementary exit test. Since the test score in grade 3 is used to alleviate the measurement error, β_1 will be an unbiased estimate of the effect of ability on the elementary exit test. β_2 provides a lower bound for the additional PB effect when the parents who help their child study for the elementary exit test in doing so also help the child perform on the grade 6 test. It only provides a lower bound since both the instrument and the instrumented variable are most likely a function of PB.

Equation (2) looks at the additional PB effect on the elementary school teacher recommendation and the main coefficient of interest is γ_3 . Due to the discrete and ordinal nature of the elementary school teacher track recommendation, for the models with this as outcome I use an ordered logit.

$$Recommendation = \gamma_0 + \gamma_1 * TestGrade6 + \gamma_2 * ExitTest + \gamma_3 * PB + \varepsilon \quad (2)$$

To also control for measurement error in this model, the analyses are also done using IV. Since IV requires a continuous dependent variable, I use the 15 options categorical track recommendation variable, rescale it using the ordered logit cut off points and use this variable for the IV. An underlying assumption of ordered logit is that the slope coefficients are the same across categories of the dependent variable, i.e. that the effects of the independent variables are the same across the different steps in the dependent variable. To test whether this is the case in the models estimated here, I first estimated generalized ordered logit models which relax the parallel lines assumption.⁶⁰ However, the parallel line assumption of the ordered logit is only violated for the elementary school exit test scores and since this is not the variable of interest ordered logit is used.

5.4 Results for the elementary school exit test score

The question asked in this chapter is whether there exists an additional PB effect on the track determinants (the elementary school exit test score and the elementary school teacher recommendation) over and above the general PB effect found on ability. The results for the PB effects on the elementary school exit test score for mathematics and

60. The command `gologit2` in Stata, with the `autofit` option, will only relax the parallel lines constraint for those variables where it is violated

reading are shown in Table 3. The first and fourth column of each panel shows the full PB effect on the math and reading elementary exit test scores. These coefficients contain both the expected PB effect on ability and the additional PB effect on the track determinants. Both in reading and in mathematics parents have a large influence on the test score, with students whose parents have a tertiary education degree scoring one standard deviation higher than students whose parents only have a primary education degree. In columns (2) and (5) the test score in grade 6 is included as a control. Any PB effect that has helped the child over the years to perform better in the test subjects should be included in this test score. As a consequence it can be seen that the effect of PB is lower in columns (2) and (5) than in columns (1) and (4). For mathematics the PB effect in column (2) is even insignificant, except for the highest PB category.

However, any remaining PB effect in mathematics and also in reading could be due to measurement error in the test score of grade 6. If this score is measured with error, it is possible that some remaining general PB effect will be captured by the PB dummies instead of by the grade 6 test score. To remove this measurement error, both columns (3) and (6) present results for models in which the grade 6 test score is instrumented by the test score in grade 3. The content of both tests is comparable and the F statistic (525.9 for mathematics and 382.0 for reading) also shows that the grade 3 is a strong instrument for the grade 6 test. Appendix P shows the first stage results.

Looking at the IV estimates in Table 3, it can first be seen that an additional PB effect on the elementary exit test score exists for reading. The effect of PB is lower when an instrument is used but for reading still a significant effect of PB remains, with students whose parents have a tertiary degree score one sixth of a standard deviation higher than students whose parents have only a primary education degree. The results show that parents have an influence on the reading section of the elementary exit test, but not on the mathematics section. This result could possibly be due to the fact that it is easier for parents to help their child with reading than with math since reading practice in the home is easily available, for instance by reading the newspaper together. However, since the results are most likely an underestimation, an additional PB effect of math could still exist although less prominent. Second, measurement error does seem to play a large role in columns (2) and (5) since the IV estimates of the effect of the grade 6 test score on the elementary exit test score are higher when the instrument is used. Interestingly, the additional effect of PB decreases much more for those students with a higher PB than for those with a lower PB. The naive estimates using ordinary least squares seem to indicate a much more favorable position for high PB students compared to lower PB students than they in reality have.

Table 3: Parental background effects on the elementary school exit test score

	Math exit test score			Reading exit test score		
	(1) OLS	(2) OLS	(3) IV	(4) OLS	(5) OLS	(6) IV
Parent: at most lower vocational education	3.12** (1.53)	0.91 (0.80)	0.44 (0.76)	3.53*** (1.12)	2.09*** (0.75)	1.43* (0.74)
Parent: at most vocational education	5.18*** (1.40)	0.46 (0.75)	-0.53 (0.77)	7.55*** (1.14)	3.48*** (0.92)	1.63* (0.99)
Parent: at most tertiary education	10.42*** (1.47)	1.84** (0.86)	0.04 (0.91)	12.55*** (1.21)	5.32*** (0.94)	2.05** (1.00)
Grade 6 test score		8.89*** (0.22)	10.76*** (0.35)		7.63*** (0.35)	11.09*** (0.51)
Constant	36.45*** (1.37)	41.29*** (0.75)	42.31*** (0.79)	67.13*** (1.09)	70.88*** (0.80)	72.58*** (0.82)
# of students	2,621	2,621	2,621	2,398	2,398	2,398
# of schools	156	156	156	149	149	149
R ²	0.09	0.68	0.66	0.10	0.51	0.43
F excl. instr.			525.9			382.0

Notes: The table presents coefficients from OLS and IV models (robust standard errors in parenthesis) on the relation between student performance on the primary exit test score and parental background and early ability. Estimates in columns (3) and (6) are from a two stage least square model with as excluded instrument for the test score in grade 6 the test score in grade 3. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The standard errors are clustered on the grade 6 schools, and student weights are used.

That the additional PB effect in columns (3) and (6) is lower than in column (1) and (4) does not mean there is no or less of a PB effect. Due to the inclusion of the grade 6 test score, all direct PB effects are captured in this variable. PB therefore most surely has an effect on the elementary exit test. What columns (3) and (6) show is that for mathematics there seems to be no additional PB effect on the elementary school exit test, while there is an additional effect on the reading exit test. The additional PB effect on reading is mainly driven by the girls, while there is no PB effect for the boys (see Appendix Q for the results).

There are three potential issues that could undermine the results on the reading test score. First, using school fixed effects leads to less robust results for reading. School fixed effects could be important to check whether school specific factors matter for the additional PB effect. One example of a confounding school factor could be specific grading practices since the grading of the tests, except for the exit test, are done by the teachers themselves. However, by including school fixed effects, part of the effect of PB is removed since the parents choose the elementary school for their child.

Including school fixed effects might therefore lead to a further underestimation of the effect of PB on the elementary exit test score. When the models of columns (1), (2), (4) and (5) of Table 3 are estimated with school fixed effects the results are robust, although smaller in size. But when instrumental variables are used, similar to columns (3) and (6)), there is no longer an additional PB effect on reading. This indicates that it might be school differences which drive the PB effects. Both coefficients drop and the standard errors increase, the last being an indication that the fixed effects models are very demanding on the data. Two reasons why these fixed effects models might show an underestimation are, first, that the amount of variation in parental education is lower within schools than across schools and, second, parents choose the elementary school for their children which means that including school fixed effects lowers the effect of PB. On the other hand, if the school fixed effects lower the additional PB effect it could also mean that the grade 6 test score does not fully capture the general PB effect, which is then picked up by the PB dummies in Table 3, but no longer when school fixed effects are included. This suggests that the aim of this paper, to disentangle the general from the additional PB effect, is hard to accomplish. All in all, the school fixed effect models render less confidence in the results on the reading test scores.

The second potential issue for the results on the reading test scores could be the sample selection procedure. To see whether the sample selection procedure influenced the results, I also use the full sample (5,643 observations for reading and 5,858 for math) and impute the grade 3 test scores using mean imputation of the lowest stratum. The reading PB effects are still there in the OLS models, but disappear when using IV. There is now also a negative PB effect for the 2nd and 3rd PB group on math. The negative math results can be explained by divergent schools which were added to the sample in the second wave, for the reading results this does not seem to be the case.⁶¹ Similarly to the school fixed effect models, these robustness checks reduce the confidence in the positive PB effects on the reading test score, allowing for the alternative conclusion that there is no additional PB effect on the elementary school exit test.

Third, the content of the reading test in grade 6 is different from the content in the elementary school exit test. A large portion of the test coincides, but the exit test contains spelling and grammar while test grade 6 does not. This leaves room for an alternative conclusion that the remaining PB effect I find on the reading elementary school exit test score is due to a general PB effect on the spelling and grammar ability of students.

61. This statement is based on the significance of imputation dummies and imputation dummy and imputed variable interactions.

5.5 Results for the track recommendation

Table 4 shows the results looking at the elementary school teacher recommendation. Also here column (1) gives the estimates of the full PB effect, displayed in odds ratios. The elementary school teacher recommendation is a categorical variable with eight options: PO & VMBO b, VMBO k, VMBO g, VMBO t, VMBO t & HAVO, HAVO, HAVO & VWO, and VWO. The estimates in column (1) of Table 4 state that for students whose parents have a tertiary education degree the odds of having the highest track recommendation as compared to the other recommendations are almost 13.5 times greater than for students whose parents only have a primary education degree. This large PB effect can also be seen in the blue solid lines of Figure 2, which shows the predicted probabilities to obtain the eight track recommendations given PB. Students with parents with at most a primary education degree have a predicted probability of 30 percent of obtaining the lowest track recommendation, while students with the highest PB have only have a predicted probability of around three percent. In column (2) of Table 4 also the grade 6 test scores are included to separate the general PB effect from the additional PB effect, and column (3) also includes the elementary school exit test scores on math and reading. Column (3) shows that for the second highest and highest PB group there remains an additional PB effect on the elementary school teacher recommendation after inclusion of test scores. Figure 2 also shows this decline in influence of PB separately for each track recommendations.⁶² The red dashed lines in Figure 2 shows that the predicted probabilities of obtaining one of the eight track recommendation when controlled for ability still differs across PB groups. The slope of the red dashed line is less steep than the blue solid line, except for the highest pre-vocational track and the lower general track, and the two higher PB groups are less likely to receive a recommendation for the lower tracks and more likely to receive a recommendation for the higher tracks as compared to the other groups. Another way to interpret these results is by looking at how much tertiary educated parents are able to bridge the gap between the lowest and the highest track, which is given in the last row of Table 4. In the model with no controls they are able to bridge 56 percent of the distance between the lowest and the highest track, while in column (3) they able to bridge 13 percent of the distance. The additional PB effect on the

62. The line indicating the model controlling for ability is almost always below the line where ability is not controlled for, illustrating the fact that including ability measures lowers the effect of parental background. For the VMBO-t, VMBO-t & HAVO, and HAVO track recommendation, this is however not the case. Karlson, Holm, and Breen (2012) show that when estimating same-sample nested logit models, it is not possible to compare the coefficients of the different models since the variance of the underlying latent (and thus estimated) variable differs between different models. To correct the estimates for this, the KHB method essentially uses the variance of the underlying variable as estimated in the largest model to estimate the coefficients for the lower level nested models resulting in larger effects of the coefficients of interest in the smaller model. This correction therefore increases the PB effect in the smaller models and makes the decrease due to the inclusion of the grade 6 and the elementary exit test scores even larger. However, the additional PB effect in the largest model still remains which is the reason the corrections were not applied to Table 4 or in Figure 2.

teacher recommendation is mostly due to the boys (see Appendix Q for the results). When the extended sample is used (5,401 observations) a PB effect exists for the second lowest and the highest PB category.⁶³

Table 4: Parental background effects on the elementary school teacher recommendation

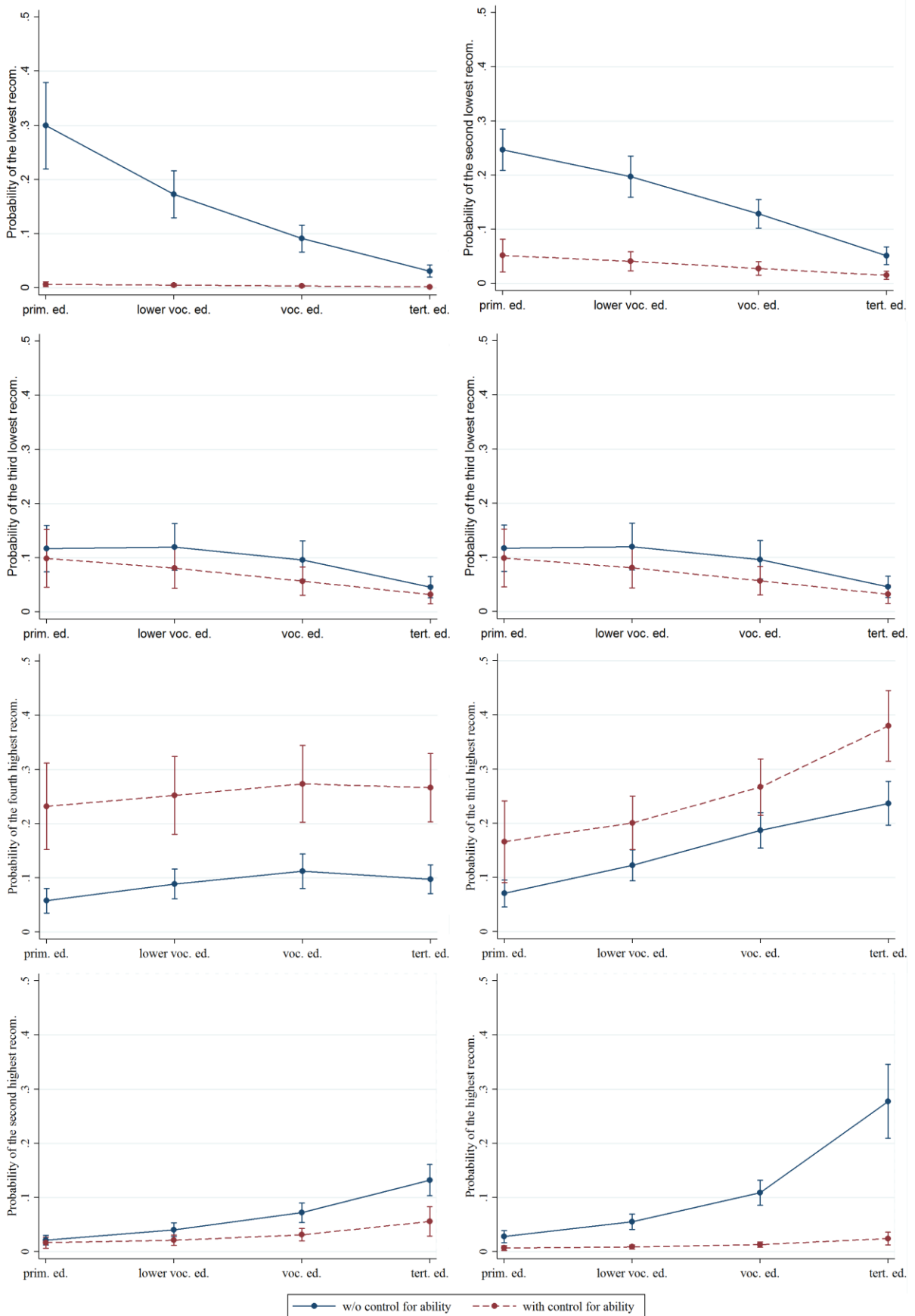
	(1) o. logit	(2) o. logit	(3) o. logit	(4) IV
Parent: at most lower vocational education	2.05*** (0.47)	1.54 (0.42)	1.28 (0.37)	1.13 (0.14)
Parent: at most vocational education	4.28*** (0.89)	2.30*** (0.65)	1.93** (0.57)	1.38** (0.16)
Parent: at most tertiary education	13.49*** (3.34)	5.16*** (1.65)	3.65*** (1.20)	1.62*** (0.24)
Test score grade 6: math		5.44*** (0.67)	2.65*** (0.43)	1.01 (0.01)
Test score grade 6: reading		4.45*** (0.40)	2.22*** (0.25)	1.02*** (0.01)
Elementary exit test score: math			1.09*** (0.01)	2.01*** (0.41)
Elementary exit test score: reading			1.14*** (0.01)	2.44*** (0.37)
# of students	2,452	2,452	2,452	2,307
# of schools	152	152	152	149
(Pseudo) R ²	0.04	0.30	0.38	0.69
Parent: HBO/WO/(cut 3- cut1)	0.56	0.20	0.13	-
F statistic math test score grade 6	-	-	-	50.66
F statistic reading test score grade 6	-	-	-	24.24

Notes: The table presents odds ratios from ordered logit models (first three columns) and adjusted coefficients from the second stage of an IV model (robust standard errors in parenthesis) on the relation between elementary school teacher recommendation and parental background and early ability. The IV model uses a continuous version of the teacher recommendation based upon the ordered logit cut points, see the text for more information, and the IV coefficient are, for comparison reasons, displayed as exp(b), similar like the ordered logit coefficients. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The standard errors are clustered on the grade 6 schools, student weights are used, and the constants are omitted from the table. The third from last row states: Being from the highest PB group bridges 13 percent (column 3) of the distance between the highest and the lowest track, which is calculated using ordered logit coefficients.

63. This also hold if the grade 6 test scores are not included, or when the overall elementary exit test score (i.e. not the reading and math parts separately) is included.

The measurement error in the grade 6 test scores cannot be removed by an IV model using the current variables, since the elementary school teacher recommendation is a categorical variable. An instrumental variable method cannot be performed using a categorical dependent variable due to its non-linearity. For illustration purposes, Column 4 of Table 4 still presents results from an IV estimation using the full categorical dependent variables with all 15 options. Appendix P shows the first stage results related to column 4 of Table 4. Since the steps in the categorical variable might not be equal for each discrete jump in the categorical variable I rescale the elementary school teacher recommendation using the cut points from an ordered logit model looking at the relation between the elementary teacher recommendation and parental education. The cut points are used to transform the track recommendation in the following way: The value for each recommendation option is the average of the two consecutive cut points. The value for the lowest recommendation option is the second lowest value minus the difference between the first and second cut point, while the value for the highest cut point is the second highest recommendation plus the difference between the penultimate and the last cut point. Column 4 of Table 4 presents the estimates for this model. Since the coefficients in column (4) are estimated with IV, the coefficients are not readily comparable to the odds ratios in the other columns. However, similar to the ordered logit models, for the highest PB category a significant positive effect of PB exists. Using IV also the second highest educated parents are able to positively influence their child's track recommendation. As said before, these results are only for illustrative purposes since IV requires a continuous dependent variable.

Figure 2: Predictions for parental background on the elementary school teacher recommendation



5.6 Conclusion

In this chapter I look at whether there exists a parental background (PB) effect on the two track determinants in the Netherlands over and above a PB effect on ability. In the tracked education system of the Netherlands, secondary schools made the track placement decision based on two performance signals: an elementary school exit test score and an elementary school teacher recommendation. Parents have no formal route to exert influence on the track decision. Naturally there always is a PB effect on performance since parents directly influence the ability of their children.

To investigate whether there is a PB effect on the two track determinants over and above a PB effect on ability, I control for observed ability in the form of a test score in grade 6. Since this test score measures ability with error, I instrument this with an earlier test score. I find that parents are able to increase their child's reading elementary school exit test score and the higher educated parents are, the more effect they have. I find no such effect on the math exit test score. The existence of a PB effect on the reading test score and not on the math test score can be explained by the easier transmission of reading skills in daily life than math skills. The results on the elementary exit test are however not fully convincing: the significant effects disappear when I use school fixed effects or an extended sample. On the other hand, the results on the elementary exit test score provide a lower bound for the full additional PB effect due to the quick succession of the grade 6 test and the exit test, which leaves room for a larger effect.

On the elementary school teacher track recommendation I find a robust additional PB effect. The odds of having the highest track recommendation as compared to the other recommendations, for students whose parents have a tertiary education degree are 3.6 times greater than for students whose parents only have a primary education degree.

Summarizing, parents seem to have an influence on the elementary school exit test score and the teacher track recommendation of their child, over and above a PB effect on the child's ability. Whether this is due to parents purposefully or unconsciously exerting influence or due to teachers providing different recommendations based on PB cannot be said based on these results. As far as I know this is the first study to look at the PB effect on track determinants, while controlling for measurement error.

