ESSAYS ON THE PRODUCTION OF HUMAN CAPITAL

Sander Gerritsen

ESSAYS ON THE PRODUCTION OF HUMAN CAPITAL

Essays over de productie van menselijk kapitaal

ISBN: 978-94-6108-747-8

Print & Lay-out: Gildeprint, Enschede, The Netherlands

ESSAYS ON THE PRODUCTION OF HUMAN CAPITAL

Essays over de productie van menselijk kapitaal

Proefschrift

ter verkrijging van de graad van doctor aan de Erasmus Universiteit Rotterdam
op gezag van de rector magnificus

Prof.dr. H.A.P. Pols

en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op

vrijdag 26 september 2014 om 11.30 uur

Sander Bernard Gerritsen
geboren te Amsterdam

ERASMUS UNIVERSITEIT ROTTERDAM

Promotiecommissie

Promotor

Prof.dr. H.D. Webbink

Overige leden

Prof.dr. H. Oosterbeek

Prof.dr. B.J. ter Weel

Dr. E.M. Bosker

Preface

My interest in doing research started in 2008, when I did my internship at the CPB, The Netherlands Bureau for Economic Policy Analysis. During this internship, which was part of my Master thesis in econometrics, I investigated the effect of early cannabis use on educational attainment. Using a dataset on Australian twins who differed in their use of cannabis, I estimated – perhaps not surprisingly – a negative effect. This master thesis made me realize that I liked doing research. After finishing my masters, I started working at the education department of the CPB in 2009, under the supervision of Dinand Webbink, who was program leader at that moment. I couldn't have had a better supervisor. He introduced me to a novel trend in the economics field I wasn't aware of at that time, and which had not been given much attention in my econometrics courses: the (quasi-)experimental literature, in which treatment evaluations play an important role. This type of research, with its focus on finding causal effects, strongly appealed to me, and during my first 2 years of employment at the CPB I worked with Dinand on numerous projects, making myself comfortable with terms like differences-in-differences, regression discontinuity and counterfactuals. The course economics of education taught by Erik Plug and Hessel Oosterbeek further fueled my interest writing a PhD. When Dinand left the CPB in 2011 to become a professor at the Erasmus University, and asked whether I was interested in doing a PhD, the choice was easy. From April 2011 onwards, I have been working on my thesis. And what can I say about writing a thesis? Of course, it is hard work, but also that it is a great learning process. I not only learned a great deal about causal impact evaluations, I also learned how to make papers suitable for scientific publication. Also, I learned that patience is a virtue: it requires hard work and a lot of time to get something published. But I am sure it all will be worth it and pay off in the end.

This thesis could not have been written without the help and support of numerous persons. First of all, I would like to thank Dinand for having asked me to be his PhD-student, and having faith in me. I really enjoyed all the meetings and fruitful discussions we had, the ideas we exchanged, his enthusiasm, but most of all his sense of humor. I also would like to thank Erik Plug, coauthor of one of my papers. The meetings we had helped to improve the paper considerably. Secondly, I am indebted to my employer, the CPB, which gave me the opportunity to write this thesis. I would like to thank Coen Teulings, Casper van Ewijk, George Gelauff and Ruud Okker. A special thanks goes to Debby Lanser and Bas ter Weel for their support and valuable comments on some of my papers. In addition to these colleagues, I would like to thank Adam Elbourne and Rob Luginbuhl for checking my English. Also I am grateful to other colleagues for the great work atmosphere. In particular I would like to thank my old roomy Suzanne Kok and Ryanne van Dalen for having fun during work and after. The spinning days on Tuesday evenings kept me in shape. Furthermore I am grateful to the members of the PhD committee, Hessel Oosterbeek, Robert Dur, Bas Jacobs and Maarten Bosker for their time and effort.

Last but not least I would like to thank family and friends who indirectly contributed to this thesis. Special thanks go to Tashi and Benjamin, for being 'paranimf'. We have had many good dinners in the past, and hopefully we will keep having them in the future. Also, I would like to thank my father, who has always supported me in whatever I do. Finally, I want to express my gratitude and love for Elske. She has been at my side since 2006, when we met each other at the radio station Amsterdam FM. Although I am happy I never made it as a journalist and instead finished this thesis, I am thankful for this brief side track in my career; otherwise I would never have got to know her.

Sander Gerritsen

Amsterdam, July 2014

Contents

1

Introduction

‘Human capital is the acquired and useful abilities of all the inhabitants or members of society. The acquisition of such talents, by the maintenance of the acquirer during his education, study, or apprenticeship, always costs a real expense, which is a capital fixed and realized, as it were, in his person. Those talents, as they make a part of his fortune, so do they likewise that of the society to which he belongs.’ (Adam Smith, 1776)

1.1 Background

Human capital has been shown to be important for economic growth (e.g. Hanushek and Woessman, 2008). Countries would not flourish and people would be less happy if individuals did not have at least some. Human capital is not fixed, however, as one can increase her human capital by investing in it. One way to do so is through education. People go to school in order to learn skills that are necessary to produce goods with economic value. Moreover, they learn skills that are relevant for later success in life. This might not only be important for their own success, but also for that of others. Persons may benefit from someone else’s education: people can learn from each other, and may be less likely to become a victim of crime if others are higher educated. This is often referred to as positive spillovers or ‘external’ effects.

However, people may underinvest in human capital. They may not have the financial resources for education, they may be shortsighted, or simply do not see the value of education. This may not only harm them in terms of a lower probability to be employed or lower wages, it may also harm the society as a whole. External effects, as discussed above, might not be internalized if people do not invest enough in their human capital. Moreover, it might lead to higher educational inequality if some underinvest while others would not. For these reasons, the government interferes with people’s decisions with respect to schooling. That is, governments intervene in the market for education. They do so by jurisdiction, i.e. setting rules and regulations, and by subsidizing education. But how does a government intervene in an effective way? That is one of the main questions that economists of education are studying.

To investigate this question, economists look at the so called educational production function (see for examples Boardman and Murnane, 1979; Todd and Wolpin, 2003). This function describes how human capital is accumulated by individuals. The amount of human capital someone accrues is related both to factors *outside* school such as her inherent ability (genes), family environment and networks, and to factors *inside* school such as the time spent in school, class size, and teacher quality. A better understanding of how these educational factors contribute to human capital can help policymakers with decision making. Does a reduction in class size increase student achievement? And if so, by how much? And to what extent does an increase in compulsory schooling lead to the acquisition of more skills? These are all relevant questions. For a policymaker that can choose from an almost undefined number of policy options, it is important to know the answers to these types of questions if she wants to devise good policies. After all, bad policies cost the taxpayer a lot.

This thesis looks at four different input factors that may contribute to the production of human capital. It investigates the importance of teacher quality, time in school, compulsory education, and school facilities for student outcomes, in particular student achievement. The main challenge in this thesis is identifying the causal effect of each of these input factors. Identifying this effect is difficult because of the complexity of the educational production function. Factors outside school often interfere with factors inside schools. For example, schools that give additional instruction time to their students may also accommodate more pupils that are more motivated to learn. If it turns out that these schools have higher student achievement than other schools, it would be difficult to attribute the higher achievement to the additional instruction time, since it could just as well be attributed to student's higher motivation. Moreover, it may not only be the motivation in which students might differ. They can also differ in many other dimensions. As with motivation, these dimensions are often not observable. This means that simple comparisons of students' achievement between schools will not lead to causal estimates of additional instruction time, as the students between the schools are not similar. This commonly known problem, frequently encountered in evaluations of inputs of the educational production function, is called the endogeneity problem. Trying to find ways to deal with this problem has been the challenge over the last two decades in economic studies about the impact of educational inputs on human capital (Angrist and Pischke, 2010). These studies address the endogeneity problem by exploiting randomized experiments or quasi-experiments in which students are exogenously assigned to a control or treatment group. This assignment can be done by the researcher herself via a lottery, in which case it is a randomized experiment. The design of education policies can also lead to exogenous variation in the treatment of students, in which case it is a quasi-experiment or a natural experiment. The exogenous assignment ensures that control and treatment groups are similar, which enables a 'valid' comparison between the two. For estimation of causal effects often instrumental variables, regression discontinuity, and differences-in-differences techniques are exploited, and the studies using these techniques are referred to as the experimental literature.

This thesis makes a contribution to this literature. In particular, it exploits education policies that were implemented in such a way that they created exogenous variation in teacher quality, time in school, compulsory education and school facilities. Using this type of variation enables an impact evaluation of each of these four inputs on student outcomes within a quasi-experimental setting. For the identification of the causal effects for three of these four inputs regression discontinuity methods are used (Chapters 3, 4 and 5). In Chapter 2 twins are exploited. Before summarizing the results of the four chapters, the intuition behind the methods will be discussed briefly.

1.2 Methods

Twins

Twins are often exploited to cancel out family influences that might confound a relationship between the treatment and the outcome. This is done by taking twin differences. In that case, the difference in their treatment is related to the difference in their outcomes. For example, in the returns to schooling literature twin differences in income are related to their differences in education (see for instance Ashenfelter and Krueger, 1994). That is, the income of the twin with more education is compared to that of her sibling with less education. Although this approach has appealing features, it often relies on a strong assumption: it is assumed that the difference in the treatment condition (educational level) between twins is exogenously determined. That is, there are no other differences between twins that are correlated with both their differences in educational level and earnings. This assumption may not be plausible, since twins that differ in education may also differ in other (un)observable characteristics that could contribute to earnings such as motivation or talent. The fundamental problem is that, also within pairs of twins, there might be self selection into treatment, as individual twins choose the amount of education themselves.

Chapter 2 in this thesis introduces a novel twin strategy in which the difference in treatment condition can be considered exogenous. It examines the effect of teacher quality on student achievement by exploiting data on twins who entered the same school but were allocated to different classrooms in an exogenous way. In many Dutch primary schools the assignment of twins to different classes is the result of an informal policy rule that dictates that twins are not allowed to attend the same class. This assignment mechanism might induce a random assignment because at school entry neither schools nor parents have much information about the ability or behavior of twins and the ability or behavior of their class mates. Moreover, because in early childhood twins are more similar than different, it seems not likely that small differences between twins will affect the way they are assigned to different classes. By using this identification strategy, it is for the first time that exogenous variation in twin differences is exploited. That is, in contrast with previous twin studies, there is no self-selection of twins into treatment. Neither twins nor their parents have much to say about the assignment of the twins to different class rooms, whereas in the returns to education literature twins might self select into the level of education.

Regression Discontinuity

Chapters 3, 4 and 5 of this thesis make use of regression discontinuity designs. In contrast to randomized experiments in which subjects are randomly assigned to treatment and control groups, subjects are assigned to treatment status by thresholds on an underlying variable. Subjects with values above a certain threshold value of the underlying variable receive treatment, whereas subjects with values below this threshold value do not. The key idea behind regression discontinuity is that subjects will be similar in (un)observed characteristics around the cutoff. As such, the effect of the treatment can be determined by comparing the outcomes of those subjects above the threshold with those below. For example, in Chapters 3 and 4[1] school entry rules are used to identify the effect of an extra year in school. These rules determine which children start school in the current school year and which children have to wait another year. For instance, if a country uses a school entry rule with a cutoff set at the first of October, this means that children born before October 1 start in the current school year, whereas children born after October 1 start in the next school year. The school entry rules create variation in time in school for children born close to the cut-off date. Students that are almost the same age differ in their time spent in school. In that case, birth month is used as underlying variable and October 1 as threshold value. Chapter 5 also exploits a regression discontinuity design by making use of a threshold in the assignment of a subsidy program targeted at improving facilities for biology, physics, and chemistry in secondary schools. The subsidy was assigned to schools based on a priority score reflecting the ambition level of schools to improve student achievement. Schools with scores below a threshold value did not receive subsidy. The assignment procedure is exploited to estimate the impact of the subsidy on student outcomes.