Does Early Educational Tracking Increase Inequality? Short and Long Term International Evidence⁶⁴

Abstract: This study investigates the short and long term effects of early tracking comparing the performance of relatively young and relatively old students in early and late tracking countries. We exploit exogenous variation in the effect of relative age across the two groups of countries causing differences in the probability of selection into the low track. Using internationally comparable data, we observe that relatively younger students are more likely to be sent to low academic tracks in countries which select early. This is expected since i –relative age differences within a cohort matter for school success and ii –this difference naturally disappears as pupils becomes older. This higher probability to go to the lower track is often seen as a negative result since students in lower tracks will have worse outcomes. However, it could also be that these students benefit since in early tracking countries the composition of the low track will be more heterogeneous with respect to ability, and relatively younger students will quickly outperform their less able older peers in the low track when the relative age effect has dissipated. Test scores after tracking reveal that relatively younger students in early tracking countries do better than their peers in late tracking countries. This is confirmed when looking at long term outcomes where we exploit within country changes in tracking age.

64. This chapter is based on joint work with Olivier Marie and Dinand Webbink. We would like to thank Danielle Checci for generously offering the longitudinal data on age of tracking to us for this chapter.

6.1 Introduction

To facilitate learning when ability is heterogeneous, all education systems track students on ability at some point. However, the age of tracking differs strongly between countries. In some countries students are tracked quite early, often at the start of lower secondary education (e.g. age 10 in Germany and Austria) while in other countries students are tracked later, sometimes not until the start of upper secondary education (e.g. age 16 in Sweden and US). The age at which students are tracked might have important short and long term consequences. The long term consequences could be especially important if tracking in different educational levels works as presorting for adult socioeconomic outcomes, for instance when access to tertiary education is restricted for those students graduating from the lower tracks.

However, the short and long term effects of the age of first tracking are not well understood since investigating the effect of the age of tracking is difficult because most of the variation is between countries and not within. Previous studies looking into the effects of age of tracking on outcomes have therefore used between country variation using a control strategy, between country variation using difference-in-difference or instrumental variables. Some exceptions for which within country variation was available are Kerr *et al.* (2013), Van Elk *et al.* (2011), Duflo *et al.* (2011), and Dustman *et al.* (2014) and these studies come to conflicting results. The main challenge for studies that exploit between country variation in age of tracking is to take account of a multitude of factors that might differ between countries. For instance, Hanushek and Woessmann (2006) use a difference in difference approach in which they use test scores from before and after tracking has taken place. They find no effect on mean achievement of students, but a negative effect of early tracking on distributional inequality. However these result are not robust to small samples changes (Jakubowski, 2009) or when as outcome inequality of opportunity is used (Waldinger, 2006).

This chapter proposes a new approach to study the effect of tracking on short and long term outcomes in which we exploit an exogenous mechanism which creates a difference in the probability of enrollment in the low ability tracks between countries with early tracking and countries with late tracking. The mechanism we use arises from school entry rules which generate differences in the age at which students start school: Within a cohort the oldest students are 11 months older than the youngest students. It has been shown that age at start matters for performance: Relatively older students perform better (Angrist and Krueger, 1992; Bedard and Dhuey, 2006; Crawford *et al.*, 2013). And in a number of studies it has also been found that younger (older) students are more likely to enroll in the low (high) ability track (Grenet, 2010; Puhani and Weber, 2006; Jurges and Schneider, 2007; Muhlenweg and Puhani, 2010;

Dustmann *et al.*, 2014). Given that age differences at the start of a tracked school system are relatively larger when countries track earlier, because a difference of 11 months in age at start is relatively larger for 10-year-olds than for 15-year-olds, the effects described above are therefore likely to be even larger in countries that track at an earlier age. Subsequently we expect that in countries with early tracking relatively young students more often enroll in the low track than in countries with late tracking.

We investigate the differences in short-term and long-term performance between relatively young and relatively old students and relate this to the timing of tracking. Specifically, we investigate whether these differences between ‘young’ and ‘old’ students are larger in late tracking countries or in early tracking countries. We use data from the Program for International Student Assessment (PISA) for 23 OECD countries with different age of tracking. We group the countries into early and late tracking countries which allows us to employ country fixed effects. We then show that those who are relatively young are more likely to be enrolled in low level tracks but, perhaps unexpectedly, they perform better at age 15 than young students in countries that track late. An explanation might be found in the effect of early tracking on the composition of students in the low track. In early tracking countries the composition of the low track will be more heterogeneous with respect to ability, and the more able young students in the lower track might benefit from this heterogeneity at the expense of the less able old students causing a larger discrepancy in outcomes from students from the lower track. To look at the long run effect, we use data from the Program for Assessment of Adult International Competences (PIAAC) and focus on males between the age of 25 and 50. We use data from 14 countries in PIAAC that changed the age of tracking between 1939 and 2012 and therefore potentially switch between being early and late tracking countries. In this setup we are able to use within country variation in both relative age and age of tracking providing us with causal estimates. We find that also here the relatively young gain in a system that tracks early: they are more likely to be above average earners and less likely to be among the bottom ten percent earners. Hence, early tracking changes the relative age effects commonly found and has long lasting effects. Although relatively young students are more often placed in the low track, they benefit from this with better long run outcomes.

This chapter is structured as follows. Section 6.2 describes the related literature on both relative age effects and early tracking. Sections 6.3 and 6.4 present the empirical strategy and the data. Following that, section 6.5 shows the different results in three steps. First, we show the effect of relative age on track placement. Then we show that this changes the ability distribution in the two tracks. And in the last step, we show that also in the long run these effects persist. Section 6.6 concludes.

6.2 Tracking and relative age

This chapter brings together two strands of the literature: the literature on relative age effects and the literature on early tracking. These strands have already been combined by a number of papers by showing that the relatively young are more likely to be placed in the lower track. Here we present the existing evidence on these topics.

6.2.1 Relative age effects

Our approach is partly inspired by the month of birth (MOB) literature (see for a review of the early non-economic literature Stipek, 2002). The MOB literature stems from the fact that within classes, composed of children of the same age, the age difference between children can be up to 11 months. In the early years, this age difference of 11 months can be quite large relative to the age in months. The literature distinguishes between disadvantages in student performance due to the age of starting school (school readiness), the age of sitting the test, and length of schooling effects. In most papers, including this one, the methodology makes it difficult to distinguish between the different effects of relative age. Gaining early school experience is found to be more beneficial than delaying schooling, while the age of sitting the test has been found to increase performance (Stipek, 2000). The theory of human capital formation based on self-productivity and dynamic complementarities provides theoretical underpinning of these effects (Cunha et al, 2006).

Angrist and Krueger (1992) provide the first economic evidence by looking at the effect of MOB on attainment when it induced students to stay in school longer. Bedard and Dhuey (2006) and Crawford *et al.* (2013) look at the effect of relative age at different points in the school career. Bedard and Dhuey (2006) find effects of relative age on 4th and 8th graders across OECD countries and college applicants in the US and Canada. Relatively young students perform on average between 4 and 12 percentile points worse in 4th grade than their relatively older peers, while the performance difference declines to 2 and 9 percentiles in 8th grade. Crawford *et al.* (2013) look at the effect of relative age in England, which is a particular interesting case to look at due to the regular national testing in schools. Similarly to Bedard and Dhuey (2006), Crawford *et al.* (2013) show a decline in the effect of relative age, now between the ages 11, 14, 16, 18 and for participation in tertiary education at age 19. For these five different points in time, the authors show the difference in the probability of obtaining a certain academic threshold. The relatively young students are less likely to reach this threshold, more so in the early years than later. By making use of variation in the cutoff point of school starting ages within England, Crawford *et al.* (2013) are able to identify the effect they find as an effect due to relative age, and not due to the age children start schooling.

Some papers also look at long run effects of relative age. Black *et al.* (2011) find, using Norwegian data, a positive effect of age of sitting the test and a small negative effect of the age of starting school on IQ score at age 18. They find only small long term effects, with a small negative effect of relative age on male earnings until the age of 30. Fredriksson and Oeckert (2013) find no effect of the age of starting school on life time earnings, although they do find heterogeneous effects for parental background.

6.2.2 *Early tracking*

Most papers looking at the effect of tracking on inequality look at inequality caused by parental background (Pekkarinen *et al.*, 2009; Brunello and Checchi, 2007; Waldinger, 2006) or distributional inequality (Hanushek and Woessman, 2006; Jakubowski, 2009) and they have either used between country variation using a control strategy, within country variation, or between country variation.

An example of a study using within country variation is Pekkarinen, Uusitalo, and Kerr (2009). These authors use a Finnish education reform in the 1970s, which postponed the age of tracking from 10 to 15, as a natural experiment to study the effects of early tracking on intergenerational income mobility. They make use of the gradual implementation of the reform in which some municipalities implemented delayed tracking earlier than others. The reform reduced the intergenerational income elasticity in a sample of males.

Brunello and Checchi (2007) use between country variation but are able to alleviate the inevitable endogeneity in cross country research. The authors use age variation, dropouts and other factors influencing time in school as instrumental variables to instrument the time spend in a tracked system and find positive effects of the tracking length on performance between age 16 and mid-20s and that the effect of parental background increases if students are tracked for a longer period.

Hanushek and Woessmann (2006) use the fact that no country tracks in elementary school to apply a cross country difference-in-difference to look at early tracking. The main assumption is that early tracking and late tracking countries have a common trend in student performance: the change in performance in late tracking countries would have been seen in the early tracking countries if they had tracked late. Hanushek and Woessmann (2006) find no effect on mean achievement of students, but a negative effect of early tracking on distributional inequality. However this result is questioned by Jakubowski (2009) and Waldinger (2006). Jakubowski (2009) uses individual level data for the same countries and finds that with small samples changes the positive results of Hanushek and Woessman (2006) turn negative or disappear. Waldinger (2006) looks at the effect of parental background on student performance

using a similar difference in difference approach. Using individual level data, he finds no effect of an interaction between early tracking and parental background once elementary school inequality is taken into account.

6.2.3 *Relative age effects and tracking*

Following the MOB literature and its effects on educational outcomes, it has been shown that relatively young students are more likely to be enrolled in the lower track: Allen and Barnsley (1993) find this effect for Canadian Hockey teams, Puhani and Weber (2006), Jürges and Schneider (2007), Muehlenweg and Puhani (2010) and Dustmann *et al.* (2014) for secondary school track allocation in Germany, and Grenet (2010) for upper secondary school choice in France.

Fredriksson and Oeckert (2013) compare the long run relative age effects in Sweden in a time when Sweden still had an extensive tracking regime (till late fifties) with those in the period after and find that the move from early tracking to late tracking reduced the long run relative age effects on educational attainment. Contrary to the findings of Fredriksson and Oeckert (2013), Dustmann *et al.* (2014) find no MOB effect Germany on later outcomes which they attribute to track mobility options in the highly stratified German system. The authors look at tracking in Germany and find a strong MOB effect on track placement, but as said this effect does not persist. Also Muehlenweg and Puhani (2010) find no signs of a decline in the relative age effects in Germany if tracking is postponed, as is possible in some schools. Muehlenweg and Puhani (2010) do find that after the age of 16, when track mobility possibilities are larger, the relative age effect declines. Thus, although it has been suggested that the relative age effect would be smaller in education systems that track later, no consistent evidence supports this. As far as we know, we are the first to look directly at the relation between the timing of tracking and the relative age effects in a cross country setting. Using our set up we are able to obtain effects of early tracking on long term outcomes across a number of countries, by exploiting the exogenous variation in the probability to go to the low track caused by a difference in the size of the relative age effect across early and late tracking countries.

6.3 Empirical strategy

In our empirical strategy we exploit exogenous variation in the effect of relative age, which itself is caused by exogenous variation in MOB and school cut off dates, across two groups of countries: those that track students in secondary school at an early age and those that track students late. This difference in relative age effects across these two groups of countries leads to exogenous variation in the probability to go to the low track for the relatively young students. Our empirical strategy consists of two parts.

First, we focus on short term differences in relative age effects between countries with early and countries with late tracking, both in terms of track placements and student performance. Next, we use within country variation in the age of tracking to obtain estimates of early tracking on long term outcomes. We do this by exploiting changes across cohorts in the age of tracking that occurred within countries in the period between 1939 and 2012.

6.3.1 Conceptual framework

Section 6.2.3 refers to evidence that relatively young students are more likely to be placed in the lower track. We explain this fact as a natural consequence of relative age effects: age difference of up to 11 months within a class cohort cause relative age effects in which the relatively young students perform worse than the relatively old students. Since track placement is often based on a performance measure, for instance a test score or a teacher recommendation, and this measure often does not control for age differences of students in one class, the track placement decision is biased. Hence, due the performance difference between relatively young and relatively old students (the relative age effect) at the time of track placement, relatively young students are more likely to be placed into the lower track. Observed performance is a function of ability, age and relative age, as shown by equation (1). Relative age ranges from 1 to 12: 1 is allocated to students who were born just before the school start cut off and 12 to the student born just after the school start cut off. Since the relative age bias fades out over time we expect that the later the track decision is made, the less biased the used performance measure will be. Consequently, relatively young students in a country that has early tracking are more likely to be placed in the lower track than the relatively young students in a country that has late tracking. It might even be that in a country that has late tracking the relative age bias has faded out completely, since we assume that after some threshold c^* the bias has disappeared, as is shown in equation (2), and thus no misallocation takes place.

$$Performance_t = \beta * Ability + \underbrace{\delta * \frac{Rel.Age}{Age_t}}_{\text{Relative age bias}} \quad (1)$$

$$\lim_{t \rightarrow c^*} \frac{Rel.Age}{Age_t} = 0 \quad (2)$$

6.3.2 Differences in relative age effects between early and late tracking countries

In the first part of our empirical strategy we show that, due to exogenous variation in MOB and school starting cut off dates, relatively young students are more likely to go to the lower track, as is consistent with the literature. But, more importantly, we show that relatively young students in countries that track early are even more likely to go to

the lower track than relatively old students than they would in countries that track later. Following this we also look at performance differences after tracking has taken place. For our short term outcomes we estimate models similar to difference-in-differences models. The first difference is the difference in performance between relatively young and relatively old students. The second difference is between early tracking and late tracking countries. Equation (3) shows the short run models in which Y_{ic} either is a dummy for being in the low track for student i in country c or the PISA math test score for student i in country c .

$$Y_{ic} = \beta_1 RelAge_{ic} + \beta_2 RelAge_{ic} * EarlyTracking_c + \mu_c + \epsilon_{ic} \quad (3)$$

Since we compare two groups of countries with each other, those that track early and those that track late, we can also include country fixed effects (μ_c) to only look at within country differences. If relative age differences are common across countries then the difference in relative age effects can be attributed to the difference in the age of first tracking. Hence, estimates of the parameter β_2 will yield the causal effect of early tracking if the relative age effect in late tracking countries would have been the relative age effect in early tracking countries if they had tracked late. Since we compare two groups of countries and are therefore able to use country fixed effects, only other characteristics common over early or late tracking countries could influence our results. Between country differences do not influence our results. However, countries might try to mitigate the consequences of this mechanism by addressing relative age differences in performance, and this might be correlated with the timing of tracking. For instance, parents in early tracking countries might be more inclined to hold relatively young students back a year, or early tracking countries might differ in acceleration or retention policies. For the long term outcomes we therefore also employ another source of variation: within country variation in age of tracking.

6.3.3 *Exploiting changes in age of tracking within countries to obtain long term outcomes*

To look at the effect of early tracking on long term outcomes we use a second approach based on changes in the age of first selection within countries. In our strategy, as explained above, we compare the relative age effect across countries that track early and those that track late. However, several countries have changed their age of first tracking in secondary school since 1939. For instance England has changed its age of first selection from 14 to 11 in 1944 and to 16 in 1966. These changes provide us with an extra source of variation and allow us to look at between cohorts b within country c changes in the effect of early tracking on individual educational and labor market outcomes using equation (4).

$$Y_{ibc} = \beta_1 RelAge_{ibc} + \beta_2 RelAge_{ibc} * EarlyTracking_{bc} + \beta_3 EarlyTracking_{bc} \quad (4) \\ + \beta_4 Age_{ibc} + \mu_c + \mu_c + \epsilon_{ibc}$$

By exploiting within country changes in age of tracking, we can make use of when a country switches from an early to a late tracking country and vice versa. For each cohort we therefore estimate the difference in relative age effects between countries that track early and countries that track late, with a different composition of the two country groups for some cohorts. Employing this extra variation helps us to exclude any time-invariant systematic differences between countries that track early and those that track late.

6.4 Data: PISA and PIAAC

To look at both short run and long run effects we make use of two datasets: PISA 2009 for the short run effects and PIAAC 2012 for the long run effects.

6.4.1 PISA

PISA 2009 contains information on 15 year old students and their schools from all 36 OECD countries.⁶⁵ Per country a representative sample of schools is selected and all the students in that school of the relevant age group participate in PISA. Students make tests in reading, mathematics, and science, and fill in a survey on their home situation, school work and other related issues. Subsequently for each student a large number of background characteristics are known, for instance age, gender, parental background, and the country specific track he or she is at age 15. Besides student background characteristics, from each student we have their test scores on mathematics, reading, and science.

Since we are interested in the difference in the relative age effect on track placement between early and late tracking countries, we only keep those countries in which students are placed into tracks at the age we observe them (age 15). This means that we delete the countries that track after the PISA tests are administered (the non-tracking countries). This leaves us with 160,733 students in 20 countries.⁶⁶ Table 1 gives an overview of these countries. The age of tracking is obtained from the OECD and Eurydice (see Table 1) and differs across countries ranging from age 10 for Germany and Austria, to 15 for France, Portugal, and four other countries. For our difference in difference strategy we use three options for the definition of an early

65. We restrict ourselves to OECD countries since the data from non-OECD countries often showed unclear school cut off dates and because our two methods for determining which national tracks can be considered low and high tracks often contradicted each other for these countries.

66. We deleted 11 students who had an unknown country specific track and students in tracking countries who were not tracked yet due to grade repetition at the time of the PISA tests (293 students).

tracking country: A country is an early tracking country if tracking takes place before the age of 12, 13, or 14. When the cut off is placed at age 12, only 6 countries from the 20 are early tracking countries; when we use age 13 as a cut off the sample is balanced with 8 countries before the cut off and 12 after; with age 14 there are 9 early tracking countries and 11 late tracking countries.

Table 1: Descriptive statistics of the PISA sample

Country	Prop. in the low track	Age of first selection	Observations
Germany	0.43	10	4979
Austria	0.48	10	6583
Czech Rep	0.49	11	6064
Slovak Republic	0.56	11	4555
Hungary	0.61	11	4605
Turkey	0.71	11	4996
Mexico	0.27	12	38250
Netherlands	0.53	12	4759
Luxembourg	0.49	13	4622
Italy	0.25	14	30780
Belgium-Flemish	0.26	14	4431
Korea	0.29	14	4987
Belgium-French	0.47	14	3620
Slovenia	0.65	14	6154
Israel	0.14	15	5761
Greece	0.17	15	4969
Japan	0.25	15	6088
France	0.45	15	4298
Portugal	0.49	15	6298
Ireland	0.64	15	3934

Sources: Age of first selection comes from OECD (2007), except for France OECD (2003) and BFL and BFR Eurydice (2013). The proportion of students in the lower track(s) comes from PISA 2009.

The countries in PISA differ widely in the number of tracks they offer to their students. The official track number ranges from 1 to 5 as can be seen in Table 1. However, in PISA for a large number of countries also further subdivisions of tracks are known. To obtain an internationally comparable track measure for all students, we employ two methods. First, we manually assign each track to the definition low or high track based on the name and description of the track provided in PISA. Second, we use the mean performance per country specific track to categorize the different tracks in all the countries as either the low or the high track. We check whether the two methods coincide and for the largest part they do. Table 2 shows as an example all the available tracks in Germany and our categorization of the German low and high tracks and Table 1 shows the proportion of students in the low track for all the countries.