1 In Chapter 4 the school entry rule is used in combination with the raising of the minimum school leaving age.

1.3 Summary of findings

Chapter 2 examines the effect of teacher quality on student achievement using a novel identification strategy that exploits data on twins who entered the same school but were allocated to different classrooms in an exogenous way. The assignment of twins to different classrooms can be viewed as a natural experiment that exposes very similar individuals to different schooling conditions. This quasi-experiment allows investigation of the causal effect of classroom quality on student outcomes using observational data. The variation in classroom conditions to which the twins are exposed can be considered as exogenous if the assignment of twins to different classes is as good as random. This assumption seems quite plausible within the institutional context of this study, which is Dutch primary education. In many Dutch schools twins are assigned to different classes due to an informal policy rule that dictates that twins are not allowed to attend the same class. This assignment mechanism might induce a random assignment because at school entry neither schools nor parents have much information about the ability or behavior of twins and the ability or behavior of their class mates. Moreover, because in early childhood twins are more similar than different, it seems not likely that small differences between twins will affect the way they are assigned to different classes. By using this identification strategy, it is shown that class room quality comes down to teacher quality, and that this quality is important. Teacher quality can be measured by teacher experience. It is found that (a) the test performance of all students improve with teacher experience; (b) teacher experience also matters for student performance after the initial years in the profession; (c) the teacher experience effect is most prominent in earlier grades; (d) the teacher experience effects are robust to the inclusion of other classroom quality measures, such as peer group composition and class size, and (e) an increase in teacher experience also matter for career stages with less labor market mobility which suggests positive returns to on the job training of teachers.

Chapter 3 uses school entry rules to provide the first estimates of the causal effect of time in school on cognitive skills for many countries around the world, for multiple age groups and for multiple subjects. These estimates enable a comparison of the performance of education systems based on gain scores instead of level scores. Data from international cognitive tests are used and variation induced by school entry rules is exploited within a regression discontinuity framework. The effect of time in school on cognitive skills differs strongly between countries. Remarkably, there is no association between the level of test scores and the estimated gains in cognitive skills. As such, a country's ranking in international cognitive tests might misguide its educational policy. Across countries it is found that a year of school time increases performance in cognitive tests by 0.2 to 0.3 standard deviations for 9-year-olds and by 0.1 to 0.2 standard deviations for 13-year-olds.

Chapter 4 evaluates the effects of raising the minimum school leaving age from 14 to 15 in the Netherlands in 1971. The policy goal was to increase the number of high school graduates. The analysis shows that the change led to a decrease in the high school dropout rate of approximately 20 percent. However, there were no benefits in terms of employment or higher wages. Several explanations for this finding have been explored. Suggestive evidence is presented in support of a skill-based explanation that no more labor-market relevant skills were learned during this extra year of school compared to those skills previously learned out of school.

Chapter 5 evaluates the effects of a subsidy program targeted at improving facilities for biology, physics, and chemistry in secondary schools. The goal of this policy was to increase the enrollment rate in science and engineering (S&E) related courses at secondary and subsequent education institutions. The subsidy was assigned to schools based on a priority score reflecting the ambition level of schools to improve student achievement. Schools with scores below a threshold value did not receive subsidy. The assignment procedure is exploited in a regression discontinuity framework to estimate the impact of the subsidy on student outcomes. It is found that the subsidy increased the enrollment rate in S&E-related courses in secondary school by 3 percentage points (equivalent to a rise of approximately 7.5%). In addition, it is found that the enrollment rate in S&E-related courses in *tertiary* education increased by 2.5 percentage points (equivalent to a rise of approximately 11%). The increased enrollment did not lead to a deterioration in student achievement as measured by students' biology, physics, and chemistry grades. This suggests that supply side policies that make S&E-related courses more attractive – such as the subsidy evaluated in this chapter – are capable of increasing the number of S&E students, while keeping the quality of the supply of S&E students constant.

2

Teacher quality and student achievement: evidence from a Dutch sample of twins[2]

2 This is joint work with Erik Plug and Dinand Webbink.

Abstract

This chapter examines the causal link that runs from classroom quality to student achievement using data on twin pairs who entered the same school but were allocated to different classrooms in an exogenous way. In particular, we apply twin fixed-effects estimation to estimate the effect of teacher quality on student test scores from second through eighth grade, arguing that a change in teacher quality is probably the most important classroom intervention within a twin context. In a series of estimations using measurable teacher characteristics, we find that (a) the test performance of all students improve with teacher experience; (b) teacher experience also matters for student performance after the initial years in the profession; (c) the teacher experience effect is most prominent in earlier grades; (d) the teacher experience effects are robust to the inclusion of other classroom quality measures, such as peer group composition and class size; and (e) an increase in teacher experience also matter for career stages with less labor market mobility which suggests positive returns to on the job training of teachers.

2.1 Introduction

The quality of teachers is considered to be a crucial factor for the production of human capital. Understanding the determinants of teacher quality is important for improving the quality of education and therefore a key issue for educational policy. A large literature has investigated the contribution of teachers to educational achievements of students, the heterogeneity between teachers and the aspects of teachers that are important (e.g. Hanushek and Rivkin, 2006; Staiger and Rockoff, 2010). A consistent finding in the literature is that teachers are important for student performance and that there are large differences among teachers in their impacts on achievement. However, little evidence has been found that any observable characteristic, save experience, explains the variation between teachers. Teacher experience only seems to matter in the initial years in the profession.[3] Hence, the literature does not yet provide clear policy advice about the type of teachers that are most effective, and therefore should be hired and kept in the education profession based on their observed characteristics.

Estimating the effect of teacher characteristics on student performance is complicated because students, teachers and resources are almost never randomly allocated among schools and classrooms. Unobserved factors correlated with both teacher characteristics and student outcomes might bias estimates using non-experimental data. In fact, recent studies have provided evidence for non-random sorting of teachers (Clotfelter et al., 2006; Feng, 2009).

Researchers have addressed this issue by using panel methods or exploiting random assignment of students and teachers into schools and classrooms. The most common approach in the literature is to estimate value-added models that focus on gains in student achievement and eliminate confounding by past unobserved parental and school inputs (Hanushek, 1971, 1992; Aaronson, Barrow and Sander, 2007; Rockoff 2004; Rivkin et al., 2005; Hanushek et al., 2005). Several recent studies exploit multiple years of information for teachers to estimate teacher fixed effects and to link these effects with teacher characteristics (Hanushek et al., 2005; Rockoff, 2004; Aaronson, Barrow and Sander, 2003; Rivkin et al., 2005). Although the most sophisticated value-added models use a three-way-fixed effects approach (student, teacher and school fixed effects) concerns remain about

3 Recent studies find gains from teacher experience beyond the initial years in the career (Wiswall, 2013; Harris and Sas 2011). Mueller (2013) finds that teacher experience moderates class size effects. He finds a class size effect only for senior teachers.

non-random assignment of students to teachers and about modeling assumptions. For instance, Rothstein (2010) finds evidence for dynamic sorting which biases the estimated teacher effects.[4] In addition, Wiswall (2013) shows that restrictive modeling assumptions generate the common finding that teacher experience beyond the initial years in the profession is not important.

A second approach in the literature focuses on classes where students are, or appear to be, randomly assigned. For instance, Clotfelter et al. (2006) use a subsample of schools that feature relatively balanced distributions of students across classrooms, based on observable characteristics. Several studies exploit data from the STAR-experiment in which students and teachers were randomly assigned to small and large classes (Krueger, 1999; Dee, 2004; Nye, Konstantopoulos and Hedges, 2004; Chetty et al., 2012). Teacher experience is found to be the only observed teacher characteristic that matters which is consistent with studies using value-added models (Staiger and Rockoff, 2010). However, the gains from teacher experience are also found after the initial years in the profession.

This chapter examines the effect of teacher quality on student achievement using a novel identification strategy that exploits data on twin pairs who entered the same school but were allocated to different classrooms in an exogenous way. By exploiting an exogenous assignment of individuals to classrooms within the same schools our approach is related to the second approach from the literature. Moreover, the longitudinal character of the data enables us to take prior achievements of students into account like in the common value-added models. The assignment of twins to different classrooms can be viewed as a natural experiment that exposes very similar individuals to different schooling conditions. This quasi-experiment allows us to investigate the causal effect of classroom quality on student outcomes using observational data. The variation in classroom conditions to which the twins are exposed can be considered as exogenous if the assignment of twins to different classes is as good as random. This assumption seems quite plausible within the institutional context of this study; Dutch primary education. In many Dutch schools twins are assigned to different classes due to an (informal) policy rule that dictates that twins are not allowed to attend the same class. At school entry schools and parents do not yet have much information about the ability or behavior of twins and the ability or behavior of their

4 Rothstein (2010) evaluates the most common value-added specifications used for the assessment of teacher performance. He finds that the assumptions underlying common value-added specifications are substantially incorrect and the estimates of teacher effects based on these models cannot be interpreted as causal. See also Guarino et al. (2013) on the validity of value-added measures of teacher performance.

class mates. Moreover, because in early childhood twins are more similar than different, it seems not likely that small differences between twins will affect the way they are assigned to different classes. In our empirical analysis we have tested this assumption and did not find evidence of the non randomness of the assignment.

Our research design exploits the exogenous assignment of twins to different classrooms. The treatment in this design is classroom quality, which is a multi-dimensional concept that might include factors such as peer quality, class size and teacher quality. In our empirical analysis we especially focus on the effects of observed teacher characteristics on student outcomes because, in applying our design, teachers seem the most obvious factor differing across classes. Typically, Dutch schools equalize classroom facilities and class composition across classes. In schools with many students and few teachers we expect little variation within twin pairs in class size and peer composition, and much within twin pair variation in teacher quality. Therefore, we will exploit the assignment of twins to different classes in particular to estimate the effects of teacher characteristics on student outcomes. For doing so, we use longitudinal data of a large representative sample of students from Dutch primary education. We have identified twins from the population based sample by using information on their date of birth, family name and school.

This chapter makes two important contributions to the current economic literature. First, we contribute to the literature on teacher quality by introducing an empirical strategy that has never been used to estimate the impact of teacher quality on student outcomes. Previous studies have relied on value-added modeling (e.g. Rivkin et al., 2005; Rockhoff, 2004) or exploited classes where students are, or appear to be, randomly assigned (e.g. Krueger, 1999; Chetty, 2005; Clotfelter et al., 2006). Second, we contribute to the economic literature that exploits data on twins by combining twins with exogenous treatment assignment. Twin differencing has been applied on various topics such as the returns to schooling or the intergenerational effects of schooling (see for instance Ashenfelter and Krueger, 1994; Behrman and Rosenzweig, 2002). These studies are based on the assumption that variation within twin pairs is exogenous but it remains unclear why twins differ.[5] As far as we know, there are no twin studies that arguably exploit exogenous variation in twin differences.

5 Li et al. (2010) exploit a twin design in which parents are forced to send one of their twins to the countryside. The treatment assignment might not be random as it is based on a parental decision.

In line with earlier studies on teacher effects we find that teacher experience is the only observed teacher characteristic that matters for student performance (Hanushek, 2011; Staiger and Rockoff, 2010; Chetty et al., 2011). Twins who are assigned to classes with more experienced teachers perform better in reading and math. On average one extra year of experience raise test scores by approximately one percent of a standard deviation. The effects of teacher experience are most pronounced in kindergarten and early grades. Our findings are remarkably consistent with the results found by Krueger (1999) and Chetty et al. (2005) using data from the STAR-experiment in which students and teachers were randomly assigned to classes. They also report linear effects of teacher experience and find that the effects of teacher experience reduce after kindergarten. The linear effects of teacher experience contrast 'the consensus in the literature' that only initial teacher experience matters (Staiger and Rockoff, 2010), but are in line with recent findings by Wiswall (2013) and Harris and Sas (2011) who also find gains from later experience.

The estimated effects of teacher experience should be interpreted carefully because they do not necessarily reflect the effect of training on the job but could be driven by other intrinsic characteristics of more experienced teachers.[6] For instance, the findings might be driven by sample attrition if less effective teachers are more likely to leave the profession.

In the empirical analysis we test for various mechanisms that might explain why experienced teachers are better teachers. We do not find evidence consistent with mechanisms that stress the importance of changes over time such as changes in the quality of teacher education or changes in outside opportunities in the labor market. However, we also find an effect of teacher experience for career stages with less labor market mobility. These estimates suggest positive returns to on the job training of teachers as it is less likely that these estimates will be biased because of selection into or out of the profession. This finding is consistent with recent studies that also have found positive return to teacher experience for later career stages (Wiswall, 2013; Harris and Sas, 2011).

Regardless the story, our estimates show that experienced teachers matter (especially for reading). More focused policies to maintain experienced teachers in the classroom appear beneficial, especially for younger students.

This chapter is organized as follows. In Section 2.2 we describe our empirical strategy and relate this to previous approaches from the literature. Section 2.3 describes the data. The main estimation results are presented in Section 2.4. Section 2.5 provides additional

6 Rockoff (2004), Kane et al. (2006), Chetty et al. (2011) and Harris & Sass (2011) have previously noted this issue.

tests about the key assumption of our empirical strategy about the random assignment of twins to classrooms. Section 2.6 explores the possible mechanisms underlying the estimated effects of teacher experience. In Section 2.7 we conclude.