Table 2: The allocation of the available tracks in Germany into high and low track

Track description	Mean math score	% of students	High track
Lower secondary with access to upper secondary (comprehensive)	379.18	3.60	0
Lower secondary without access to upper secondary, providing a basic general education (grades 5 - 10)	396.63	0.98	0
Lower secondary without access to upper secondary, providing a basic general education (grades 5 - 10)	401.35	1.43	0
pre-vocational training year upper secondary level	410.00	0.88	0
Lower secondary without access to upper secondary, providing a basic general education (gr 5 -10)	415.39	0.50	0
Lower secondary without access to upper secondary, providing a basic general education (grade 5 - 10)	424.99	17.03	0
Vocational school	451.87	1.43	0
Lower secondary without access to upper secondary, providing an extensive general education (grades 5 - 10)	465.95	2.63	0
Lower secondary with access to upper secondary (comprehensive)	481.49	7.41	0
Lower secondary with/without access to upper secondary, providing an extensive general education (gr 5 - 10)	483.64	1.87	0
Lower secondary without access to upper secondary, providing an extensive general education (grades 5 - 10)	487.30	2.11	0
Lower secondary with access to upper secondary (comprehensive)	487.51	1.45	0
Vocational school upper secondary level	500.56	1.25	0
Lower secondary without access to upper secondary, providing an extensive general education (grades -5 10)	518.60	24.82	1
Lower secondary with access to upper secondary (comprehensive)	572.95	2.31	1
Lower secondary with access to upper secondary	595.30	29.89	1
Upper secondary level of education	604.23	0.02	1
Upper secondary level of education	673.15	0.40	1

Source: Own calculations with PISA 2009 data.

The relative age of students is constructed using their MOB and the school starting month for each country. We use the school starting month for most of our countries from Gerritsen and Webbink (2013). We supplement this dataset with cutoffs found in the data for five countries for which we could find no school starting month or for which the data showed a different starting month than those collected by Gerritsen and Webbink (2013).⁶⁷ We define relative age in such a way that the youngest in each cohort have a value 1 and the oldest have a value 12.

67. The four countries for which we use empirical school starting dates are Ireland, Israel, Mexico, and Turkey.

6.4.2 PIAAC

To look at long run outcomes, we use data on adults taken from the Program for International Assessment of Adults Competencies (PIAAC). PIAAC is an international survey among the adult populations of 24 OECD countries. The survey collects test score data on literacy, numeracy and problem solving, and on the education and work history of the participants and other life outcomes. The sample consists for each country of a representative sample of the entire working age population age 16 to age 65. We restrict our sample to males age 25 to 50. This leaves us with 38,809 participants in 21 countries.⁶⁸ PIAAC contains a wide variety of questions on education and work history of the participants, but unfortunately does not contain the secondary school track the individuals attended. For ease of interpretation of our difference in difference analyses, we construct dummy variables on education and employment. Descriptive statistics on all variables are displayed in Table 3. It displays a subset of the available variables in PIAAC and groups them into education, employment, and earnings outcomes. Education outcomes are available for the largest number of observations (31,585-46,519 males) and contain a dummy for whether or not the respondent obtained a tertiary education degree and the age at which respondent's left formal education. The employment variables contain whether or not the respondent was employed, whether he was self-employed, whether the job he was in required a tertiary education degree and whether he has a supervisory role in his job or not. Limited earnings information was available for 26,918 males and this was transformed into a dummy for earning above the mean, above the mean of the own age group and belonging to lowest ten percent earners by age group.

For our causal estimates on whether relative age differences in life outcomes are larger in early tracking countries, we supplement the PIAAC data with time changing data on countries age of tracking. We use this panel data on tracking to categorize early tracking countries separately for each cohort, as some countries have changed their tracking regime over time. This time changing data is collected by Braga, Checchi, Meschi (2013) and was generously shared by the authors for this chapter. The dataset contains time changing data from 1929 to 2000 on a number of education system characteristics and shows quite some changes over time. For instance, Flanders changed the age of first selection from age 14 to age 15.5 in 1971 and to age 12 in 1985 and England changed from age 14 to age 11 in 1944 and again to age 16 in 1966. For students entering the tracked school system after 2000, we use the age of first

68. Of the 24 countries in PIAAC we delete three countries for which we have missing data: For Cyprus we do not know when the school starting month is and for Austria and Russia only the age in years is available in PIAAC. From the 20 countries we use in the short run analyses, 9 countries are not included in PIAAC: Flemish Belgium, Greece, Hungary, Israel, Luxembourg, Mexico, Portugal, Slovenia, and Turkey. Vice versa, Canada, Denmark, England, Spain, Finland, Norway, Poland, and Sweden were not included in the short run analysis but are included in the long run analysis because these countries track after the PISA tests are administered and were thus excluded in the short run analyses.

selection as used in the PISA analyses as described above. We allocate the age of first selection into tracks to each participant in which he/she entered the tracked system. For example, using the example of Flanders also mentioned above, a person born in 1956 in Flanders was tracked at age 14, while a person born the year later in 1957 was tracked at age 15.5. Among the 14 countries, 17 times countries increased the age of tracking, while the age was decreased 5 times. Nine countries actually moved from an early to a late tracking country or vice versa when we define early tracking as tracking before the age of 13. The school starting months in this long run dataset are taken from the same sources as described for the short run analyses.

Table 3: PIAAC descriptive statistics

Variable Description	Obs.	Mean	St. Dev.
EDUCATION			
Highest level of formal education obtained is or above ISCED 5, i.e. tertiary education	22778	0.45	0.50
Continuous respondent's age when leaving formal education (min. 12, max.33)	18304	21.51	4.39
EMPLOYMENT			
Required education level in current work is ISCED 5 or above	15400	0.46	0.50
Employed vs. Unemployed/Out of the labor;	22774	0.75	0.44
Supervising status at job or business	16728	0.26	0.44
Self-employed vs. not	16728	0.09	0.28
EARNINGS			
Monthly earnings (including bonuses for wage and salary earners and self-employed) above mean	15395	0.25	0.43
Monthly earnings (including bonuses for wage and salary earners and self-employed) above mean by five age group	15395	0.25	0.43
Monthly earnings (including bonuses for wage and salary earners and self-employed) below 10% by age group	15395	0.10	0.30
<i>Notes:</i> Males aged 25 to 50 in 14 countries.			

6.5 Results

6.5.1 Short run results

Tracking

Previous studies using German, French, and Canadian data have shown that relatively young students are more likely to go to the lower track. We find the same pattern using cross country data, as illustrated by Figure 1 and Table 4. Figure 1 shows the link between relative age and the probability of enrolment in a low ability track for early tracking countries (tracking occurs before the age of 13) and late tracking countries (tracking occurs on or after the age of 13). We observe in both types of countries that the probability of enrolment in the low ability track is higher when students are

relatively younger as compared to those students who were born just after the school starting cut off. However, and most important, this difference in the probability to go to the low track for young and old students is much higher in early tracking countries.

Figure 1: Probability of Going to Low Track by Month of Birth for Early Vs Late Tracking Countries

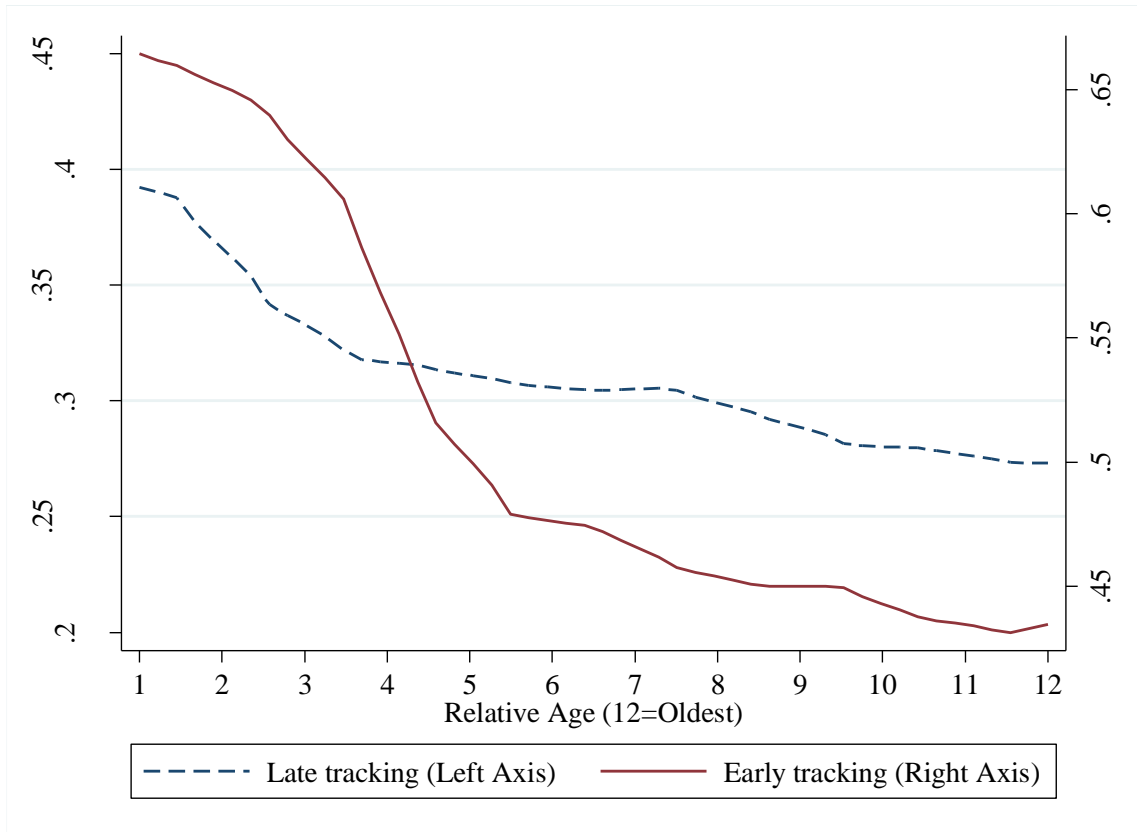


Table 4 shows this result more formally, and estimates equation (3). For three different definitions of early tracking, based on the age threshold, being in the low track is related to relative age and early tracking. As can be seen from the table, the relatively oldest are less likely to be in the low track, as is consistent with the literature. A new insight however is that Table 4 also shows that probability to be in the low track is different for students in countries with early tracking as compared to students in countries with late tracking. When the definition of early tracking is tracking before the age of 13 or 14 (the two right columns of Table 4), relatively older students are even less likely to go to the low track. Or in other words, relatively young students are more likely to go to the low track and even more so in a country that tracks before the age of 13 or 14. Controlling for parental background does not change these results, although students with higher parental background are less likely to go to the low track for any relative age (not shown). In this chapter we exploit the exogenously induced difference between early and late tracking countries for investigating the short term and long term consequences of (early) tracking. From here on we use the threshold at 13 since for this threshold the number of countries before and after the cut off is more balanced (8 before, 12 after) and since we are mainly interested in the largest

difference in the probability to go to the lower track and we aim to show the consequences of this difference.⁶⁹

Table 4: Short Run Evidence: Impact of Relative Age and Tracking on the Probability of Being in Low Track

	Threshold at age:	12	13	14
		Low track	Low track	Low track
Relative Age * Early Tracker Country Dummy		0.011*** (0.001)	-0.016*** (0.001)	-0.016*** (0.001)
Relative Age (12=oldest)		-0.020*** (0.000)	-0.009*** (0.000)	-0.009*** (0.000)
Constant		0.523*** (0.002)	0.524*** (0.002)	0.524*** (0.002)
# of students		160,733	160,733	160,733
# of countries		20	20	20
R ²		0.11	0.11	0.11
# of countries with early tracking		6	8	9

Notes: The table presents coefficients (standard errors in parenthesis) from fixed effect models, as depicted in equation (3), using individual weights provided by PISA. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Relative age ranges from 1 for the youngest, to 12 for the oldest.

Math scores

Table 5 shows the effects of early tracking for the relatively young and old students on the PISA mathematics, reading, and science test score at age 15. As is consistently found by others, the relatively old students perform better than the relatively young students. However, in countries that track early the difference in performance between the relatively old and young students is almost half of what it is in countries that track late. So although the relatively young students are more likely to be in the lower track in countries that track early, they do perform better relative to their counterparts in countries that track late.

69. We also did the following analyses with the early tracking threshold at 12 or 14 and the results are qualitatively very similar.

Table 5: Short Run Evidence: Impact of Relative Age and Tracking on Student Performance

	Math	Reading	Science
Relative Age* Early Tracker Country Dummy	-0.682*** (0.133)	-0.690*** (0.134)	-1.050*** (0.133)
Relative Age (12=oldest)	1.234*** (0.094)	1.361*** (0.095)	1.571*** (0.094)
Constant	480.700*** (0.488)	478.900*** (0.489)	482.900*** (0.487)
# of students	160,733	160,733	160,733
# of countries	20	20	20
R ²	0.18	0.13	0.20

Notes: The table presents coefficients (standard errors in parenthesis) from fixed effect models, as depicted in equation (3), using individual weights provided by PISA. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Early tracking countries are those who track before the age of 13. Relative age ranges from 1 for the youngest, to 12 for the oldest.

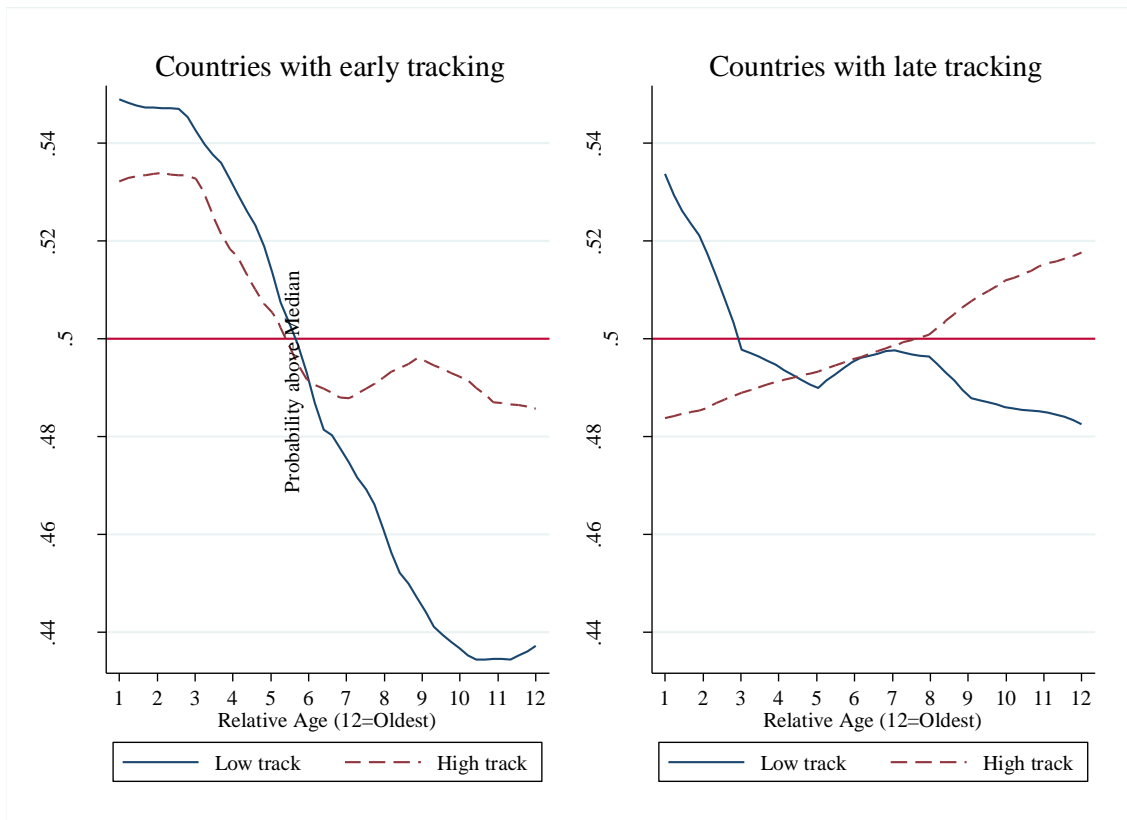
6.5.2 Mechanism

The short run results from Section 6.5.1 might seem surprising given that being placed into the lower track is thought to have negative consequences for performance. However, under the assumption that ability is not correlated to month of birth and that schools do not correct for the relative age bias when placing students into tracks, relatively young students are more often placed in a track that is too low for them in a country that tracks early than in a country that track later. This causes the ability distribution in the low track to differ between countries that have early and countries that have late tracking. In education systems that track late the high track consists purely of high ability students and the low track consists purely of low ability students, but in countries that track early the low track consists of low ability students and of high ability relatively young students. The reason that the high ability relatively young students are in the low track in a systems that tracks early is that at the time of track placement their performance measure was not a unbiased measure of their ability, but it had a downwards bias due to the then still prevalent relative age bias. Over time, their performance became a less biased measure of their ability level due to the fading out of the relative age bias. As a consequence of this fading out the performance level of the more able relatively young students surpasses the performance level of the low ability relatively old students making them the best of the track.

Figure 2 illustrates this principle. Figure 2 shows the probability to have a PISA mathematics test score above the median score within each track over relative age and for students in countries that track early and those that track late. In countries that track late, the relatively young students hardly outperform the relatively old students in the

low track, while the relatively old students hardly outperform the young in the high track. If we turn to students in countries that track early we see that in the high track, the relatively young do slightly better than the relatively old, similarly to the pattern we saw in countries that track late. However, we also see that the relatively young perform more often above the median than the relatively old students in the low track.

Figure 2: Probability to have a math score above the median in the own track over low and high track for countries with early vs late tracking



To investigate this further we turn to Table 6 which presents similar results to Table 5 but now controls for the track level of the student. After inclusion of the track level, the interaction between relative age and early tracking increases from -0.68^{***} to -2.08^{***} for math, and similarly for reading and science. Given their track level, relatively old students have up to a quarter standard deviation lower math scores in early tracking countries than their peers in late tracking countries. Since the relatively young students in early tracking countries are more likely to be in the lower track, controlling for the lower track even more clearly shows that the relatively old students take up a lower position in the math ability distribution than the relatively young students in early tracking countries. In general, being in the low track is bad for performance, with those in the low track scoring almost a full standard deviation lower than students in the higher track.

Table 6: Impact of Relative Age and Tracking on Student Performance in PISA 2009 Controlling for Track Level

	Math	Reading	Science
Relative Age* Early Tracker Country Dummy	-2.078*** (0.119)	-2.101*** (0.119)	-2.383*** (0.120)
Relative Age (12=oldest)	0.396*** (0.084)	0.513*** (0.084)	0.770*** (0.085)
Low track	-89.920*** (0.438)	-90.900*** (0.439)	-85.850*** (0.442)
Constant	527.800*** (0.491)	526.500*** (0.492)	527.800*** (0.495)
# of students	160,733	160,733	160,733
# of countries	20	20	20
R ²	0.35	0.32	0.35

Notes: The table presents coefficients (standard errors in parenthesis) from fixed effect models, as depicted in equation (3) including a control for the low track, using individual weights provided by PISA. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Relative age ranges from 1 for the youngest, to 12 for the oldest.

6.5.3 Long run effects

The mechanism described in Section 6.5.2 could have long run consequences for two reasons, which have opposing effects on long run outcomes. First, being in the low track might negatively influence the performance of the relatively young students since the curriculum is designed for less able students and might learn the able relatively young students less than the high track would have. So we could expect a negative effect of the larger probability to be placed into the lower track for relatively students in countries that track early. Students placed in the low track are also most likely to finish secondary school with a lower educational degree which could have large long run negative effects due to attainment differences.

Second, the relatively young students turn out to be more able than the relatively old students in the lower track in countries that track early. When in countries that track early students are placed into tracks, the observed performance of the relatively young students is a biased measure of their ability due to the relative age bias. Their ability level is actually higher than observed at the time, but this bias will disappear over time. When this bias has disappears and the true ability of the relatively young students is mirrored in increased performance, the relatively young students turn out to be the best in their class and even in their track. This might motivate the students and cause even better performance resulting in a positive effect of being in the lower track. The effect running through being the best of the track could even cancel out the lower initial attainment and explain the findings of Muehlenweg and Puhani (2010) and

Dustmann *et al.* (2014) who find no long term relative age effect in Germany due to track mobility.

In the following analyses we estimate the two effects combined and we are not able to separate out the two effects since we do not have the secondary school track of individuals in PIAAC. Since we do not have information on the track of the respondents in secondary school, we look at differences in outcomes due to relative age and early and late tracking only. The following results are therefore similar to those in Table 5 where we also did not control for track level. However, since both effects have opposite signs we are able to say which of the two effects dominates.

Table 7: Impact of Relative Age and Tracking on Long Run Outcomes (Outcome variables for males aged 25 to 50)

	(1) Tert. Edu. Ob.	(2) Employed	(3) Earnings above mean	(4) Earnings above mean by age group	(5) 10% lowest earners by age group
Relative Age* Early Tracking Dummy	0.002 (0.003)	0.000 (0.002)	-0.002** (0.001)	-0.002** (0.001)	0.003* (0.002)
Relative Age	0.001 (0.001)	-0.001 (0.001)	0.001** (0.000)	0.001** (0.000)	-0.000 (0.001)
Early Tracking Dummy	0.005 (0.025)	-0.058*** (0.021)	0.004 (0.008)	0.009 (0.009)	0.004 (0.014)
Age	-0.005*** (0.001)	0.004*** (0.000)	0.001*** (0.000)	0.000** (0.000)	0.000 (0.000)
Constant	0.674*** (0.028)	0.607*** (0.019)	0.210*** (0.009)	0.234*** (0.007)	0.098*** (0.015)
# of students	22,778	22,774	15,395	15,395	15,395
# of countries	14	14	14	14	14
R ²	0.05	0.04	0.89	0.90	0.18
Country*cohort clusters	377	377	371	371	371

Notes: The table presents coefficients (robust standard errors in parenthesis) from fixed effect models, as depicted in equation (4). The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The standard errors are clustered at country birth year. Relative age ranges from 1 for the youngest, to 12 for the oldest.

Table 7 displays the causal effects of relative age and early tracking on a number of long run outcomes for males between the age of 25 and 50. We make use of within country variation on both relative age and age of tracking. The within country variation on age of tracking creates changes in the composition of the group of countries who track early or late for some cohorts. These analyses are therefore not

confounded by time-invariant systematic differences between countries that track early or those that track late. Table 7 shows that we do not find a significant effect of relative age and early tracking on educational attainment (column 1) or employment (column 2), but we do find effects on earnings at later ages (columns 3-5). Thus relative age and early tracking do not seem to matter for obtaining a tertiary degree and for being employed, but do matter for earnings at later ages. In general, relatively old students have a small earnings advantage compared to the relatively young: The oldest are one percent more likely to have earnings above the average. But column 3 also shows that being a relatively old student in school and being tracked early means that at later ages it is more likely to have earnings below the average, as compared to peers in a late tracking country: with early tracking those who were relatively old when in school are two percent less likely to have earnings above the average than relatively old students in late tracking countries.

The relatively old students are also disadvantaged with early tracking if we look at earnings by five age categories (column 4) and they are more likely to be among the bottom ten percent earners (column 5). From the two effects we expected to find, a long run negative effect for the relatively young of being placed into the low track and a positive effect of being among the best performers from the low track when on the labor market, it seems that the second effect dominates leading to positive effects of early tracking for those students at later ages.

The robustness of these long run results are tested by using different methods and by sample changes. The analyses as presented in Table 7 are for males between the age 25 and 50 and were conducted using country fixed effects, while the standard errors were clustered on 371 country birth year clusters. The same results are naturally also obtained without the clusters or when using 25 equal sized age groups. The results also stay the same when we use weights to ensure all countries have the same weight in the estimations. Since the number of observations differ between countries from 1,049 (Sweden) to 7,279 (Canada) sample size per country might have impacted the results.

Sample changes were also made for the age categories. For a sample of males between the ages of 25 to 45, 25 to 55, and 25 to 65 similar results were obtained for having earnings above the mean of earnings above the mean by age group. The results on the ten percent lowest earners are also found in these samples. For women between the age of 25 and 50 relative age and early tracking only have a negative impact on the continuous age of leaving education. There is no effect on labor market outcomes, which is not surprising given the specific and very diverse labor market paths of women.

6.6 Conclusions

In this chapter we look at the effects of early tracking on both short run and long run outcomes. We focus on inequality which arises due to relative age when countries use tracking in the education systems. This type of inequality can arise due to two established facts. First, relative age differences within a school class, which arise due to school starting cut offs, matter for educational outcomes. Second, the relative age effect naturally disappears over time when the age difference becomes relatively smaller as compared to the complete life span. Combining these two facts, we focus on relative age differences between countries that track early and those that track late. We look at short run outcomes at the age of 15, namely track placement and PISA test scores, and long run education and labor market outcomes for males between 25 and 50 in countries which participated in PIAAC.

We find that relatively young students are more likely to end up in the lower track, and even more so in countries that track early. Having been placed in the lower track is often accompanied by negative outcomes as it can restrict entry into tertiary education. However, we find that the relatively young with early tracking do better than with late tracking. The reason that the relatively young students turn out to benefit from early tracking is as follows: When more able relatively young students end up in the lower track, the ability distribution of the lower track differs between countries with early and countries with late selection: The relatively young are more often the more able students in the lower track in countries that track early. In the lower track the relatively young perform better on math, reading and science, as captured by the PISA tests, than the relatively old as compared to the high track.

For the long run analyses we use within country variation in relative age when in school and age of tracking for those aged between 25 and 50 in 2012. The use of within country variation in tracking allows us to alleviate both country and education system heterogeneity and make causal statements on the effects of relative age and early tracking on long run outcomes. Being relatively young in a system which tracks students early leads to better long run labor market outcomes compared to being relatively young in a system with late tracking. The mechanism points towards an explanation in which the young in an early tracking country are among the best of their track and graduation cohort and therefore are advantaged in the labor market when employers select the best graduates for the available jobs.

The advantage for the relatively young in countries with early tracking means a disadvantage for the relatively old students in the lower track. These students are grouped together with young students who turn out to be more able than them and the young outperform them at all stages in life: early cognitive outcomes and labor market

outcomes. Summarizing, we can say that the relative age effect which caused more relatively young students to be placed into the lower track has long lasting effects. Although the relative age effect has not disappeared in late tracking countries, it is much smaller than the effect in early tracking countries. Early tracking therefore induces an inequality in outcomes based on month of birth.

Conclusion

7.1 Summary of the main findings

This thesis contains studies into multiple aspects of tracking students in secondary education. The first two chapters after the Introduction relate to the effects of tracking on student performance and inequality and how implementation of tracking plays a role in these effects. Chapter 2 shows a positive effect of tracking on student performance at age 15 in European countries when using an instrumental variable approach. Chapter 3 uses a sample of OECD countries and compares countries with no tracking (comprehensive systems) with countries that do track and have up to five tracks available to students. When school principals consider prior performance when accepting the student to the school and students are in a tracked system with four or five tracks, tracking has a positive relation with student's performance. The intuition behind these results is that if prior performance of students is taken into account when deciding on track placement less misallocation of students across tracks occurs and tracks are more homogenous in ability. This seems especially the case when a country allows for enough formal differentiation through tracking by providing students with more than three tracks. The larger number of tracks and track placement based upon prior performance leads to a large number of homogenous tracks, which might very well be the reason for the better outcomes of the students. As discussed in the Introduction, homogenous classes might aid the teacher in maximizing the learning of all students in the class as students are closer to the average ability level the teacher is aiming for. Using prior performance when accepting the student to the school also leads to a lower parental background (PB) effect since the influence of parents is lower.

Chapter 4 provides an answer to the question whether being in the high track is good for outcomes of the marginal student and reports that for some outcomes it is. In the Netherlands the marginal student who is on the threshold of being allowed or not to go to the high track has higher reading and IQ scores and a higher perceived probability to obtain the degree when this student goes to the high track. For other outcomes, as mathematics scores or personality, being in the high or low track has no impact.

Chapter 5 and 6 look at two specific forms of inequality: Inequality due to PB and inequality due to month of birth. Both types of inequalities are well documented in western education systems. However, Chapter 5 looks at whether there is an additional PB effect on the two assignment variables for track placement in the Netherlands over and above the PB effect on ability. Additional influence on track assignment variables, in this case an elementary school exit test and the elementary school teacher track recommendation, will increase the inequality in a system. In the Netherlands especially for students from high educated parents an additional PB effect on the track assignment variables is found. Chapter 6 does more than repeat the finding that

relatively young students are more likely to be placed in the lower track. It adds to the literature the result that this is even more so the case in countries with early tracking, as opposed to late tracking. Due to this increases probability to go to the low track for relatively young students, the ability distribution of the low track becomes wider since more high ability relatively young students are placed in the low track. This provides these students with an advantage since they are the best performers of the track. This has negative effects for the relatively old students in early tracking countries: They have lower earnings at later ages and even are more likely among the bottom ten percent earners.

7.2 Conclusions

From the studies presented in this thesis two overarching conclusions arise. First, tracking students into secondary school is not by definition bad for student outcomes. Previous literature has presented conflicting evidence which pointed either to a positive (e.g. Ariga and Brunello, 2007, Brunello and Checchi, 2007), a negative (e.g. Van Elk *et al.*, 2011; OECD, 2010; Hanushek and Woessmann, 2006) or no (e.g. Jakubowski, 2009; Pekkarinen, 2008) relation of tracking with student performance and inequality. This thesis adds to this debate, for instance, by looking at specific circumstances in which tracking could be beneficial or not. Chapters 3 to 5 of this thesis suggest that the implementation of tracking is important in achieving desired outcomes and that also the degree of tracking matters. Contrary to much of the policy discussions, tracking can be positive for student performance and equality when track placement is done based on prior performance and a country has more than three tracks. These two elements cause the tracks to be more homogenous in ability, which is potentially the cause of the positive outcomes.

However, the use of prior performance to decide on track placement, does not guarantee optimal allocation of students across tracks. For an optimal allocation of students there needs to be an optimal mapping from the prior performance measure to the different tracks. Chapter 4 looks at this mapping in the Netherlands and finds that the marginal student is to some extent better off in the high track and thus in this specific case an optimal mapping has not been reached. This does not lead to the recommendation that many more students should be allowed to enter the high track, since in Chapter 4 only the effects for the marginal student was looked at. Whether allowing more lower ability students to enter the high track, or reversely, whether allowing more of the higher ability students to leave the low track for the high track has any effects on the peers in each track has not been looked at. Nor has it been looked at what the effects for the marginal student might be if many more marginal students would also switch from the low to the high track. Chapter 4 does show that at

least for students near the threshold being admitted to the higher track causes better outcomes. This suggests that the high track, with its different curriculum, higher ability peers, and perhaps different resources, is also beneficial for at least a specific subset of students in the lower track. One way to increase outcomes for this group is to recreate the same circumstances for them as for those in the high track. As far as policy is concerned, Chapters 3 and 4 highlight that adopting a new policy and implementing it are two different things.

Inequality based on parental background is very common and in Chapter 3 it is argued that this inequality is lower in education systems that track students and base track placement on ability as measured by prior performance. In these systems parents are only able to influence the track placement of their children by influencing the performance of their children and not by influencing the outcomes (track placement) directly. However, Chapter 5 provides evidence that considering prior performance to decide on track placement does not fully remove the effects of parental background. In the Netherlands, where at the time of this research an elementary school exit test and an elementary school teacher track recommendation were mandatory to decide on track placement, an additional effect of parental background is found on the two track placement assignment variables. This shows that when parents are indeed limited in their direct influence on the outcomes of their children, they might find ways to increase their indirect influence. This might be a conscious choice, or could be considered regular parenting: Parents will almost always seek out the best outcomes for their children and will, if direct influence is not possible, prepare them to obtain these. Thus requiring the use of prior performance to decide on track placement is no silver bullet that ensures improved performance and lower inequality. Just like with the considerations of using prior performance to decide on track placement as outlined in the paragraph above, using prior performance will also not likely remove the full effect of parental background.

The second conclusion, which arises from Chapter 6, is that relative age should be taken into account when formulating the mapping from prior performance to tracks. Relative age, or month of birth combined with the national cutoff date to start elementary school, influences track placement in that it introduces a bias in the prior performance of relatively young students. As a consequence from this bias the relatively young students are more likely to be placed in the low track. In turn, due to distributional shifts, this leads to a disadvantage for the relatively old in educational outcomes at age fifteen and in the labor market. To prevent this from happening, multiple things can be done. First, the relative age bias can be ruled out, for instance, by using different performance thresholds for relatively young and old students or by using an age correction on those performance measures used for track placement.

Second, school admission can be done by a rolling admission, although this might be difficult to accomplish. Third, stimulation of track mobility when the first track allocation turns out to be incorrect might repair the earlier damage done by misallocation. Fourthly, the age at which tracking takes places can be postponed until the relative age effect no longer affects track placement. It must be said however that our results show that also in education systems that track late, relative age has an effect on later outcomes. Thus both early and late tracking countries are confronted with inequalities caused by the birth month of students.

7.3 Limitations

To ensure robust findings, in the chapters of this thesis numerous checks were conducted. Still, caution must be taken in interpreting the findings from this thesis. All chapters come with their own dis- and advantages and all are dependent on the specific circumstances and data availability. Chapter 2 uses a sample of European countries and presents average effects from these countries. Furthermore, it uses an instrumental variable approach which, in the case of bad instruments, could in serious cases lead to more biased estimates than OLS. The exclusion restriction, which aims at preventing this bias, is tested and indeed more aspects than the level of tracking in education systems are affected by the instrument. This should be kept in mind when interpreting the results. However, the direction and size of the effect estimated with IV does not deviate that much from the OLS estimates and are also to some extent in line with previous studies. Chapter 3 is based on a sample of OECD countries. To avoid relying solely on cross country variation, the main models are supplemented with within country models. This is done to alleviate the country heterogeneity from cross country analyses. When students with different ability levels would sort between schools where the school principals do or do not consider prior performance on accepting the students to the school, this could influence the results. Therefore also models using solely the variation of the use of prior performance by school principals between countries are presented. The results from Chapter 4 and 5 are specific to the Dutch context as data from the Netherlands used and must be interpreted as such, although the results could be indicative for other contexts. Chapter 6 uses data from OECD countries and therefore presents average effects and is not able to zoom into country specific circumstances.

7.4 Future research

This thesis contributes to a deeper understanding of the effects of tracking on student performance and inequality. However, it also opens up new questions. The analyses in this thesis raise questions of great consequence for policy makers and other

stakeholders: What is the optimal number of tracks? What are the optimal thresholds for prior performance to enter into these tracks? What is the optimal age of tracking to ensure the relative age bias is not affecting track placement? Or how should prior performance be corrected for to eliminate the relative age bias in track assignment measures? Also questions on the effect of curriculum differences or curriculum changes can ensure a deeper understanding of the different effects that tracking secondary school students might have.

The chapters of this thesis have not discussed all questions related to tracking, student performance and inequality. This thesis focused almost exclusively on cognitive outcomes and inequality, and to some extent on non-cognitive and labor market outcomes. Other outcomes, like civic outcomes, were not taken into account but might very well be important for the stakeholders in education. And tracking students in secondary education might have an impact on the development of those outcomes.

Furthermore, since currently the only cross nationally cognitive tests for students are on general education subjects (languages, mathematics and science), it might be that comprehensive systems have an advantage in cross national research as compared to tracked systems. In tracked education systems, especially those with vocationally orientated tracks, part of the student population might encounter little general education in secondary school, leading automatically to worse cognitive outcomes on these subjects. However, the vocationally oriented students might very well outperform students in other countries in vocationally oriented tests. Whether this is the case and how this influences the debate on the performance differences between comprehensive and tracked education system is open for further research.

As touched upon in the Introduction, little attention is given to the different channels through which the effects of tracking reported in this thesis run: The homogenous classes which are easier to teach but have different peers, the curriculum effects and possible resource differences. Future research should also shed more light on these topics which were not investigated in the chapters of this thesis.

Finally, the identification strategies used in this thesis are the best available given the data and the questions. This does not imply that the answers given to the questions in this thesis are final. Like in any scientific effort, they will have to be replicated using data from other regions and using different methodologies.

Addendum: Valorization

A number of countries have changed the tracking regime for students in secondary education in the twentieth century (see also Chapter 6) and still every few years there are political and societal debates on the desirability of tracking students. This thesis discusses in the different chapters questions that are often leading in the political and societal debate on tracking: (1) Does tracking increase the average cognitive development? and (2) Does tracking increase inequality among students? This section discusses five policy recommendations following from this thesis which could aid in these debates, focusing on recommendations that do not require a complete overhaul of the education system.

- *To try to increase student performance in a highly tracked education system, society could decide to base track placement on prior performance. (Based on Chapter 3)*

Chapter 3 looks at whether the relation of tracking and student performance is different under different circumstances, namely when school principals consider prior performance when accepting students or not. It shows that tracking has a more positive relation with student performance if prior performance is taken into account. This suggests that using prior performance for track placement increases performance in a tracked system.

However, Chapter 3 does not present causal evidence saying that highly tracked education systems that introduce track placement based on prior performance achieve better results after the change, but the results can be interpreted as suggestive for such a causal inference. Further research is needed to confirm this recommendation.

- *If society wants to reduce the influence of parental background on student performance, one possible way to do this might be by using prior achievement to select students into tracks. (Based on Chapter 3)*

Students in highly tracked countries that are in schools where the principal considers prior performance when accepting the student experience a lower effect of parental background than students in a comprehensive system. This is intuitive since if students are accepted based on prior performance the track choice of the parents is only one of the factors involved. Or it could be that in those systems parental choice is not even taken into account at all. By making it mandatory for school principals to consider prior performance for track placement, the influence of parents might be reduced. However, as Chapter 3 does not present causal evidence this recommendation is based on suggestive evidence only and should also be interpreted that way.

- *Secondary school principals in the Netherlands should be more lenient in accepting the marginal student to the highest track. (Based on Chapter 4)*

Chapter 4 shows that in the Netherlands the marginal student on the threshold from going to the higher track (VWO) and the middle track (HAVO) performs better on a number of dimensions in the higher track, while it is not harmed on other dimensions, at least not on the dimensions looked at in this study. The students perform better on a reading and IQ test and they also have a more positive self-perceived probability to obtain the degree. No negative effects on other cognitive and non-cognitive skills were found. This means that there seem to be only benefits for going to the higher track for the marginal student. For the marginal student it would therefore be better if school principals are more lenient in accepting this student to the higher track.

However, students who are already in the higher track might suffer negative consequences from allowing one extra (low performing) student to the class. In the sample of Chapter 4 this would mean raising the median class size from 19 to 20 and shifting the mean ability level somewhat down. This peer effect is not looked at in Chapter 4 and thus no definite answer can be given here. On the other hand, given the literature on peer effects it is not unreasonable to assume that allowing one extra low performing student to the class has a low impact on the outcomes of the other students, especially if the mean class ability does not decrease much from such a change (Sacerdote, 2011).

- *If society wants to reduce the influence of parental background on track placement, not only the influence of parents on the child should be addressed but also the influence of the parents on the teacher of the child. (Based on Chapter 5)*

Chapter 5 investigates the influence of parents on the elementary exit test score and the elementary school teacher track recommendation in the Netherlands. I find that on both performance measures there is an influence of parents over and above the general influence they have on their child's ability, although only the influence on the recommendation seems to be very robust. The results suggest that either teachers are influenced by parents to provide a different recommendation than justified by the performance of the child, or teachers might (unconsciously) make different considerations for children with specific parental characteristics. The results in Chapter 5 show that parents have an influence on track placement through two channels: through their influence on their child's ability and through an additional influence on the teacher recommendation. If parental influence is deemed undesirable, it is important to not only counteract parental influence on the ability of their children but also the additional parental influence on the teacher.

In a broader context, the results from Chapter 5 can be seen as only one example of a study into the influence of parental characteristics on teachers. There is also indication

that teachers behave differently towards children of different parental backgrounds and their parents by, for instance, lowering expectations (Jussim *et al.*, 1996) or by providing a specific environment most suitable for specific groups of parents (Lareau, 1987). If parental influence is deemed undesirable by society then also these routes of parental influence should be addressed.

- *Test scores for young students should be corrected for their relative age when one wants to compare results across students. (Based on Chapter 6)*

Chapter 6 looks at the effect of early tracking and relative age on short and long term outcomes. The results show that relatively young students are more likely to go to the lower track and more so in countries that track early. The reasoning for this is that, as captured in the relative age effect, relatively young have lower student performance than relatively old students in the early school years. When early on students are selected into tracks based on this student performance and no correction for age in month is made, relatively young students have lower scores than relatively old students, are thus more often deemed to be the weaker students, and thus more likely to be placed in the lower track.

Chapter 6 looks only at the effect of relative age and early tracking on outcomes, but the relative age effect itself also has an effect on outcomes, for instance by lowering the chances of relatively young students to go to the higher track. However, also in earlier stages, day to day contact between educators and children, and in different fields, like the sport field, the relative age effect plays a role. Due to this performance difference (relatively young and old) children might not be challenged or supported as much as they need to excel. To prevent this and to be able to accurately compare or rank the performance of students, performance needs to be corrected for relative age differences.