2.2 Empirical strategy

The basic framework in the economic literature that studies the effects of teachers models student achievement as a function of family, peer, community, teacher and school inputs and student ability (Hanushek and Rivkin, 2006). Student achievement at any point in time is seen as a cumulative result of the entire history of all inputs and the individual's initial endowment (e.g. innate ability). A common approach for modeling this so-called educational production function is to assume that the cumulative achievement function is additively separable and linear (e.g. Boardman and Murnane, 1979; Todd and Wolpin, 2003; Harris and Sas, 2011). Estimating the effect of teachers is complicated because in any actual application we will generally not be able to control for all relevant school, family or student characteristics. If some omitted variables are correlated with the relevant teacher characteristics, then the estimated parameters will be biased. The major threat to identification is the non-random sorting of students among schools and classrooms.

Researchers have used two types of empirical strategies for identifying the effects of teacher characteristics. The first, and most common, approach is based on value-added modeling. The second approach exploits situations where students are, or appear to be, randomly assigned. The most common approach in the literature is to include measures of prior achievement and estimate value-added models. These models focus on gains in student achievement or the rate of learning over specific time periods. Recent studies exploit the availability of multiple years of information for teachers for estimating teacher fixed effects which are linked with teacher characteristics (Hanushek, 1992; Hanushek et al., 2005; Rockoff, 2004; Aaronson, Barrow and Sander, 2003; Rivkin et al., 2005). A second approach in the literature identifies teacher effects by exploiting situations in which students are, or appear to be, randomly assigned to classrooms and teachers (e.g. Clotfelter et al., 2006; Chetty et al., 2012). It might be expected that unobserved factors will not bias the estimates due to the random assignment of students.

In this chapter we exploit the assignment of twins to different classrooms for estimating the causal effect of class inputs on student achievement. This assignment can be viewed as a natural experiment that exposes very similar individuals to different class room conditions. Our approach is most related to 'the random assignment studies' but by including previous test scores as controls we are also able to estimate value-added models. For explaining our empirical strategy we use as a starting point the basic economic framework which relates class inputs to measures of educational performance. Consider the following specification of a traditional educational production function:

$$(2.1) \qquad Y_{ijcs} = \alpha_1 X_i + \alpha_2 Z_j + \alpha_3 SQ_s + \alpha_4 CQ_c + x_i + z_j + sq_s + cq_c + \varepsilon_{ijcs}$$

where indices i, j, c and s stand for pupil i born in family j in classroom c at school s. Observable educational output Y represents the test scores on reading or math. Observable inputs of the educational production function contain individual attributes X, family characteristics Z, and various measures of classroom quality CQ and school quality SQ. For ease of notation, we keep the specification very general in the sense that any class attribute could be represented by CQ. It could be teacher's experience for example, but also class size or composition. The error term consists of analogous unobservable inputs of the educational production function x, z, sq, cq and an idiosyncratic effect ε which is uncorrelated with all these observable and unobservable determinants. In this chapter we are interested in estimating α_4 which represents the structural effect of an observable class input on pupil test scores.

If we use conventional cross section data to estimate equation (2.1), least squares estimation might not yield an unbiased estimate of α_4 for multiple reasons. First, there might be a non random assignment of pupils to classes, i.e. $Cov(CQ,x) \neq 0$. This means for example that worse performing pupils are more often assigned to classes with better peers or better teachers. Second, there might be parental influences on the child's school and classroom, i.e. $Cov(CQ,z) \neq 0$. For instance, higher educated parents may select better schools or class rooms because they may be more involved with their children than lower educated parents. Third, better schools may attract better teachers (and better pupils), i.e. $Cov(CQ,sq) \neq 0$, because differences in quality of school management may cause different schools to attract different teachers. Fourth, schools may use multiple inputs to manipulate classroom environment, i.e. $Cov(CQ,cq) \neq 0$, because schools may decide to compensate classes with high fractions of low ability pupils by reductions in class size and/or extra aide. To sum up, estimating an equation like (2.1) with non experimental data is likely to induce bias due to unobservable characteristics of pupils, parents, and teachers and schools (i.e. school management).

Our empirical strategy for identifying α_4 exploits differences within pairs of twins.

If we suppress subscripts and take twin differences, our empirical model can be rewritten as

$$(2.2) \qquad \Delta Y = \alpha_1 \Delta X + \alpha_2 \Delta Z + \alpha_3 \Delta SQ + \alpha_4 \Delta CQ + \Delta x + \Delta z + \Delta sq + \Delta cq + \Delta \varepsilon$$

Identification of α_4 now rests on four assumptions:

(A.1) twins share family background, i.e. $\Delta Z = 0$ and $\Delta z = 0$

(A.2) twins enter the same school, i.e. $\Delta SQ = 0$ and $\Delta sq = 0$

(A.3) twins are exogenously allocated to different classrooms, i.e. $\Delta CQ \neq 0$ but $Cov(\Delta CQ, \Delta x) = 0$

(A.4) observable and unobservable class attributes are unrelated, i.e. $Cov(\Delta CQ, \Delta cq) = 0$

Assumptions (A.1) and (A.2) are satisfied by design. Assumption (A.3) seems also plausible because of the assignment procedures in Dutch education. We will discuss the plausibility of this assumption shortly. However, assumption (A.4) seems not plausible. If we assume that the assumptions (A.1), (A.2) and (A.3) hold we can simplify the empirical model to:

$$(2.3) \qquad \Delta Y = \alpha_4 \Delta CQ + \Delta cq + \Delta \varepsilon$$

Twin fixed-effect estimation will therefore give us the following estimator:

$$\alpha_4^{FE} = \frac{Cov(\Delta Y, \Delta CQ)}{Var(\Delta CQ)} = \alpha_4 + \frac{Cov(\Delta CQ, \Delta cq)}{Var(\Delta CQ)}$$

Hence, this twin fixed effect estimator not only captures the impact of any observable classroom characteristic but also the impact of every unobservable characteristic that is correlated with it. This can be interpreted as the broad impact of classroom quality.