References

- Acemoglu, D., D. Cantoni, S. Johnson, and J. Robinson (2011) The Consequences of Radical Reform: The French Revolution. *American Economic Review*, 101, pp. 3286–3307.
- Allen, J. and R. Barszley (1993) The Interaction of Ability, Maturity, and Training in Systems with Age-Dependent Recursive Selection. *The Journal of Human Resources*, 28(3), pp. 649-659.
- Ammermueller, A. (2005) Educational Opportunities and the Role of Institutions. *ZEW Discussion Paper*, no. 05-44.
- Angrist, J. and A. Krueger. (1991) Does Compulsory School Attendance Affect Schooling and Earnings? *The Quarterly Journal of Economics*, 106, pp. 979-1014.
- Archer, M. (1979) *Social Origins of Educational Systems*. SAGE Publications Ltd: London.
- Ariga, K. and G. Brunello (2007) Does Secondary School Tracking Affect Performance? Evidence from IALS. *IZA Discussion Paper*, no. 2643.
- Barg, K. (2013) The influence of students' social background and parental involvement on teachers' school track choices: reasons and consequences. *European Sociological Review*, 29(3), pp. 565-580.
- Becker, S. and L. Woessmann (2009) Was Weber Wrong? A Human Capital Theory of Protestant Economic History. *The Quarterly Journal of Economics*, 124(2), pp. 531-596.
- Bedard, K. and E. Dhuey (2006) The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects. *Quarterly Journal of Economics*, 121(4), pp. 1437-1472.
- Betts, J. and J. Shkolnik (2000) The effects of ability grouping on student achievement and resource allocation in secondary schools. *Economics of Education Review*, 19, pp. 1-15
- Black, S, P. Devereux, and K. Salvanes (2005) Why the Apple Doesn't Fall Far: Understanding Intergenerational Transmission of Human Capital. *American Economic Review*, 95, pp. 437-449.
- Black, S., P. Devereux, and K. Salvanes (2011) Too Young To Leave the Nest? The Effects of School Starting Age. *The Review of Economics and Statistics*, 93(2), pp. 455–467.
- Boli, J., F. Ramirez, and J. Meyer (1985) Explaining the Origins and Expansion of Mass Education. *Comparative Education Review*, 29(2), pp. 145-170.
- Borghans, L., R. Diris, W. Smit, and J. De Vries (2011) *The labor market effects of tracking: a regression discontinuity approach*. Thesis University Maastricht Ron Diris.
- Boudon, R. (1974). *Education, inequality and social opportunity*. New York: Wiley.
- Braga, M., D. Checchi, and E. Meschi (2013) Educational policies in a long-run perspective. *Economic policy*, 28(73), pp. 45-100.
- Brunello, G., and D. Checchi (2007). Does school tracking affect equality of opportunity? New international evidence. *Economic Policy*, (October), 781–861.
- Byun, S. and H. Park (2012) The Academic success of East Asian American Youth The Role of Shadow Education. *Sociology of Education*, 85(40), pp. 40-60.
- Cantoni, D. and N. Yuchtman (2014) Medieval Universities, Legal Institutions, and the Commercial Revolution. *The Quarterly Journal of Economics*, 129(2), pp. 823–887.
- CBS (2012) *Jaarboek onderwijs in cijfers 2012*. Centraal Bureau voor de Statistiek: Den Haag/Heerlen.
- CITO (2014) *Cito dossier Eindtoets 2014*. Arnhem: CITO.
- CITO Score (2014) *Schooladvies*. <http://www.Cito-score.nl/schooladvies.html>. Last accessed 24-02-2014.
- Clark, D. (2010) Selective Schools and Academic Achievement. *The B.E. Journal of Economic Analysis & Policy Advances*, 10(1), article 9.

References

- Crawford, C., L. Dearden, and C. Meghir (2013) When You Are Born Matters: Evidence for England. *IFS Report*, no. R80. London: Institute for Fiscal Studies.
- Cunha, F., J. Heckman, L. Lochner, and D. V. Masterov (2005) Interpreting the Evidence on Life Cycle Skill Formation. *NBER Working Paper*, No. 11331.
- Dittmar, J. (2011) Information Technology and Economic Change: The Impact of The Printing Press. *The Quarterly Journal of Economics*, 126(3), pp. 1133-1172.
- Dollmann, J. (2011) Verbindliche und unverbindliche Grundschulempfehlungen und soziale Ungleichheiten am ersten Bildungsübergang. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 63(4): 431-457.
- Driessen, G., L. Mulder, and J. Roeleveld (2012) *Cohortonderzoek COOL5-18. Technisch rapport basisonderwijs, tweede meting 2010/11*. Nijmegen: ITS / Amsterdam: Kohnstamm Instituut.
- Driessen, G., L. Mulder, G. Ledoux, J. Roeleveld, and I. van der Veen (2009) *Cohortonderzoek COOL5-18. Technisch rapport basisonderwijs, eerste meting 2007/08*. Nijmegen: ITS / Amsterdam: SCO-Kohnstamm Instituut.
- Dronkers, J. and R. Korthals (under review) Tracking in the Netherlands. Ability selection or social reproduction? in H.-P. Blossfeld (eds.).in: *Differentiation in Secondary Education and its' Short- and Longer-Term Effects on Inequality of Educational Opportunities*.
- Dronkers J, R. van der Velden, and A. Dunne (2012) Why are Migrant Students Better Off in Certain Types of Educational Systems or Schools than in Others? *European Educational Research Journal*, 11(1), pp. 11-44.
- Duflo, E. P. Dupas, and M. Kremer (2011) Peer effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. *American Economic Review*, 101, pp. 1739-1774.
- Dustmann, C. (2004) Parental background, secondary school track choice, and wages. *Oxford Economic Papers*, 56, pp. 209-230.
- Dustmann, C., P. Puhani, and U. Schönberg (2014) The Long-Term Effects of Early Track Choice. *IZA Discussion Paper Series*, nr. 7897.
- Education Inspectorate (2013) *Onderwijsverslag 2012-2013*. Inspectie van het onderwijs: Den Haag.
- Essser, H. (under review) Educational Systems and Educational Inequality. What is With Ability Tracking? in H.-P. Blossfeld (eds.).in: *Differentiation in Secondary Education and its' Short- and Longer-Term Effects on Inequality of Educational Opportunities*.
- Figlio, D. and M. Page (2002) School Choice and the Distributional Effects of Ability Tracking: Does Separation Increase Inequality? *Journal of Urban Economics*, 51, pp. 497-514.
- Fredriksson, P. and B. Oeckert (2013) Life-Cycle Effects of Age at School Start. *The Economic Journal*, 124, pp. 997-1004.
- Fuchs, T. and L. Woessmann (2007) What accounts for international differences in student performance? A re-examination using PISA data. *Empirical Economics*, 32, pp. 433-464
- Gerristen, S. and D. Webbink (2013) How much do children learn in school? International evidence of the effects of time in school. *CPB Discussion Paper*, nr. 255. Den Haag: Girard, H. and A. Bastide (1963) La stratification sociale et la démocratisation de l'enseignement. *Population*, 18(3), pp. 435-472.
- Grab, A. (2003) *Napoleon and the Transformation of Europe*. Palgrave MacMillan: England.
- Grenet, J. (2010) La date de naissance influence-t-elle les trajectoires scolaires et professionnelles? Une évaluation sur données françaises *Revue économique*, 60(3), pp. 589-598.
- Guyon, N., E. Maurin, and S. McNally (2012) The effect of tracking students by ability into different schools. *The Journal of Human Resources*, 47(3), pp. 684-721.
- Hall, C. (2012) The effects of reducing tracking in upper secondary school. *Journal of Human Resources*, 47(1), pp. 238-269.

- Hanushek, E. (1986) The Economics of Schooling: Production and Efficiency in Public Schools. *Journal of Economic Literature*, 24(3), pp. 1141-1177.
- Hanushek, E. (1997) Assessing the Effects of School Resources on Student Performance: An Update. *Educational Evaluation and Policy Analysis*, 19(2), pp. 141-164.
- Hanushek, E. and L. Woessmann (2006) Does Educational Tracking Affect Performance and Inequality? Difference-in-Difference Evidence Across Countries. *Economic Journal*, 116, pp. C63-C76
- Hanushek, E. H., S. Link, and L. Woessmann (2011) Does School Autonomy Make Sense Everywhere? Panel Estimates From PISA. *NBER Working Paper Series*, no. 17591.
- Heckman, J.J. and Y. Rubinstein (2001) The Importance of Non-Cognitive Skills: Lessons from the GED Testing Program. *American Economic Review*, 91(2), pp. 145-149.
- Heckman, J.J., R. Pinto, and P. Savelyev (2012) Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes. *NBER working paper*, no. 18581, Cambridge: NBER.
- Heller, S., H. Pollack, R. Ander, and J. Ludwig (2013) Preventing Youth Violence and Dropout: A Randomized Field Experiment. *NBER working paper*, no. 19014, Cambridge: NBER.
- Hoekstra, M. (2009) The Effect of Attending the Flagship State University on Earnings: A Discontinuity-Based Approach. *The Review of Economics and Statistics*, 91(4), pp. 717-724.
- Horn, D. (2009) Age of selection counts: a cross-country analysis of educational institutions. *Educational Research and Evaluation*, 15(4), pp. 343-366.
- Horn, D. (2013) Diverging performances: the detrimental effects of early educational selection on equality of opportunity in Hungary. *Research in Social Stratification and Mobility*, 32(June), pp. 25-43
- Ichou, M. and L.-A. Vallet (2011). Do all roads lead to inequality? Trends in French upper secondary school analysed with four longitudinal surveys. *Oxford Review of Education*, 37, pp. 167-194.
- Imbens, G. and T. Lemieux (2007) Regression Discontinuity Designs: A Guide to Practice. *Journal of Econometrics*, 142(2), pp. 615-635.
- Jackson, K. (2010) Do better students benefit from attending better schools? Evidence from rule-based student assignments in Trinidad and Tobago. *The Economic Journal*, 120(December), pp. 1399-1429.
- Jackson, M. (2013). Introduction: How is inequality of educational opportunity generated? The case for primary and secondary effects. In: Jackson, M. (Ed.) *Determined to Succeed? Performance versus Choice in Educational Attainment*, pp.1-33. Stanford: Stanford University Press.
- Jakubowski, M. (2010) Institutional tracking and achievement growth: Exploring difference-in-differences approach to PIRLS, TIMSS and PISA data. In: *Quality and Inequality of Education* (2010), pp 41-81.
- Juerges, H., K. Schneider, and F. Buechel (2005) The Effect of Central Exit Examinations on Student Achievement: Quasi-Experimental Evidence from TIMSS Germany. *Journal of the European Economic Association*, 3(5), pp. 1134-1155.
- Juerges, H. and K. Schneider (2007) What Can Go Wrong Will Go Wrong: Birthday Effects and Early Tracking in the German School System. *CESifo working papers*, no. 2055. CESifo.
- Jussim, L. and K. Harber (2005) Teacher expectations and Self-Fulfilling Prophecies: Knowns and Unknowns, Resolved and Unresolved Controversies. *Personality and Social Psychology Review*, 9(2), pp. 131-155.
- Jussim, L., J. Eccles, and S. Madon (1996). Social perception, social stereotypes, and teacher expectations: Accuracy and the quest for the powerful self-fulfilling prophecy. *Advances in experimental social psychology*, 29, pp. 281-388.

References

- Karlson, K., A. Holm, and R. Breen (2012) Comparing regression coefficients between same-sample nested models using logit and probit: A new method. *Sociological Methodology*, 42, pp. 286-313.
- Kerr, S., T. Pekkarinen, and R. Uusitalo (2013) School Tracking and Development of Cognitive Skills. *Journal of Labor Economics*, 31(3), pp. 577-602.
- Kingdom of the Netherlands (1981) *Wet op het primair onderwijs*. Hoofdstuk I. Basisonderwijs. Artikel 42. Onderwijskundig rapport. Law book. The Netherlands.
- Korthals, R. (2012) Selection and tracking in secondary education: A cross country analysis of student performance and educational opportunities, *ROA Research Memorandum*, no. 014, Maastricht University: Research Centre for Education and the Labour Market.
- Korthals, R. (2015) The Pre-Tracking Effects of Parental Background. *ROA Research Memorandum*, no. 005, Maastricht University: Research Centre for Education and the Labour Market.
- Landvoigt, T. G. Muehler, and F. Pfeiffer (2007) Duration and Intensity of Kindergarten Attendance and Secondary School Track Choice, *ZEW Discussion Papers*, no. 07-051.
- Lareau, A. (1987). Social class differences in family-school relationships: the importance of cultural capital. *Sociology of Education*, 60 (2), 73-85.
- Lavy, V. and E. Sand (2015) On the origins of Gender Human Capital Gaps: Short and Long Term Consequences of Teachers' Stereotypical Biases. *NBER Working Paper Series*, no. 20909.
- Lee, D. and T. Lemieux (2010) Regression Discontinuity Designs in Economics, *Journal of Economic Literature*, 48(June), pp. 281-355.
- Lucas, S. (1999) *Tracking Inequality: Stratification and Mobility in American High Schools*. *Sociology of Education Series*. Teachers College Pr.
- Muhlenweg, A. and P. Puhani (2010) The Evolution of the School-Entry Age Effect in a School Tracking System. *The Journal of Human Resources*, 45(2), pp. 407-438.
- OECD (2003) *Literacy Skills for the World of Tomorrow - Further results from PISA 2000*. OECD: Paris.
- OECD (2007) *Science competencies for tomorrows world*. OECD: Paris.
- OECD (2010) *School Factors Related to Quality and Equity. Results from PISA 2000*. OECD: Paris.
- Panichella, N. and M. Triventi (2014), Social inequalities in the choice of secondary school. Long-term trends during educational expansion and reforms in Italy. *European Societies*, 16(5), pp. 666-693.
- Pekkarinen, T. (2008) Gender Differences in Educational Attainment: Evidence on the Role of Tracking from a Finnish Quasi-experiment. *Scandinavian Journal of Economics*, 110(4), pp. 807-825.
- Pekkarinen, T., R. Uusitalo, and S. Kerr (2009) School tracking and intergenerational income mobility: Evidence from the Finnish comprehensive school reform. *Journal of Public Economics*, 93, pp. 965-973.
- Pop-Eleches, C. and M. Urquiola (2011) Going to a better school: Effects and behavioural responses. *NBER working paper series*, no. 16886.
- Puhani, P. and A. Weber (2006) Does the early bird catch the worm? Instrumental variable estimates of early educational effects of age of school entry in Germany. *Empirical Economics*, 32, pp. 359-386.
- Ramirez, F. and J. Boli (1987) The Political Construction of Mass Schooling: European Origins and Worldwide Institutionalization. *Sociology of Education*, 60, pp. 2-17.
- Roberts, B., K. Walton, and W. Viechtbauer (2006) Patterns of Mean-Level Change in Personality Traits Across the Life Course: A Meta-Analysis of Longitudinal Studies. *Psychological Bulletin*, 132(1), pp. 1-25.

- Sacerdote, B. (2011) Peer effects in education: How Might They Work How Big Are They and How Much Do We Know Thus Far. In: Hanushek, E.A., Machin, S., and Woessmann, L. (eds) *Handbook of the Economics of Education*, Volume 3, Elsevier.
- Santon, R. and L. MacKay (2010) *Grote atlas van de wereldgeschiedenis*. Paragon: UK.
- Schneider, S. L. and N. Tieben (2011). A healthy sorting machine? Social inequality in the transition to upper secondary education in Germany. *Oxford Review of Education*, 37, pp. 139–166.
- Schuetz, G., H. Ursprung, and L. Woessmann (2008) Education Policy and Equality of Opportunity. *Kyklos*, 61(2), pp. 279-308.
- Staiger, D. and J. Stock (1997) Instrumental Variables Regression with Weak Instruments. *Econometrica*, 65(3), pp. 557-586.
- Stipek, D. (2002) At what age should children enter kindergarten? A question for policy makers and parents. *Social Policy Report*, 16(2).
- Stock, J. and M. Yogo (2005) Testing for Weak Instruments in Linear IV Regression. In: D. Andrews, J. Stock, T. Rothenberg (eds) *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg*, pp. 80-108. New York: Cambridge University Press.
- Sullivan, A., A. Heath and C. Rothon (2011), Equalisation or inflation? Social class and gender differentials in England and Wales. *Oxford Review of Education*, 37, pp. 215–40
- Tieben, N., P. de Graaf, and N. D. de Graaf (2009) Changing effects of family background on transitions to secondary education in the Netherlands. Consequences of educational expansion and reform. *Research in Social Stratification and Mobility*, 27, pp. 77–90.
- Van Elk, R., M. van der Steeg, and D. Webbink (2011) Does the Timing of Tracking Affect Higher Education Completion? *Economics of Education Review*, 30(5), pp.1009-1021
- Walldinger, F. (2006) Does Tracking Affect the Importance of Family Background on Students' Test Scores? *Working Paper*. paper.www.cepr.org/meets/wkcn/3/3527/papers/Walldinger.pdf, Last accessed 05-12-2012.
- Woelfel, O. and G. Heineck (2012) Parental risk attitudes and children's secondary school track choice. *Economics of Education Review*, 31, pp. 727– 743.
- Woessmann, L. (2003) Schooling Resources, Educational Institutions and Student Performance: the International Evidence. *Oxford Bulletin of Economics and Statistics*, 65, pp. 0305-9049.
- Woessmann (2007) Fundamental Determinants of School Efficiency and Equity: German States as a Microcosm for OECD Countries. *IZA Discussion Paper*, nr. 2880.
- Woessmann, L., E. Luedemann, G. Schuetz, and M. West (2007) School Accountability, Autonomy, Choice, and the Level of Student Achievement: International Evidence from PISA 2003. *OECD Education Working Papers*, no. 13.

Data sources

- Centennia Software (2002) *Centennia Historical Atlas: French Revolutionary and Napoleonic Era from 1789 to 1819*. www.historicalatlas.com
- Braga, M., D.Checci and E. Meschi (2013) Educational policies in a long-run perspective. *Economic policy*, 28(73), pp. 45-100.
- SCO-Kohnstamm Instituut Amsterdam, ITS Radboud Universiteit Nijmegen, CITO Arnhem, GION Groningen (2008), *Cohortonderzoek Onderwijsloopbanen van 5-18 jaar - COOL 5-18 - Basisonderwijs 2007; Eerste meting basisonderwijs 2007*. Accessed at <https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:44290/tab/2>
- Kohnstamm Instituut, Universiteit van Amsterdam, ITS, Radboud Universiteit Nijmegen (2012), *Cohortonderzoek Onderwijsloopbanen van 5-18 jaar - COOL 5-18 - Basisonderwijs 2010/2011; Tweede meting basisonderwijs 2010/2011*. Accessed at <https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:49841>

References

- Eurydice (2013) http://eacea.ec.europa.eu/education/eurydice/eurypedia_en.php GADM database of Global Administrative Areas (2013) *Codes and software for making maps in Stata*. www.gamd.org.
- Google maps (2013) <http://maps.google.com/>
- Horloge Parlante (2013) <http://www.horlogeparlante.com/nl/distance.php>.
- Institut zur Qualitätsentwicklung im Bildungswesen (2013) *PISA Extended Germany 2006*. Berlin: Forschungsdatenzentrum am Institut zur Qualitätsentwicklung im Bildungswesen.
- Kultusminister Konferenz (2013) *Übersicht über die Bildungsgänge und Schularten im Bereich der allgemeinen Bildung*. http://www.kmk.org/fileadmin/pdf/Bildung/AllgBildung/SchemaBildungsgaenge_und_Schularten-Stand_2012-12.pdf Last accessed 05-08-2013.
- OECD (2006) *Program for International Student Assessment 2006*. OECD: Paris. www.oecd.org/pisa/
- OECD (2007) *PISA 2006: Science Competencies for Tomorrows World*. Table 5.2 Structural features of school systems. OECD: Paris.
- OECD (2009) *Program for International Student Assessment 2009*. OECD: Paris. www.oecd.org/pisa/
- OECD (2010) *School Factors Related to Quality and Equity. Results from PISA 2000*. OECD: Paris.
- OECD (2011) *Education at a Glance 2011: OECD Indicators*. Table D5.1a. National examinations at the lower secondary level (2009). OECD: Paris.
- OECD (2013) *Programme for the International Assessment of Adult Competencies 2012*. OECD: Paris.
- Regional Education Monitor Limburg (2013) *KAANS data*. <http://www.kaans.nl/>
- UNESCO (1968a) *The Yearbook of Education*. Vol. XXX. UNESCO: Geneva.
- UNESCO (1968b) *National Report: Educational Developments in 1967-8*. United Kingdom. International Conference on Education; 34th; Geneva. http://www.ibe.unesco.org/National_Reports/UK/nr_pp_xxk_1968_ef.pdf Last accessed: 20-09-2013.
- UNESCO (1971a) *National Report: Organization of Education: Denmark*. International Conference on Education; 33rd; Geneva. http://www.ibe.unesco.org/National_Reports/Denmark/nr_mf_dk_1971_e.pdf Last accessed: 20-09-2013.
- UNESCO (1971b) *National Report: Educational Organization and Development in the Federal Republic of Germany in 1970/71*. International Conference on Education; 33rd; Geneva. http://www.ibe.unesco.org/National_Reports/Germany/nr_mf_gw_1971_ef.pdf Last accessed: 20-09-2013.
- UNESCO (1971c) *National Report: Report on the Organization of Education During 1970-1971, Hungary*. International Conference on Education; 33rd; Geneva. http://www.ibe.unesco.org/National_Reports/Hungary/nr_mf_hu_1971_ef.pdf Last accessed: 20-09-2013.
- UNESCO (1971d) *National Report: Report on the Educational Movement in 1970-1971, Italy*. International Conference on Education; 33rd; Geneva. http://www.ibe.unesco.org/National_Reports/Italy/nr_mf_it_1971_fe.pdf Last accessed: 20-09-2013.
- UNESCO (1971e) *National Report: Educational Trends in the Netherlands 1969-1970-1971*. International Conference on Education; 33rd; Geneva. http://www.ibe.unesco.org/National_Reports/Netherlands/nr_mf_ne_1971_ef.pdf Last accessed: 20-09-2013.
- UNESCO (1973a) *National Report: Organization of Education in Finland in 1971-1973*. International Conference on Education; 34th; Geneva.

- http://www.ibe.unesco.org/National_Reports/Finland/nr_mf_fi_1973_ef.pdf Last accessed: 20-09-2013.
- UNESCO (1973b) National Report: Report on the Organization of Education in the United Kingdom 1971-1973. International Conference on Education; 34th; Geneva.
http://www.ibe.unesco.org/National_Reports/UK/nr_mf_xk_1973_ef.pdf Last accessed: 20-09-2013.
- UNESCO (1975) National Report: Development of Education in 1973-74 and 1974-75, Sweden. International Conference on Education; 35th; Geneva.
http://www.ibe.unesco.org/National_Reports/Sweden/nr_mf_sw_1975_e.pdf Last accessed: 20-09-2013.
- World Bank (2012) *World Development Indicators & Global Development Finance Database*. Last accessed 23-03-2012. <http://databank.worldbank.org/data/home.aspx>

Appendices

Appendix A: Summary of the historical case studies of Ramirez and Boli (1987) (Chapter 2)

Prussia

Prussia already provided for quite some schooling in the 18th century. For instance, in 1716 Frederick William I “made attendance at village schools compulsory for all children not otherwise provided with instruction” but funds were not provided for. But after 1806, when Prussia lost to the French army and was subjected to large influence from France, Prussia made great strides in its education efforts. The German philosopher Johann Gottlieb Fichte in 1807 “claimed that universal state-directed, compulsory education would teach all Germans to be good Germans and would prepare them to play whatever role [...] fell to them in helping the state reassert Prussian power” (p. 5). In the same year the Bureau of Education was established and between 1817 and 1825 a state administration was set up and taxes to finance education were imposed.

Austria

Austria introduced its first universal compulsory school law in 1774 but the implementation was “frustrated by the reactionary policies in the aftermath of the French Revolution and the Napoleonic Wars” and education came back in religious control. In 1866 education came back in state control, but only after 1869 the first compulsory school law was accepted.

Denmark

In Denmark multiple attempts were made towards compulsory schooling in the 18th century, but money was never allocated to make the attempts successful. The final, and successful, attempt to introduce a Danish education system was between 1789 and 1814 under commission of Frederick VI. This period coincide with the fall of Denmark in the early 19th century. N.F.S. Grundvig, a clergymen, saw education “as a means for Denmark to regain its spiritual and national strength”.

Sweden

Before and during the Napoleonic War there more numerous proposals in Sweden on education, but they were either religiously inspired or they were defeated by clergy, aristocracy or king, who was a former general from Napoleon. After the Napoleonic Wars there was more room for liberal reform and a school reform bill in 1842 was adopted. “In Sweden, the crisis brought about by the Napoleonic Wars was quite acute,

but it did not immediately generate a state educational system, despite efforts in that direction [...] [due to] the resistance of the relatively powerful and independent peasantry (together with the standard resistance of the aristocracy and clergy) that slowed the movement towards state-controlled education” (p. 7).

Italy

Italy had its “first serious calls for state-directed schooling [...] in 1796 when Austria’s counterattack on France seemed likely to quench the [French] Revolution” (p.7). But little was achieved. Only after Italian unification was real progress made on establishing a mass public education system. However, “The Piedmontese state strove to build a national education system before national unification had been achieved” (p. 7). It was thus the north-western state of Piedmont, closest to France, that was first to seriously consider mass education in Italy.

France

“In France, education under the *ancient regime* consisted mainly of secondary schools run by religious orders for the middle and upper class and the school runs by Catholic Church”. Although Napoleon frequently pointed towards the importance of education, he ignored primary education and focused purely on elite secondary schooling (*lycées* and *Grandes Ecoles*). Only after 1833 did France made the first efforts towards mass primary education, but these efforts failed due to the “Revolution of 1840 and the regime of Louis Napoleon in 1852” (p.8). Between 1870 and 1881 an elementary school system was established.

England

In England, education was restricted to the elite and the “the movement towards state-sponsored mass education was inhibited by the very success of England’s navy and merchants [...]. Schooling developed slowly in private hands and classical liberal restraints on state action kept the state out of education much longer in England than elsewhere.” (p. 8). The first successful reform was in 1833 and only stipulated state grants for schools. After 1860 the situation changed and debate also shifted to providing education for the working classes. In 1870 the Elementary Education Act was established which provided mandatory primary education. Church control of schools was ended in 1902.

Appendix B: The country borders (Chapter 2)

The borders of the maps used in this chapter are drawn by using the command `spmap` in Stata making use of the Shape files and the dBase databases of the European countries found on www.gamd.org. Thus the maps here are drawn using the current

borders of the European countries. This means that country borders in the historical maps could (and most probably will) deviate from the historical borders in a number of ways. For simplicity it is chosen here to ignore these changes.

In a few instances was it possible to divide countries into regions more suitable for either the historical categorization of the French treat (Germany and Italy) or for the display of education systems (Germany, Belgium and Great Britain).

Germany and Italy

Germany and Italy were in the 1800s not yet the united countries as they are today. In Germany a large number of states, city states and kingdoms existed and the borders and the independency of these regions changed numerous times. To be able to match the current German states to the regions in the 1800s a division of Germany is made. In this chapter Westphalia refers to the current German states Hesse and North Rhine-Westphalia. The Rhineland are the German states Saarland and Rhineland-Palatinate. Prussia consists of Mecklenburg-Vorpommern, Berlin, Brandenburg, and Saxony-Anhalt. Bavaria is Bavaria and Baden-Wurttemberg. Hanover consists of the German states Hamburg, Bremen, and Lower Saxony, while Saxony refers to Saxony and Thuringia.⁷⁰ Schleswig-Holstein belonged to Denmark around 1800 and is therefore display as a single territory in the historical maps and is excluded from the maps depicting standardization. The standardization maps are made using PISA2006 data and it would be very difficult and to some extent arbitrary to somehow add the Danish and Schleswig-Holstein data to create one “Greater Denmark”.

Italy was similar to Germany in the 1800s in the sense that it consisted of a number of states and kingdoms and was only later united in a single country. This chapter distinguishes four regions in Italy in the 1800s: The North West consisting of current day Liguria, Piedmont, and Sardinia, The Republic of Italy consisting of Bolzano, Trento, Emilia-Romagna, Friuli-Venezia Giulia, Lombardy, and Veneto, The Kingdom of Naples: Basilicata, Campania, Puglia, Provincia Sicilia. The other Italian provinces are Abruzzo, Lazio, Marche, Molise, Tuscany, Umbria, Valle d'Aosta, and Calabria. Of these eight provinces the first six can be regarded as the Papal State while Valle d'Aosta could be assigned to the North West and Calabria was part of the Kingdom of Naples. However, in PISA2006 it not possible to distinguish between these eight provinces and therefore these provinces are excluded from the maps and analysis using standardization.

70. The division of German states into these regions, instead of using all German states separately, is used since it is not allowed to display results using PISA2006 with individual German states. It is allowed however to use clusters of states, which is what is done in this chapter.

Belgium and Great Britain

Both Belgium and Great Britain are divided in two regions that better represent the education systems (and also the cultural regions). Belgium is divided in the Flemish region and the French region (including the German speaking parts), while Great Britain is divided in Scotland and the region consisting of England, Wales, and Northern-Ireland.

Appendix C: Historical categorization of the French influence by Grab (p. 18, 2003) (Chapter 2)

Pays réunis: “countries annexed to France and directly ruled by Napoleon”.

Avignon (1791)

Savoy (1792)

Nice (1793)

Belgium (1795)

Luxemburg (1795)

Geneva (1798)

Piedmont (1802)

The Rhineland (1802)

Liguria (1805)

Kingdom of Etruria (Tuscany) (1808)

Parma (1808)

Rome and its environs (1809)

The Illyrian Provinces (1809)

Kingdom of Holland (1810)

Hanover and Hamsa cities of Hamburg, Bremen and Lubeck (1810)

The Grand Duchy of Oldenburg (1810)

Pays conquis: conquered countries, or “satellite states that were entrusted to French rulers”.

Swiss Confederation (created in 1803)

Republic of Italy (1802)/Kingdom of Italy (1805)

Kingdom of Naples (1806)

Grand Duchy of Berg (1806)

Kingdom of Holland (1806)

Kingdom of Westphalia (1807)

Duchy of Warsaw (1807)

Kingdom of Spain (1808)

Pays allies: “allied countries whose territory was expanded by Napoleon in some cases, and continued to be governed by their native rulers”.

Most of the members of the Rhenish Confederation (including Saxony and the South-German states of Bavaria, Wurttemberg, Baden)

In four cases, this chapter deviates from the categorizations of countries made by Grab (2003). Since Grab (2003) does not include Scandinavia in his book, Finland, Sweden, and Norway are not included in his categorization. In this chapter Finland, Sweden, and Norway are regarded as *pays allies*. Portugal is categorized as *pays réunis* since Portugal was invaded by Napoleon in 1807 to ensure England could not trade on the continent (Centennia Historical Atlas).

Appendix D: Standardization (Chapter 2)

Section 2.2 shows that political pressure from Napoleon seems to have had little effect on the extent of (input and output) standardization. Therefore Section 2.3 proceeds with only differentiation. For completeness this Appendix presents the same analyses for standardization as for differentiation in the main text.

Data

Standardization on inputs

Next to the background characteristics mentioned in Section 2.3, in PISA 2006 also asked some questions to the school principal on who has authority for the textbooks, course content, teacher appointments, teacher employment conditions, and budget. These last five variables are the main independent variables, which are transformed into education system characteristics on standardization by using country averages. Table D1 shows the exact definitions of the variables and some descriptive statistics. These five standardization measures are combined into two means, namely one on the administrative standardization and one on the content related issues of standardization. These five are the variables used in Section 2.2 to look at the relation between Napoleonic pressure and education systems.

Table D1: Descriptive statistics of the standardization variables

2006-PISA	Mean	Standard deviation
Responsibility for teacher hire: ratio of schools vs state	0.11	0.25
Responsibility for course content: ratio of schools vs state	0.45	0.40
Responsibility for salary increases: ratio of schools vs state	0.41	0.45
Responsibility for formulate budget: ratio of schools vs state	0.79	0.37
Responsibility for textbook use: ratio of schools vs state	0.36	0.42
<i>Notes:</i> Variable obtained at the school level (from the school principle). Shown mean and standard deviations are averaged country data. State means national education authority; school means teacher, principal, or school governing board. The higher the variable, the more standardization.		

Results

As expected, a weaker picture can be found in Table D2 which looks at the effect of standardization on student performance, than in Table 8 in Section 2.3 for differentiation. Here the first stage never has an F statistics of above 10 (the standard rule of thumb and no effect is found in the second stage for any specification for either administrative standardization or content related standardization.

The previous results on standardization often show negative effects of standardization. For instance, Hanushek, Link, and Woessmann (2011) who show, using a panel study, that decentralization of (administrative and content related) inputs has a positive effect on student performance in developed countries. Woessmann, Luedemann, Schuetz, and West (2007) find that the negative association between standardization and performance holds especially when schools are also held accountable for their outputs. Woessmann (2003) shows a positive association between content related standardization and student performance, and a mixed picture for administrative standardization.

Table D2: The effect of standardization on mathematics test scores

	Administrative standardization					Content related standardization				
	(0)	(1)	(2)	(3)	(5)	(0)	(1)	(2)	(3)	(5)
1st stage										
Distance		0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)		0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Empire			0.014 (0.086)		0.027 (0.091)			0.069 (0.047)		0.073 (0.051)
Pays réunis				0.003 (0.113)	0.007 (0.110)				0.113 (0.067)	0.125* (0.065)
Pays conquis				0.074 (0.119)	0.0811 (0.120)				-0.059 (0.071)	-0.041 (0.069)
Pays allies				0.052 (0.121)	0.045 (0.127)				0.034 (0.071)	0.015 (0.077)
2nd stage										
Admin	-23.230 (27.160)	-4,042.490 (76,790)	1,326.350 (7,492)	-36.130 (145.100)	73.580 (201.700)					
Content						12.6900 (34.4000)	366.200 (362.400)	378.000 (238.700)	131.500 (92.1500)	177.100* (100.500)
# of students	157,200	157,200	157,200	157,200	157,200	157,200	157,200	157,200	157,200	157,200
# of countries	29	29	29	29	29	29	29	29	29	29
R ² 2 nd stage	0.328			0.327	0.280	0.326	0.043	0.024	0.294	0.265
Partial R ² 1 st stage		0.00	0.00	0.02	0.02		0.02	0.06	0.18	0.22
F test 1 st stage		0.003	0.016	0.188	0.153		1.849	1.908	2.311	1.912
<i>Notes:</i> The table presents coefficients (standard errors in parenthesis) on the relation between standardization and student performance, instrumenting in models (2) to (5) tracking with the political pressure from Napoleon. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Model 2 is the preferred IV model. The models include imputation dummies and imputation variable interaction terms, and clustering is done on the country level. The control variables are as explained in the main text. Durbin and the Wu-Hausman tests are satisfied.										

Appendix E: Full second stage model of Table 9 (Chapter 2)

	Math	
	Coef.	SE
Grade of student: 7 under modal grade	ref	
6 under modal grade	-88.110***	(14.950)
5 under modal grade	11.580*	(6.962)
4 under modal grade	-81.600***	(7.384)
3 under modal grade	-187.300***	(22.140)
2 under modal grade	-133.200***	(17.700)
1 under modal grade	-73.400***	(16.810)
modal grade	-23.060*	(12.960)
1 above modal grade	-2.309	(13.790)
2 above modal grade	37.980**	(18.410)
3 above modal grade	-183.400***	(16.650)
4 above modal grade	-63.150***	(17.050)
Gender of student: Male	ref	
Female	-25.400***	(1.406)
Number of books in the household: 0-10 books	ref	
11-25 books	10.910***	(1.929)
26-100 books	29.220***	(2.963)
101-200 books	41.070***	(3.917)
201-500 books	55.590***	(4.190)
More than 500 books	53.230***	(4.668)
Track level student: General	ref	
Vocational	-43.080***	(9.652)
School average number of books	36.250***	(4.261)
Agglomeration school: Village	ref	
Small town	-0.001	(4.419)
Town	-3.052	(5.157)
City	-5.268	(4.812)
Large city	-12.640**	(5.984)
School size	0.014**	(0.006)
Ability grouping in the school: Not for any subjects	ref	
For some subjects	-5.201	(4.079)
For all subjects	-10.080**	(4.916)
GDP per capita in 2005	-0.001	(0.001)
Number of tracks	16.330**	(7.630)
Constant	√	
Imputation dummies and interactions	√	
Observations (# of students)	157,200	
Clusters (# of countries)	29	
R ² 2 nd stage	0.33	
Partial R ² 1 st stage	0.42	
F 1 st stage	27.61	
Durbin p	0.00	
Wu-Hausman p	0.00	

Notes: The table presents coefficients (standard errors in parenthesis) on the relation between tracking and student performance, instrumenting tracking with the political pressure from Napoleon. The first stage is not depicted but can be found in Table 9 of Chapter 2. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Appendix F: The exclusion restriction (Chapter 2)

Tables F1 to F4 display the results of similar analyses as in Section 2.2 but now with the GINI coefficient, the number of police officers per 100,000 inhabitants, the expenditures on education and the proportion of women seats in parliament. These are just four examples on which Napoleon is reasonably to be expected to have had an influence. None of these is consistently found to be explained by the political pressure from Napoleon, except for the GINI coefficient. To use the political pressure from Napoleon as an instrument for education systems, Napoleon should not have influenced anything else that influences student performance today. Since Napoleon at least influenced the GINI coefficient, this exclusion restriction is violated.

Table F1: The effect of political pressure from Napoleon on the GINI coefficient

	(1)	(2)	(3)	(4)	(5)
Distance	-0.003* (0.001)	-0.003* (0.001)	-0.001 (0.001)		-0.001 (0.001)
Empire		-0.042 (1.349)			0.123 (1.196)
Pays réunis			-0.959 (1.510)	-0.533 (1.325)	-0.937 (1.560)
Pays conquis			1.363 (1.414)	1.436 (1.389)	1.402 (1.496)
Pays allies			-4.652** (1.737)	-5.019*** (1.610)	-4.630** (1.793)
Constant	33.550*** (1.343)	33.560*** (1.460)	33.050*** (1.892)	32.120*** (1.129)	32.980*** (2.069)
# of countries	26	26	26	26	26
Adjusted R ²	0.114	0.0752	0.353	0.371	0.321
p F test	0.051	0.156	0.010	0.004	0.023

Notes: The table presents coefficients (standard errors in parenthesis) on the relation between political pressure from Napoleon and GINI coefficient in a country. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Model 2 is the preferred model.

Source: The GINI index is the most recent number (2008-1999) from the World Bank.

Table F2: The effect of political pressure from Napoleon on the number of police officers per 100,000 inhabitants

	(1)	(2)	(3)	(4)	(5)
Distance	-0.050 (0.033)	-0.052 (0.034)	-0.007 (0.041)		-0.006 (0.041)
Empire		27.040 (33.470)			30.540 (33.560)
Pays réunis			8.238 (40.450)	10.950 (36.640)	14.920 (41.260)
Pays conquis			-9.369 (40.790)	-8.641 (39.780)	-0.979 (41.960)
Pays allies			-97.130* (49.340)	-100.000** (45.560)	-94.330* (49.620)
Constant	364.500*** (35.370)	353.700*** (38.030)	343.300*** (51.450)	336.000** (29.580)	323.000** (56.240)
# of countries	29	29	29	29	29
Adjusted R ²	0.043	0.03	0.09	0.13	0.09
p F test	0.144	0.254	0.178	0.094	0.218
<p><i>Notes:</i> The table presents coefficients (standard errors in parenthesis) on the relation between political pressure from Napoleon and the number of police officers per 100,000 inhabitants in a country. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Model 2 is the preferred model.</p> <p><i>Source:</i> The number of police officers (per 100 000 inhabitants) for 2008 is from Eurostat.</p>					

Table F3: The effect of political pressure from Napoleon on the expenditure on education

	(1)	(2)	(3)	(4)	(5)
Distance	0.0001** (0.000)	0.001** (0.000)	0.001* (0.000)		0.001* (0.001)
Empire		-0.268 (0.378)			-0.295 (0.411)
Pays réunis			0.185 (0.492)	-0.155 (0.475)	0.120 (0.505)
Pays conquis			-0.176 (0.496)	-0.267 (0.515)	-0.257 (0.514)
Pays allies			0.0234 (0.600)	0.387 (0.590)	-0.004 (0.608)
Constant	4.359*** (0.398)	4.466*** (0.429)	4.272*** (0.626)	5.186*** (0.383)	4.467*** (0.689)
# of countries	29	29	29	29	29
Adjusted R ²	0.12	0.11	0.03	-0.06	0.01
p F test	0.034	0.087	0.331	0.693	0.411
<p><i>Notes:</i> The table presents coefficients (standard errors in parenthesis) on the relation between political pressure from Napoleon and the expenditures on education in a country. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Model 2 is the preferred model.</p> <p><i>Source:</i> The expenditures on education are from the World Bank (Public spending on education, total (% of GDP)) for 2005.</p>					

Table F4: The effect of political pressure from Napoleon on the number of women seats in parliament

	(1)	(2)	(3)	(4)	(5)
Distance	0.003 (0.004)	0.003 (0.004)	-0.001 (0.004)		-0.001 (0.004)
Empire		-0.813 (3.711)			-0.249 (3.530)
Pays réunis			4.502 (4.181)	4.702 (3.786)	4.448 (4.340)
Pays conquis			2.759 (4.216)	2.813 (4.110)	2.691 (4.414)
Pays allies			14.280*** (5.100)	14.060*** (4.708)	14.260** (5.219)
Constant	24.570*** (3.877)	24.900*** (4.217)	22.420*** (5.317)	21.890*** (3.057)	22.590*** (5.915)
# of countries	29	29	29	29	29
Adjusted R ²	-0.01	-0.05	0.15	0.18	0.11
p F test	0.440	0.729	0.101	0.048	0.179

Notes: The table presents coefficients (standard errors in parenthesis) on the relation between political pressure from Napoleon and the number of women seats in parliament in a country. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Model 2 is the preferred model.