As in any quasi-experimental design there are deviations from the ideal experimental design in which a specific treatment is randomly assigned to an experimental group of students. The first, and crucial, issue in this design is whether the assignment of twins is truly exogenous (assumption (A.3)). The second issue in our design is about the treatment variable. What is the treatment considering the fact that classroom quality is multi-dimensional? Both issues should be considered within the institutional context of Dutch primary education. In Dutch primary education parents and pupils are free to choose their school. All schools receive funding from the government based on the number and socioeconomic background of the pupils. Primary school consists of eight grades of which grade 1 and grade 2 are equivalent to kindergarten. Children are allowed to enroll in primary education on their fourth birth day which induces a rolling admission in grade 1. Compulsory education starts at the age of 5. Most schools mix first and second graders. After grade 2 children are reassigned to different classes. The composition of these classes remains quite stable until the end of primary education in grade 8 in which pupils take a nationwide test.

Is the assignment of twins truly exogenous?

The key identifying assumption of our approach is the random assignment of twins to different classrooms. Many schools in Dutch primary education employ a policy of separating twins in different classes.[7] This separation already takes place when the twins enroll in grade 1. The rolling admission of pupils in grade 1 implies that class size and classroom composition are volatile and only partly observed by parents. After finishing the school year in which pupils enrolled they spend two complete school years in grade 1 and 2.

Hence, in total most students spend more than two years in grade 1 and 2. During this whole 'kindergarten stage' pupils keep the same teacher(s) and are not reassigned to other classes. This school policy is likely to induce random assignment at school entry since in kindergarten there is no (or very little) information on class quality, such as the quality of the class mates, that parents can use to determine which type of class suits their twins best. In addition, twins in early childhood are more similar than different and it seems unlikely that small differences between twins affect the way in which they are assigned to different classes. From this we expect that the assignment of newly entering twins, which creates the classes of grade 1 and 2, can be considered as exogenous. We will run twin-fixed effect regressions and interpret the corresponding estimate as the broad impact of classroom quality.

A reassignment of students in Dutch education takes place in the transition from grade 2 to grade 3. At this stage there will be more information available about the twins and their class mates, although it might be expected that the twins are very similar. This implies that we are not fully sure whether the assignment is still exogenous for third and higher graders.

To address this concern we will estimate value-added models which include previous test scores of our twins. We will compare the results of the basic random assignment

7 The Dutch Society for Parents of Multiples advises parents to follow their own opinion, but believes that separation stimulates the individualization of the twins (Geluk & Hol, 2001). Most schools explicitly put their twin policy on their website. Many schools assign twins to different classes based on the belief that putting them in the same class will harm them, although recent research suggest that this is not the case (see for example Webbink et al., 2007).

specifications with the results of the value-added specifications. The potential bias due to non random assignment of twins is expected to be small if these results are very similar. In addition, we will perform several tests on the empirical importance of endogenous classroom assignment.

What is the treatment?

In our design we compare the performance of twins who are assigned to different classes. The treatment in this design is classroom quality, which is a multi-dimensional concept. The literature on class quality typically focuses on the impact of peers, teacher quality and class size. In the Dutch institutional context we expect that teacher quality will be the most important component of this treatment. Due to the rolling admission class size and peer composition are volatile in grade 1 and 2, whereas teacher quality is fixed. In addition, Dutch schools typically equalize facilities, peer composition and class size across classes. In schools with many students and few teachers we expect much within twin pair variation in teacher quality and little within twin pair variation in class size and classroom composition. Therefore, we expect that the assignment of twins to different classes in grade 1 and 2 can mainly be interpreted as an assignment to different teachers. For grade 3 to 8 we also expect that teacher quality is the main component of the treatment. With many students and few teachers in schools we expect only little within twin pair variation in class size and class composition. Most variation in classroom quality will come from differences in teacher quality. A similar interpretation has been used in the literature that investigates the effect of school and teacher quality through the estimation of classroom fixed effects on achievement gains. The resulting classroom differences in average achievement gain have been interpreted as reflecting teacher quality, since the teacher is the most obvious factor differing across classrooms (Hanushek, 1992). In this study we will therefore investigate teacher quality effects in more detail and focus on components such as experience, gender and fulltime or part-time employment of teachers.

2.3 Data

The data come from the longitudinal biannual PRIMA project (Driessen et al., 2004). The PRIMA project consists of a panel of approximately 60,000 pupils in 600 schools. The participation in the project is voluntary. The main sample, which includes approximately 420 schools, is called the reference sample, which is representative for the Dutch student population in primary education. An additional sample includes 180 schools for the over-sampling of pupils with a lower socioeconomic background (the low SES sample). After each wave of the project some schools drop out and some new schools are included.[8] This means that the panel structure only holds for a subsample of the dataset.[9] We use all six waves of the PRIMA survey including data on pupils, parents, teachers and schools from the school years 1994-95, 1996-97, 1998-99, 2000-01, 2002-03 and 2004-05. Within each school, pupils in grades 2, 4, 6 and 8 (average age: 6, 8, 10, 12 years) are tested in reading and math. Information on teachers is also collected but the main focus of the project is to follow pupils (and not teachers) during primary education.

Our identification strategy is based on differences within pairs of twins. The PRIMA-data does not contain direct information on twins versus singletons. We have identified twins by matching on family name, date of birth, school and year of the survey. If there are two pupils with exactly the same values on these matching variables they are considered to be twins. In the total sample of the PRIMA-data we have identified 623 records of twin pairs that were assigned to different classrooms and for which (reading) test scores and teacher data are available. Because of the longitudinal character of the data some twin pairs will be observed more than once; the number of unique twin pairs in our data is 495. The total sample of 623 twin pair observations consists of 448 same sex pairs (219 pairs of boys, 229 pairs of girls), 173 opposite sex pairs and 2 pairs with unknown gender. More twin pairs have been identified in earlier grades than in later grades; 235 pairs in grade 2, 175 pairs in grade 4, 132 pairs in grade 6 and 81 pairs in grade 8. If one of the twins is retained or accelerated we will not observe a pair because of the sampling structure of the PRIMA-project (only grade 2, 4, 6 and 8). This might explain the lower number of pairs in later grades.