Source: The proportion of seats held by women in national parliaments (%) for 2008 is from the World Bank.

Appendix G: Analyses using data from 1970 (Chapter 2)

This appendix shows analyses similar to those done in Section 2.4, but now using data on tracking and student performance from 1970 instead of 2006.

Data

Student and school level data

Student performance and student and school level controls are obtained in 1970 from the International Association for the Evaluation of Educational Achievement (IEA) First International Science Study (FISS). FISS is the second internationally comparable study, performed by the IEA.⁷¹ For 2006, as before, I use data from the Program for International Student Assessment (PISA).

71. In 1964 another study was held by the IEA but no information on standardization was collected and only 6 European countries participated.

For FISS 1970 data was collected in 19 countries and regions, of which ten are used in this paper. These ten countries, or regions, are Flemish Belgium, French Belgium, England, Finland, West-Germany, Hungary, Italy, the Netherlands, Scotland, and Sweden. Together this amounts to 25,409 students in 1033 schools. The more extended PISA 2006 sample used in the main text of Chapter 2 contains the ten countries used in FISS. In these ten countries there are 41,790 students in 1,671 schools who participated in PISA 2006. The age target of both data sets is quite similar: FISS 1970 were 14-year-old students, while PISA 2006 were 15-year-old students. Like in the main text, missing values are replaced by either school or country averages and the analyses are done with student weights and clustering of errors on country level. This clustering on country level implies that in the restricted sample only 10 observations exists.

Both FISS and PISA contain a science test score, which is in this paper standardized over all countries to be able to compare the results over time. So while in the main text, I focus on the mathematics test score, here I focus on science since FISS only contains a science test score. FISS and PISA also provide, on a limited scale, similar information on the student and the school. Both data sets contain data on the controls used in the main text: the gender of the student, the number of books in the household (as a proxy for social economic background), the grade of the student, the school size, the location of the school, the track type of the school or students, and whether the school has ability grouping.

Tracking

For tracking in 1970 I constructed one differentiation measure based upon national reports on the state of education from UNESCO in the early seventies (or late sixties). For differentiation in 1970 a combination of the number of tracks and age of selection is made. Table G1 shows differentiation in 2006 and 1970 for the sample of ten countries.

Some countries made changes in the level of differentiation in their education systems between 1970 and 2006 (see also Chapter 5). For instance, political opinion in England, Sweden and Finland in 1970 was in favor of a comprehensive system and thus in the years following 1970 these countries saw a move towards a comprehensive system. However, in 1970 Scotland was the only country which had already a comprehensive system in place for its 14-years-pupils. England, Sweden and Finland followed in the years after 1970. Another example of a country that changes its level of differentiation is Belgium which had a less differentiated system in 1970, with only a general and vocational track and selection at age 14, than in 2006, with four tracks and selection at age 12.

Table G1: Differentiation in the restricted sample

Country	Number of tracks in 2006	Age of first selection in 2006	Differentiation in 2006	Differentiation in 1970
Scotland	1	16	0	0
England	1	16	0	1
Sweden	1	16	0	1
Finland	1	16	0	1
Italy	3	14	2	3
Hungary	3	11	3	3
Flemish Belgium	4	14	2	2
The Netherlands	4	12	3	3
(West-)Germany	4	10	3	3
French Belgium	5	14	3	2

Source: Differentiation in 1970 is from the 1968 yearbook of education of UNESCO and the different national reports of UNESCO of 1968, 1971, 1973, and 1975. Differentiation in 2006 is a combination of the number of tracks in 2006 and the age of first selection in 2006. See Table 1 of Chapter 2 for further sources.

Results

Table G2 presents the results on differentiation using the FISS 1970 restricted sample and a comparable dataset of PISA 2006, containing the same 10 countries⁷². The full 2006 sample revealed a significant positive effect of the number of tracks on student performance and an insignificant negative effect of the age of first selection. The restricted 2006 sample shows a weaker relation between both the age of first selection and the number of tracks and student performance: The coefficient on the number of tracks is still positive, although insignificantly, but the coefficient on the age of first selection is no longer consistently negative (not shown, but available on request). Table G2 shows that using this restricted sample in both 1970 and 2006 countries that were conquered by France have more differentiation. And for the combined measure of differentiation there is no effect on student performance in 2006 (right panel) although the first stage is still strong, while there is a consistent negative effect in 1970 (left panel). The first stage shows similar results as before, with only a changing sign for *pays conquis*.

In 1970 one step up the differentiation measure leads to one-tenth to one quarter standard deviation lower science test score. Similarly to before, this is equivalent to a move from a household with no books to a household with 1-10 books (.265***).⁷³ Again, it should be kept in mind, that, due to the clustering, these models effectively only have ten observations.

72. West-Germany is Germany minus the states Berlin, Saxony-Anhalt, Brandenburg, Mecklenburg-Vorpommern, Saxony and Thuringia.

73. In 1970 the variable number of books in the household has five categories: none, 1-10 books, 11-25 books, 26-50 books, and 51 or more.

Table G2: The effect of tracking on science test scores using the FISS 1970 and restricted PISA 2006 sample

Model:	1970					2006				
	(0)	(1)	(2)	(3)	(4)	(0)	(1)	(2)	(3)	(4)
1st stage										
Distance		-0.001*** (0.000)	-0.001** (0.000)	0.000 (0.000)	-0.025*** (0.005)		-0.001*** (0.000)	-0.001*** (0.000)	-0.001** (0.000)	-0.001** (0.000)
Empire			0.632 (0.425)		54.840*** (10.440)			1.029 (0.659)		-0.955 (0.542)
Pays reunis				0.898* (0.422)	-84.270*** (16.010)				1.218* (0.546)	2.441*** (0.689)
Pays conquis				1.117*** (0.177)	33.150*** (6.060)				0.386 (0.505)	-0.195 (0.454)
Pays allies				-0.048 (0.174)	-19.630*** (3.739)				0.833 (0.455)	1.092** (0.393)
2nd stage										
Differentiation	-0.083* (0.045)	-0.231 (0.147)	-0.251** (0.114)	-0.101* (0.060)	-0.117** (0.050)	0.010 (0.062)	0.056 (0.136)	0.078 (0.116)	0.006 (0.071)	-0.009 (0.065)
# of students	25,409	25,409	25,409	25,409	25,409	73,171	73,171	73,171	73,171	73,171
# of countries	10	10	10	10	10	10	10	10	10	10
R ² 2 nd stage	0.230	0.215	0.211	0.229	0.229	0.353	0.351	0.348	0.353	0.353
Partial R ² 1 st stage		0.214	0.279	0.704	0.874		0.438	0.547	0.881	0.907
F test 1 st stage		13.74	5.219	43.12	46.45		40.75	11.22	126.6	108.1
<p><i>Notes:</i> Control variables are a constant, the gender of the student, the number of books in the household, the grade of the student compared to the modal grade in the country, the school size, the location of the school, the track type of the school or students, and whether the school has ability grouping. Durbin and the Wu-Hausman tests are satisfied, except for model 5 in 2006. Clustering on country level. Imputation dummies and interactions are included. Model 2 is the preferred model.</p>										

Appendix H: Nonlinear models (Chapter 3)

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

Table H1: Non-linear models looking at student performance

Dependent variable	Reading	Mathematics	Science
School considers prior performance			
<i>Sometimes</i>	-6.56 (2.53)	-4.69** (2.36)	-5.83*** (2.15)
<i>Always</i>	-4.97 (3.21)	-4.04 (3.77)	-5.47 (4.93)
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)	2.45*** (4.11)
Always*4-5 tracks	21.79*** (5.48)	21.92*** (5.14)	26.74*** (6.28)
(pseudo)LL	-43755	-43732	-44088
# of students	187768	187,768	187,768
# of countries	31	31	31
<i>Notes:</i> Coefficients with standard errors in parenthesis. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. See also notes of Table 3 of Chapter 3.			

Table H2 shows similar models to those in the first 3 columns of Table 6 but also here including the number of tracks in a country as three dummy variables. The interactions between parental background and the number of tracks are significantly negative. This means that the relation between parental background and student performance is lower in countries with more tracks. Also this result is similar to those presented in the main text.

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

Dependent variable	Reading	Mathematics	Science
School considers prior performance			
<i>Sometimes</i>	-1.68 (2.02)	-1.08 (1.88)	-0.62 (2.08)
<i>Always</i>	6.87* (3.28)	5.76 (2.98)	8.24* (3.67)
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)
Parental background* 2-3 tracks	-12.50*** (2.34)	-12.66*** (2.56)	-14.18*** (2.55)
Parental background* 4-5 tracks	-10.90*** (3.29)	-10.59*** (3.30)	-11.33*** (3.40)
(pseudo-)LL	-43742	-43719	-44074
# of students	187768	187,768	187,768
# of countries	31	31	31
<i>Notes:</i> Coefficients with standard errors in parenthesis. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. See also notes of Table 3 of Chapter 3.			

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

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

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

	(1) Between countries, between schools	(2) Within country	(3) Between countries
School considers prior performance			
<i>Sometimes</i>	-5.97** (2.54)	-4.37** (2.17)	35.24 (39.05)
<i>Always</i>	-10.13** (4.11)	-2.67 (4.10)	-66.67** (30.42)
Number of tracks (0-4)	-5.74** (2.70)	-	5.47 (12.54)
Sometimes*Number of tracks	2.83** (1.15)	2.54** (1.05)	-36.50 (23.61)
Always*Number of tracks	8.64*** (1.82)	6.03*** (1.96)	19.95 (18.06)
Country FE	-	√	-
(pseudo)LL	-43759	-43727	-43758
# of students	187768	187768	187768
# of countries	31	31	31
<i>Notes:</i> Coefficients with standard errors in parenthesis. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Column (1) shows the main model as depicted in column (2) of Table 4. Column (3) measures the school variables “school principal consider prior performance” on a national level and thus depicts the proportion of schools (between 0 and 1) in the country with school which say they always or sometimes consider prior performance. See also notes of Table 3 of Chapter 3.			

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

	(1) Between countries, between schools	(2) Within country	(3) Between countries
School considers prior performance			
<i>Sometimes</i>	-5.99** (2.44)	-5.40** (2.20)	78.61 (55.61)
<i>Always</i>	-10.90** (5.06)	-5.07 (4.88)	-86.87* (46.45)
Number of tracks (0-4)	-2.89 (3.31)	-	6.40 (16.08)
Sometimes*Number of tracks	3.45*** (1.08)	3.33*** (0.99)	-42.89 (32.60)
Always*Number of tracks	9.68*** (2.04)	7.39*** (2.08)	34.62 (24.08)
Country FE	-	√	-
(pseudo)LL	-44092	-44054	-44092
# of students	187768	187,768	187,768
# of countries	31	31	31
<i>Notes:</i> Coefficients with standard errors in parenthesis. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Column (1) shows the main model as depicted in column (2) of Table 4. Column (3) measures the school variables “school principal consider prior performance” on a national level and thus depicts the proportion of schools (between 0 and 1) in the country with school which say they always or sometimes consider prior performance. See also notes of Table 3 of Chapter 3.			

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

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

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

	(1) Reading	(2) Math	(3) Science
School considers entrance requirements			
<i>Sometimes</i>	-6.86*** (2.20)	-4.51*** (2.16)	-5.37*** (2.01)
<i>Always</i>	-7.61 (4.78)	-5.84 (5.25)	-6.34 (6.34)
Number of tracks (0-4)	-7.85** (3.71)	-7.94** (3.96)	-5.15 (4.64)
Sometimes*Number of tracks	3.32** (1.62)	2.78 (1.90)	2.17 (1.77)
Always*Number of tracks	6.31** (2.65)	3.63 (2.25)	5.01* (2.76)
(Pseudo) LL	-28,138	-28,132	-28,358
# of students	131,423	131,423	131,423
# of countries	14	14	14
<i>Notes:</i> The table presents coefficients from random effects models (standard errors in parenthesis) on the relation between student performance and whether or not schools consider prior performance when selecting students and the number of tracks in a country. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The control variables are as described in the text. The models include imputation dummies and imputation variable interaction terms.			

Appendix K: Outcome variables (Chapter 4)

<i>Grade 7 and 9</i>	
IQ 1	Number of IQ questions correct as percentage of total number of questions.
IQ 2	Number of IQ questions correct as percentage of total questions completed.
Openness	<ul style="list-style-type: none"> I use difficult words I have lots of ideas I learn things quickly I have little imagination
Competitive spirit	<ul style="list-style-type: none"> I would like to get high marks Later I want to be good at my job
Conscientiousness	<ul style="list-style-type: none"> I do my chores immediately I often leave my stuff hanging around I always stick to my appointments I sometimes forget I have to do something I am very precise in what I do
Extraversion	<ul style="list-style-type: none"> I talk a lot (negatively formulated in grade 7) I am quite among strangers I am the pacesetter at parties I like to be around lots of people
Agreeableness	<ul style="list-style-type: none"> I try to help people I am interested in others (negatively formulated in grade 7) I empathize with others I am a friendly person
Neurotism	<ul style="list-style-type: none"> I easy get upset I am often stressed My temper shifts often I regularly have a gloomy mood
Perseverance	<ul style="list-style-type: none"> I continue until it is done I stop easily if it gets too difficult If I start something, I finish it If something is harder than expected, I soon loose heart
Social skills	
Social skills (action)	<ul style="list-style-type: none"> Own appreciation for: drawing, painting or making music Own appreciation for: looking for something on the computer Own appreciation for: writing without mistakes Own appreciation for: mental arithmetic

Appendices

Social skills (social)	Own appreciation for: own ability to concentrate
	Own appreciation for: to choose nice clothes and look good
	Own appreciation for: own ability to discuss
	Own appreciation for: own ability to comfort someone
	Own appreciation for: own ability to give your opinion
	Own appreciation for: own ability to win an argument
	Own appreciation for: own ability to get my way
	Own appreciation for: own ability to interact with other students

School wellbeing

School wellbeing student

I like going to this school
I hate this school
I like it at this school
I am bored at this school

School wellbeing teachers

The teachers like me
If I want the help of a teacher, I also receive it
The teacher think I am smart
The teacher do their very best for me

School wellbeing classmates

I have here many friends
Some students bully me
The students here like me

Estimated probability of completing secondary school

What is your estimation that you will finished your current degree?/Do you think you will manage in your next school?

Only in grade 9

Estimated probability of obtaining a job 1

How large do you think the probability is that you can find a job easily if you finish your current degree?

Estimated probability of obtaining a job 2

How large do you think the probability is that you can find a job easily if you finish your next degree?

Civic engagement: Democratic behavior

Test score

Civic engagement: International mind set

Test score

School motivation

I will drop out without finishing school
As soon as possible, I'll stop learning
I will learn a profession, but outside school
I am very motivated to continue learning
I am going to learn interesting things
I am going to continue learning because I like it
I am going to continue learning for a very long time

As soon as I can get a job, I will drop out of school

School satisfaction

- I feel secure at school
- The atmosphere at school is nice
- The teacher treat me with respect
- There are clear rules at school
- The variation in the teaching styles is large
- The teacher explain things well
- I learn a lot from the teachers
- The teachers take into account what I can and cannot do
- I am sufficiently challenged to do my best in school
- I am satisfied about my mentor
- I know who to turn to at school in case of problems
- If needed, I receive extra tutoring
- I am assisted in making important choices regarding my studies
- Teachers clearly tell me how my results are
- I am informed about things relevant for students
- The opinion of students counts at this school

Math test score 1

Test score

Math test score 2

Test score

Reading test score 1

Test score

Reading test score 2

Test score

Appendix L: Other model specifications (Chapter 4)

Supplementing the analyses done in Chapter 4, in this Appendix we use different model specifications, where we not use not both the elementary school exit test score and the elementary school teacher recommendation but either one (model IV-1 and IV-2 in the tables below), or we add an interaction between the two in the second stage (model IV-5) and an interaction between the two indicator functions in the first stage (model IV-4). The results are very robust to model specification, although the F statistics varies over the models.

Table L1: The effect of being in the high track on IQ in grade 9

Model:	OLS	IV-1	IV-2	MAIN IV	IV-4	IV-5
High Track	0.195* (0.101)	0.794 (0.485)	0.638*** (0.205)	0.444* (0.229)	0.445* (0.229)	0.192 (0.388)
Test score	0.039*** (0.010)	0.020 (0.022)		0.032*** (0.010)	0.032*** (0.010)	-0.041 (0.081)
Recommendation	0.045* (0.023)		0.063* (0.033)	0.023 (0.025)	0.023 (0.025)	-2.463 (2.790)
Test score* Rec.						0.005 (0.005)
Constant	- 21.920*** (5.038)	-11.020 (11.720)	-1.312*** (0.481)	- 17.790*** (5.476)	- 17.770*** (5.456)	20.790 (42.900)
# of students	1,175	1,175	1,175	1,175	1,175	1,175
# of schools	17	17	17	17	17	17
R ²	0.13	0.07	0.09	0.11	0.12	0.13
F excl. instr.	-	18.57	9.421	15.25	10.44	3.258
<i>Notes:</i> The table presents coefficients (robust standard errors in parentheses) from OLS and IV models using different specification depicting the effect of being in the higher track on IQ. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.						

Table L2: The effect of being in the high track on reading in grade 9

Model:	OLS	IV-1	IV-2	MAIN IV	IV-4	IV-5
High Track	0.098 (0.085)	0.950*** (0.266)	1.022*** (0.357)	0.859** (0.353)	0.867** (0.355)	0.314 (0.519)
Test score	0.044*** (0.005)	0.014 (0.011)		0.023** (0.009)	0.022** (0.009)	-0.146 (0.093)
Recommendation	0.041** (0.018)		0.010 (0.042)	-0.016 (0.027)	-0.016 (0.027)	-5.787* (3.313)
Test score* Rec.						0.011* (0.006)
Constant	-24.230*** (2.940)	-7.790 (6.073)	-0.274 (0.549)	-12.050** (5.181)	-11.920** (5.252)	77.920 (49.930)
# of students	1,927	1,927	1,927	1,927	1,927	1,927
# of schools	22	22	22	22	22	22
R ²	0.13			0.01	0.01	0.13
F excl. instr.	-	17.26	20.64	17.11	13.25	2.126

Notes: The table presents coefficients (robust standard errors in parentheses) from OLS and IV models using different specification depicting the effect of being in the higher track on reading scores. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table L3: The effect of being in the high track on the track in grade 9

Model:	OLS	IV-1	IV-2	MAIN IV	IV-4	IV-5
High Track	0.254** (0.0900)	0.747*** (0.185)	0.683*** (0.179)	0.482*** (0.162)	0.471*** (0.164)	0.505** (0.243)
Test score	0.0353*** (0.00641)	0.0491*** (0.0103)		0.0290*** (0.00606)	0.0293*** (0.00609)	0.0342 (0.0577)
Recommendation	0.144*** (0.0234)		0.159*** (0.0337)	0.126*** (0.0275)	0.127*** (0.0276)	0.312 (1.998)
Test score * Rec.						-0.000349 (0.00374)
Constant	-16.26*** (3.390)	-21.51*** (5.539)	2.498*** (0.488)	-12.59*** (3.346)	-12.76*** (3.366)	-15.41 (30.83)
# of students	2,866	2,866	2,866	2,866	2,866	2,866
# of schools	0.432	0.333	0.368	0.417	0.418	0.414
R ²	22	22	22	22	22	22
F excl. instr.	-	20.43	23.34	19.64	16.30	2.054

Notes: The table presents coefficients (robust standard errors in parentheses) from OLS and IV models using different specification depicting the effect of being in the higher track on the track in grade 9. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table L4: The effect of being in the high track on the self-perceived probability of obtaining the degree in grade 9

Model:	OLS	IV-1	IV-2	MAIN IV	IV-4	IV-5
High Track	-0.0438 (0.0744)	0.949** (0.411)	0.496** (0.228)	0.601** (0.248)	0.603** (0.248)	0.420 (0.451)
Test score	0.0134 (0.00862)	-0.0314 (0.0237)		-0.00558 (0.01000)	-0.00564 (0.00998)	-0.0586 (0.111)
Recommendation	0.0367 (0.0326)		-0.0208 (0.0402)	-0.0256 (0.0393)	-0.0258 (0.0394)	-1.865 (3.852)
Test score * Rec.						0.00345 (0.00722)
Constant	-3.567 (4.283)	20.88* (12.63)	4.400*** (0.555)	7.460 (5.324)	7.489 (5.316)	35.76 (59.16)
# of students	994	994	994	994	994	994
# of schools	16	16	16	16	16	16
R ²	0.017					
F excl. instr.	-	17.35	8.203	11.42	7.769	4.071

Notes: The table presents coefficients (robust standard errors in parentheses) from OLS and IV models using different specification depicting the effect of being in the higher track on the self-perceived probability to obtain the degree. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Appendix M: Including controls for gender and parental background (Chapter 4)

Chapter 4 present OLS, RD, and RD panel results without any controls (Tables 5 to 9). In this appendix controls are added for parental background and gender. Table M1 shows the first stage using the IQ sample, while Tables M2 and M3 show the second stage for the models looking at IQ, the reading score, the track in grade 9 and the self-perceived probability to obtain the degree, which are the four outcomes for which we find significant effects using IV. The results presented here are very similar to those in Tables 5 to 9 in the main text.

Table M1: First stage results controlling for gender and parental education

Controlled for: Dependent variable:	Gender		Parental education	
	High track	High track	High track	High track
I(Test score \geq 544)	0.14*** (0.04)	0.14*** (0.04)	0.14** (0.05)	0.14** (0.05)
I(Recommendation \geq 18)	0.34** (0.15)	0.33** (0.15)	0.31* (0.15)	0.31* (0.15)
Test score	0.01* (0.00)	0.01* (0.00)	0.01*** (0.00)	0.01*** (0.00)
Recommendation	0.06** (0.02)	0.06** (0.02)	0.05** (0.02)	0.05** (0.02)
Grade 6 IQ		0.00 (0.01)		0.00 (0.01)
Control	-0.02 (0.02)	-0.02 (0.02)	0.01 (0.02)	0.01 (0.02)
Constant	-4.28** (1.80)	-4.13** (1.77)	-6.28*** (1.59)	-6.14*** (1.80)
# of students	1,175	1,175	923	923
# of schools	17	17	17	17
R ²	0.44	0.44	0.43	0.43

Notes: The table presents first stage coefficients (robust standard errors in parentheses) from IV models using the IQ sample, controlling for either gender or parental education. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table M2: The effect of being in the high track on outcomes, controlling for parental education.

	IQ			Self-perceived probability to obtain the degree			Reading		Track grade 9	
	OLS	IV	Panel IV	OLS	IV	Panel IV	OLS	IV	OLS	IV
High track	0.20 (0.12)	0.44* (0.24)	0.49* (0.26)	-0.081 (0.08)	0.48** (0.23)	0.45** (0.22)	0.089 (0.10)	0.84* (0.44)	0.24** (0.09)	0.44** (0.19)
Test score	0.04*** (0.01)	0.03*** (0.01)	0.01 (0.01)	0.01 (0.01)	-0.00 (0.01)	-0.00 (0.01)	0.05*** (0.01)	0.03** (0.01)	0.04*** (0.01)	0.03*** (0.01)
Recommendation	0.05* (0.02)	0.03 (0.03)	0.02 (0.03)	0.03 (0.03)	-0.02 (0.03)	-0.02 (0.03)	0.05* (0.02)	-0.01 (0.04)	0.12*** (0.02)	0.11*** (0.03)
Grade 6 variable			0.28*** (0.04)			0.14*** (0.04)				
Parental education	0.04 (0.03)	0.03 (0.02)	0.03 (0.02)	0.08*** (0.02)	0.07*** (0.03)	0.07*** (0.03)	0.02 (0.02)	-0.01 (0.03)	0.07*** (0.01)	0.07*** (0.01)
Constant	-20.62*** (4.99)	-16.36*** (5.22)	-6.17 (5.81)	-3.82 (3.91)	6.19 (5.15)	5.35 (5.08)	-25.93*** (3.37)	-13.88** (6.67)	-16.85*** (3.81)	-13.62*** (4.08)
# of students	923	923	923	947	947	947	1,580	1,580	2,186	2,186
# of schools	17	17	17	16	16	16	22	22	22	22
R ²	0.12	0.12	0.18	0.03			0.14	0.03	0.43	0.41
F excl instr.		9.74	9.71		12.54	11.81		10.34		12.17

Notes: The table presents coefficients (robust standard errors in parentheses) from OLS, IV and panel IV models depicting the effect of being in the higher track on outcomes, controlling for parental education. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table M3: The effect of being in the high track on outcomes, controlling for gender.

	IQ			Self-perceived probability to obtain the degree			Reading		Track grade 9	
	OLS	IV	Panel IV	OLS	IV	Panel IV	OLS	IV	OLS	IV
High track	0.20*	0.44*	0.47**	-0.04	0.60**	0.57**	0.12	0.82**	0.27***	0.47***
	(0.10)	(0.23)	(0.22)	(0.07)	(0.24)	(0.23)	(0.09)	(0.37)	(0.09)	(0.15)
Test score	0.04***	0.03***	0.02*	0.01	-0.00	-0.00	0.04***	0.03***	0.04***	0.03***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Recommendation	0.04*	0.02	0.02	0.04	-0.03	-0.03	0.04*	-0.01	0.14***	0.12***
	(0.02)	(0.03)	(0.02)	(0.03)	(0.04)	(0.04)	(0.02)	(0.03)	(0.02)	(0.03)
Grade 6 variable			0.25***			0.16***				
			(0.03)			(0.04)				
Gender	0.05	0.05	0.03	0.05	0.07	0.07	0.32***	0.34***	0.15***	0.16***
	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.03)	(0.03)
Constant	-22.12***	-18.09***	-9.11*	-3.85	7.05	6.06	-24.75***	-13.57**	-16.92***	-13.56***
	(4.95)	(5.41)	(4.81)	(4.22)	(5.17)	(5.05)	(3.17)	(5.39)	(3.40)	(3.26)
# of students	1,175	1,175	1,175	994	994	994	1,896	1,896	2,796	2,796
# of schools	17	17	17	16	16	16	22	22	22	22
R ²	0.13	0.12	0.17	0.02			0.16	0.07	0.44	0.43
F excl instr.		15.63	15.55		11.83	11.27		18.56		21.63

Notes: The table presents coefficients (robust standard errors in parentheses) from OLS, IV and panel IV models depicting the effect of being in the higher track on outcomes, controlling for gender. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Appendix N: Representative sample (Chapter 5)

The main sample used in Chapter 5 is a subsample from a large representative sample from Dutch secondary school students. Since a large number of observations have missing values on crucial variables, the main sample contains only about thirty percent of the original sample. To still ensure representativeness, weights are employed in the analyses in this chapter. The weights are constructed using the representative sample and are based upon the education of the parents, gender, and two school composition variables: the school average of parental education and the school immigrant percentage. Table N1 shows some descriptive statistics for the full representative sample, the restricted sample used in this chapter and the weighted version of the restricted sample. The last column shows the difference in the mean between the two restricted sample and the full representative sample.

Table N1: Descriptive statistics of the different samples

	Representative sample			Obs	Sample			Sample (weighted)		
	Obs	Mean	Std. Dev.		Mean	Std. Dev.	Dif in mean with rep. sample	Mean	Std. Dev.	Dif in mean with rep. sample
Parental education	9076	3.17	0.81	2602	2.95	0.91	0.22***	3.20	0.79	-0.03
Gender	9293	1.51	0.5	2606	1.5	0.50	0.01	1.51	0.50	0.00
Month of Birth	9252	6.55	3.42	2588	6.6	3.38	-0.04	6.61	3.40	-0.05
Birth year	9252	1998.61	0.55	2588	1998.59	0.54	0.02	1998.62	0.53	-0.01
Immigration status	7887	0.04	0.19	2253	0.07	0.26	-0.03***	0.03	0.17	0.01*
School size	9444	15.84	12.31	2621	14.18	10.53	1.66***	15.2	12.18	0.64**
School parental education	9271	3.16	0.4	2621	2.92	0.47	0.24***	3.15	0.40	0.02**
School immigrant percentage	8978	0.06	0.2	2586	0.12	0.28	-0.05***	0.06	0.19	0.01

Notes: The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Appendix O: Results without weights (Chapter 5)

Table O1 and O2 present the same models as Table 3 and 4 of Chapter 5 but now not using the weights. The results are qualitatively the same as described in the main text.

Table O1: Parental background effects on the elementary school exit test (without weights)

	Math exit test score			Reading exit test score		
	(1)	(2)	(3)	(4)	(5)	(6)
Parent: at most lower vocational education	2.33** (0.983)	0.25 (0.61)	-0.11 (0.60)	4.11*** (0.98)	2.16*** (0.65)	1.41** (0.68)
Parent: at most vocational education	4.72*** (0.98)	-0.19 (0.62)	-1.03* (0.62)	7.87*** (0.99)	3.42*** (0.71)	1.70** (0.75)
Parent: at most tertiary education	10.25*** (1.02)	1.19* (0.67)	-0.36 (0.69)	13.25*** (1.08)	4.75*** (0.73)	1.47** (0.74)
grade 6 test		9.13*** (0.19)	10.69*** (0.30)		8.26*** (0.31)	11.45*** (0.43)
Constant	36.41*** (0.97)	41.64*** (0.61)	42.54*** (0.63)	65.92*** (1.00)	70.60*** (0.64)	72.41*** (0.66)
# of students	2,621	2,621	2,621	2,398	2,398	2,398
# of schools	156	156	156	149	149	149
R ²	0.10	0.70	0.68	0.12	0.54	0.47
F excl. instr.			559.3			819.4

Notes: The table presents coefficients from OLS and IV models (robust standard errors in parenthesis) on the relation between student performance on the primary exit test score and parental background and early ability. Estimates in columns (3) and (6) are from a two stage least square model with as excluded instrument for the test score in grade 6 the test score in grade 3. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The standard errors are clustered on the grade 6 schools, and no weights are used.

Table O2: Parental background effects on the elementary school teacher recommendation (without weights)

	(1) o. logit	(2) o. logit	(3) o. logit	(4) IV
Parent: at most lower vocational education	1.93*** (0.31)	1.52* (0.38)	1.29 (0.35)	1.19 (0.14)
Parent: at most vocational education	3.45*** (0.58)	1.73** (0.48)	1.41 (0.41)	1.37*** (0.17)
Parent: at most tertiary education	10.87*** (2.10)	3.64*** (1.10)	2.57*** (0.79)	1.47*** (0.20)
Math test grade 6		6.29*** (0.68)	3.14*** (0.48)	2.56*** (0.51)
Reading test grade 6		4.51*** (0.38)	2.30*** (0.23)	2.33*** (0.34)
Elementary exit test: math			1.08*** (0.01)	1.00 (0.01)
Elementary exit test: reading			1.13*** (0.01)	1.02*** (0.01)
# of students	2,452	2,452	2,452	2,307
# of schools	152	152	152	149
(Pseudo) R ²	0.06	0.41	0.49	0.70

Notes: The table presents odds ratios from ordered logit models (first three columns) and adjusted coefficients from the second stage of an IV model (robust standard errors in parenthesis) on the relation between elementary school teacher recommendation and parental background and early ability. The IV model uses a continuous version of the teacher recommendation based upon the ordered logit cut points, see the text for more information, and the IV coefficient are, for comparison reasons, displayed as exp(b), similar like the ordered logit coefficients. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The standard errors are clustered on the grade 6 schools, no weights are used, and the constants are omitted from the table. The third from last row states: Being from the highest PB group bridges 13 percent (column 3) of the distance between the highest and the lowest track, which is calculated using ordered logit coefficients.

Appendix P: First stage results from Table 2 and 3 (Chapter 5)

Table P1 presents the first stage results related to column (6) from Table 3 and column (4) from Table 4.

Table P1: First stage models related to Tables 3 and 4 from Chapter 5.

Dependent Variable:	(1) Math score grade 6	(2) Reading score grade 6	(3) Math score grade 6	(4) Reading score grade 6
Parent: at most lower vocational education	0.08 (0.09)	0.08 (0.09)	0.98 (0.05)	0.95 (0.05)
Parent: at most vocational education	0.21** (0.10)	0.33*** (0.08)	1.07 (0.06)	1.07 (0.06)
Parent: at most tertiary education	0.44*** (0.10)	0.57*** (0.09)	1.11 (0.07)	1.13** (0.06)
Grade 6 test score: math	0.64*** (0.03)		1.25*** (0.03)	1.01 (0.02)
Grade 6 test score: reading		0.51*** (0.03)	1.01 (0.02)	1.25*** (0.04)
Elementary exit test score: math			1.06*** (0.00)	1.02*** (0.00)
Elementary exit test score: reading			1.01*** (0.00)	1.04*** (0.00)
Constant	-0.24** (0.10)	-0.30*** (0.08)	0.05*** (0.01)	0.03*** (0.00)
# of students	2621	2398	2466	2398
# of schools	156	149	153	149
R ²	0.47	0.35	0.71	0.56

Notes: The table presents first stage coefficients from IV models (robust standard errors in parenthesis). Column (1) is the first stage from columns (3) in Table 3, column (2) is the first stage of column (6) from Table 3 and columns (3) and (4) are the first stages from column (4) from Table 4. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The standard errors are clustered on the grade 6 schools, and student weights are used.

Appendix Q: Boys versus girls (Chapter 5)

In Chapter 5 I find a positive effect of parental background on the elementary school exit test score for reading and on the elementary school teacher track recommendation. Table P1 presents the results for boys and girls separately relating to column (6) from Table 3 and column (3) from Table 4. Like mentioned in the main text of Chapter 5, I find that the positive effect on the elementary school exit test score for reading is mainly due to the girls, while for the elementary school teacher track recommendation it is mainly due to the boys. Channels for this difference are outside the scope of this thesis.

Table Q1: Results for boys and girls separately.

Dependent Variable: Sample:	Reading exit test score		Track recommendation	
	Boys	Girls	Boys	Girls
Parent: at most lower vocational education	0.82 (0.94)	2.10* (1.24)	0.49 (0.51)	0.04 (0.29)
Parent: at most vocational education	1.71 (1.12)	1.72 (1.41)	0.90* (0.51)	0.47 (0.29)
Parent: at most tertiary education	1.73 (1.20)	2.54* (1.49)	1.53*** (0.53)	1.09*** (0.34)
Grade 6 test score: math			1.13*** (0.17)	0.86*** (0.23)
Grade 6 test score: reading	11.35*** (0.83)	10.77*** (0.65)	0.83*** (0.16)	0.72*** (0.15)
Elementary exit test score: math			0.08*** (0.02)	0.11*** (0.02)
Elementary exit test score: reading			0.14*** (0.01)	0.13*** (0.01)
# of students	1,182	1,215	1,222	1,229
# of schools	147	145	149	147
(Pseudo) R ²	0.364	0.475	0.3043	0.381

Notes: The table presents coefficients (robust standard errors in parenthesis) from IV models in columns (1) and (2) and odds ratios from ordered logit models in columns (3) and (4). The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The constants are omitted from the table. See respective tables for more notes.

Biography

Roxanne Korthals obtained a bachelor and master in economics at the University of Amsterdam (the Netherlands) and a research master in economics at the European University Institute (Florence, Italy). She started her PhD at the Research Center for the Labor Market and the Economics department at Maastricht University in September 2011. Her thesis is part of the four year research program *Educational systems and the four central functions of education* funded by the Dutch Science Foundation (NWO). Roxanne started a post-doc position at the Economics department at Maastricht University on April 1st, 2015.

ROA Dissertation Series

1. Lex Borghans (1993), *Educational Choice and Labour Market Information*, Maastricht, Research Centre for Education and the Labour Market.
2. Frank Cörvers (1999), *The Impact of Human Capital on International Competitiveness and Trade Performance of Manufacturing Sectors*, Maastricht, Research Centre for Education and the Labour Market.
3. Ben Kriechel (2003), *Heterogeneity Among Displaced Workers*, Maastricht, Research Centre for Education and the Labour Market.
4. Arnaud Dupuy (2004), *Assignment and Substitution in the Labour Market*, Maastricht, Research Centre for Education and the Labour Market.
5. Wendy Smits (2005), *The Quality of Apprenticeship Training, Conflicting Interests of Firms and Apprentices*, Maastricht, Research Centre for Education and the Labour Market.
6. Judith Semeijn (2005), *Academic Competences and Labour Market Entry: Studies Among Dutch Graduates*, Maastricht, Research Centre for Education and the Labour Market.
7. Jasper van Loo (2005), *Training, Labor Market Outcomes and Self-Management*, Maastricht, Research Centre for Education and the Labour Market.
8. Christoph Meng (2005), *Discipline-Specific or Academic? Acquisition, Role and Value of Higher Education Competencies*, Maastricht, Research Centre for Education and the Labour Market.
9. Andreas Ammermüller (2007), *Institutional Effects in the Production of Education: Evidence from European Schooling Systems*, Maastricht, Research Centre for Education and the Labour Market.
10. Bart Golsteyn (2007), *The Ability to Invest in Human Capital*, Maastricht, Research Centre for Education and the Labour Market.
11. Raymond Montizaan (2010), *Pension Rights, human capital development and well-being*, Maastricht, Research Centre for Education and the Labour Market.

12. Annemarie Nelen (2012), *Part-Time Employment and Human Capital Development*, Maastricht, Research Centre for Education and the Labour Market.
13. Jan Sauermann (2013), *Human Capital, Incentives, and Performance Outcomes*, Maastricht, Research Centre for Education and the Labour Market.
14. Harald Ulrich Pfeifer (2013), *Empirical Investigations of Costs and Benefits of Vocational Education and Training*, Maastricht, Research Centre for Education and the Labour Market.
15. Charlotte Büchner (2013), *Social Background, Educational Attainment and Labor Market Integration: An Exploration of Underlying Processes and Dynamics*, Maastricht, Research Centre for Education and the Labour Market.
16. Martin Humburg (2014), *Skills and the Employability of University Graduates*, Maastricht, Research Centre for Education and the Labour Market.
17. Jan Feld (2014), *Making the Invisible Visible, Essays on Overconfidence, Discrimination and Peer Effects*, Maastricht, Research Centre for Education and the Labour Market.
18. Olga Skriabikova (2014), *Preferences, Institutions, and Economic Outcomes: an Empirical Investigation*, Maastricht, Research Centre for Education and the Labour Market.
19. Gabriele Marconi (2015), *Higher Education in the National and Global Economy*, Maastricht, Research Centre for Education and the Labour Market.
20. Nicolas Salamanca Acosta (2015), *Economic Preferences and Financial Risk-Taking*, Maastricht, Research Centre for Education and the Labour Market.
21. Ahmed Elsayed Mohamed (2015), *Essays on Working Hours*, Maastricht, Research Centre for Education and the Labour Market.
22. Roxanne Amanda Korthals (2015), *Tracking Students in Secondary Education: Consequences for Student Performance and Inequality*, Maastricht, Research Centre for Education and the Labour Market.