Our main dependent variables are scores on tests for languages and arithmetic which were developed as part of the PRIMA-project. The language test for children in second grade, which is equivalent to infant school, measures the understanding of words and concepts. The arithmetic test for these children focuses on the sorting of objects. These tests are taken in class. The test for children in grades 4, 6 and 8 all come from a system for following pupil achievements in primary education developed by the CITO group. The aim of these tests is to observe to what extent students master various elements of the curriculum. The tests for the same grade levels are identical each year. This ensures that the

8 There are no significant differences between the schools that drop out and the schools that remain in the project (Roeleveld and Vierke, 2003).

9 Other reasons are pupils changing schools or pupils not being present at the time tests are taken.

comparison of achievement levels over time is possible. The scores are also comparable between grades. The scales of the raw scores for language and arithmetic have no clear meaning. We have standardized these test scores by wave and grade with the mean and standard error of the reference sample.

The main explanatory variables in this chapter are a set of class input factors. First, we have information on teacher characteristics: experience (measured in years) and gender. In addition, the data provide information whether the class is taught by one fulltime teacher or by two part-time teachers. In case of two teachers we use the experience of the teacher that was present at the time of the survey. Class size is reported by the teacher but is also available from the PRIMA-register. Moreover, we use two measures of the composition of the class: fraction of girls and fraction of native Dutch pupils. The latter is a proxy for the socioeconomic status of twin's class mates.

Table 2.I shows the descriptives for the samples of twins in grade 2, in grade 4-8, and for the total sample of twins. In addition, the last columns of table 2.I show the descriptives for the total sample of the PRIMA-project.

Table 2.I: Descriptive statistics of estimation samples (for reading)

	Twin samples						PRIMA-sample	
	Grade 2		Grades 4, 6 and 8		Total		Total	
	mean	sd	mean	sd	mean	sd	mean	sd
Twin characteristics								
reading score	-0.39	1.07	-0.16	1.07	-0.25	1.08	-0.15	1.02
math score	-0.33	0.94	-0.11	1.05	-0.20	1.02	-0.12	1.01
girl	0.50	0.50	0.49	0.50	0.49	0.50	0.50	0.50
Teacher & class room								
experience in education (years)	15.36	10.43	16.60	11.30	16.13	10.99	18.21	10.58
female	0.98	0.15	0.63	0.48	0.76	0.43	0.66	0.47
multiple classroom teachers	0.52	0.50	0.40	0.49	0.45	0.50	0.50	0.50
split level classroom	0.83	0.38	0.21	0.41	0.44	0.50	0.47	0.50
class size	24.07	4.50	23.54	4.95	23.74	4.79	24.39	5.78
observations	470		776		1246		330350	
Previous test scores twins & class composition (for subsample in grades 4-8, excluding mixed grades)								
reading score T-2	-	-	-0.28	0.93	-	-	-0.12	1.01
math score T-2	-	-	-0.13	0.98	-	-	-0.05	0.98
girl share in class (%)	-	-	50.49	10.02	-	-	50.06	11.72
native share in class (%)	-	-	60.63	35.92	-	-	64.99	34.17
observations			276				97557	

The number of observations differs between grades because of the longitudinal structure of the data and the sampling strategy of the PRIMA-project. Previous test scores are only available for grade 4 and higher, and for pupils (schools) who participated in previous waves of the project. The bottom panel of table 2.I shows that we have previous test scores for one third of the sample of twins in grade 4 and higher. In addition, information about the peers is not available for classrooms with multiple grades, which is often the case in grade 2. Teacher characteristics have been measured for most classrooms. This implies that the estimation sample for models that exploit the longitudinal character of the data or models that include peer characteristics will be smaller than models that only use cross-sectional data and focus on teacher characteristics. The means of the test scores in our twin sample are negative which means that twins perform below the average of the student population of the reference sample. The means of the test scores for the total PRIMA-sample are also negative as we used the reference sample for the standardization and the total sample also includes the low-SES sample.

Table 2.II shows the variation of the class inputs within pairs of twins, which is the variation that is crucial for our identification strategy. The variance in teacher characteristics within twin pairs is much larger than the variation in class size. At the 95th percentile the difference in teacher experience within a twin pair is 26 years, for class size this difference is 4 pupils.

Table 2.II: Distribution of twin differences in class room characteristics in total estimation sample for reading (# Twin pairs=623)

	Percentiles						
	5th	**25th**	**50th**	**75th**	**95th**	**mean**	**sd**
Δ reading	-1.78	-0.64	-0.05	0.56	1.56	-0.07	1.02
Δ math	-1.64	-0.60	-0.08	0.50	1.47	-0.08	0.93
Δ girl	-1.00	0.00	0.00	0.00	1.00	0.02	0.53
Δ experience (in years)	-26.00	-10.00	0.00	8.00	24.00	-0.78	14.57
Δ female (teacher)	-1.00	0.00	0.00	0.00	1.00	0.01	0.46
Δ multiple class room teachers	-1.00	0.00	0.00	0.00	1.00	0.03	0.67
Δ split level classrooms	-1.00	0.00	0.00	0.00	1.00	0.02	0.41
Δ class size	-4	-1	0	1	4	-0.06	2.81

2.4 Main estimation results

In this section we show the main results of our empirical analysis. We start by presenting the results for students in grade 2. For these students we are most confident about the assumption that twins are exogenously assigned to classrooms because at school entry there is hardly any information about the student and his/her classmates that might lead to selection into classrooms. Table 2.III shows the twin-fixed effect estimates of teacher characteristics on student performance in reading and math using different specifications.[10] We investigate the effect of teacher experience, gender of the teacher and having one fulltime or two part-time teachers in the classroom.

Table 2.III: Twin-fixed effect estimates of teacher quality effects on student test scores in grade 2

Independent variables:	**(1) reading**	**(2) math**	**(3) reading**	**(4) math**	**(5) reading**	**(6) math**
teacher experience	0.014***	0.015***	0.014***	0.015***	0.014***	0.014***
	(0.005)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)
female teacher			-0.174	0.290	-0.106	0.271
			(0.314)	(0.291)	(0.342)	(0.281)
two teachers			0.065	0.025	0.065	0.020
			(0.098)	(0.076)	(0.099)	(0.076)
split level classroom			-0.228	-0.103	-0.231	-0.151
			(0.183)	(0.180)	(0.192)	(0.189)
girl (in opposite sex twin pair)			0.162	0.169	0.156	0.160
			(0.120)	(0.115)	(0.119)	(0.112)
class size			0.005	0.033	0.004	0.033
			(0.032)	(0.027)	(0.031)	(0.028)
% girls in class					-0.006	-0.001
					(0.004)	(0.003)
% natives in class					0.004	-0.004
					(0.003)	(0.004)
# twin pairs	235	236	235	236	235	236
R-squared	0.035	0.052	0.049	0.074	0.061	0.079

Notes: Each column shows the results of an OLS-regression. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 Covariates have been imputed, see footnote 10.

10 To improve the power of our analysis we have used a sample for which missing values for several covariates have been imputed. In case of a missing value on a covariate we have assumed that there is no difference within a pair of twins. For reading we have imputed 31 observations, for math we have imputed 27 observations. Teacher experience has not been imputed. The main estimation results do not change when we use the smaller samples without imputations. Estimations results are available upon request.

The first two columns of table 2.III show the estimated effects of having a more or less experienced teacher in the classroom in models without controls. The estimates show that one additional year of teacher experience in the classroom increases performance in reading or math with 1.4 or 1.5 percent of a standard deviation. The other teacher characteristics are included in the models in column (3) and (4). These models also control for split level classrooms, class size and gender of the student. The observed teacher characteristics, gender of teacher and the number of teachers in the class room, do not affect student performance. We also observe that the inclusion of the new variables does not change the estimated effects of teacher experience in the classroom. Column (5) and (6) additionally controls for differences in classroom composition, in particular the proportion of girls and the proportion of native students in the classroom. Again, we observe that including these controls does not change the effect of teacher experience. Hence, the estimates for students in grade 2 suggest that teacher experience in the classroom is an important determinant of student performance and teacher experience seems to be the only observed teacher characteristic that matters. These findings are consistent with previous results from the literature on teacher quality (see Section 2.1 and 2.2).

The returns to teacher experience beyond the initial years in the profession

The recent literature is not consistent about the returns to experience during various stages of the teaching profession. Many studies have found that experience only matters in the initial years in the profession (e.g. Rivkin et al., 2005; Rockoff, 2004) and there seemed to be a consensus about this finding in the literature (Staiger and Rockoff, 2010; Wiswall, 2013). However, recent studies also find gains from teacher experience in later years of the career (Harris and Sas, 2011; Wiswall, 2013). Moreover, Wiswall (2013) shows that restrictive modeling assumption in previous studies have generated the common finding that experience only matters in the first years of the profession. Using an experience variable with a limited number of categories within a panel setup that includes only a few years of information on teachers seriously reduces the variance that can be exploited in the estimation. He finds high returns to later experience using an unrestricted experience model for student performance in math. For student performance in reading he finds low returns to later experience. Previous studies based on the data from the STAR-project report linear effects of teacher experience (Krueger, 1999; Chetty et al., 2012). Our estimates also suggest a linear effect of experience on student achievements (see also figure 2.1A and figure 2.1B).

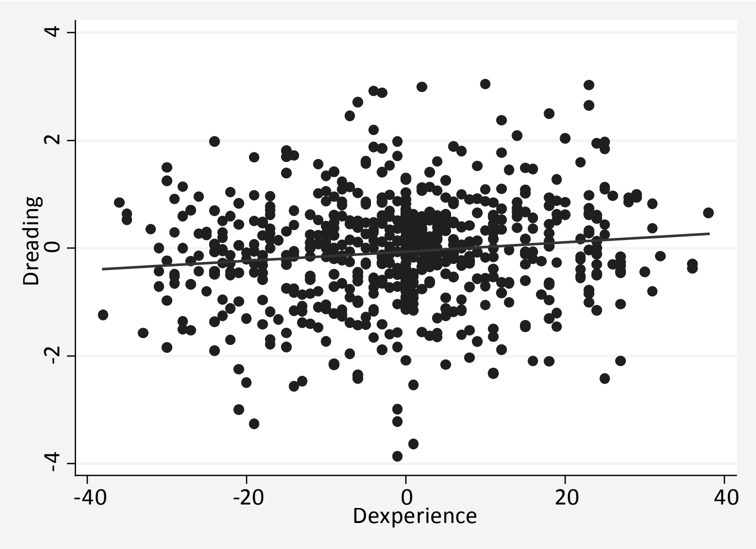


Figure 2.1A. Student performance in reading by teachers experience within pairs of twins

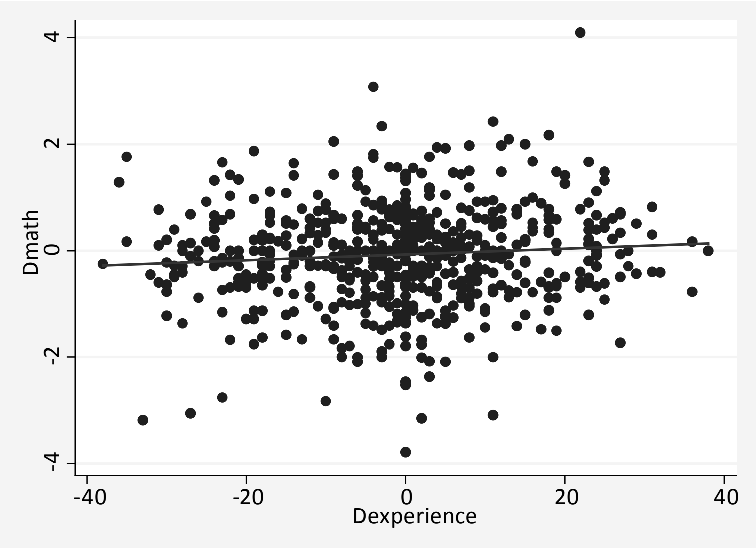


Figure 2.1B: Student performance in math by teachers experience within pairs of twins

We have also experimented with higher order term of experience but we did not find significant results for these specifications.[11]

Results for the full sample of students from grade 2 to 8

In the next step of our analysis we use the full sample of twins from grade 2, 4, 6 and 8. As noted in Section 2.2, for the full sample of twins we are less confident about the assumption that students have been randomly assigned to classrooms because students in Dutch primary education are re-assigned to classes after grade 2. It might be expected that teachers, parents and students will have more information after grade 2 about themselves and other students which might lead to non- random selection into classes. This selection might bias the estimated effects for the full sample. To address this issue we not only estimate the 'random assignment specifications' from table 2.III but also estimate value-added specifications in which we control for previous test scores. By combining a value-added specification with our experimental design we aim to mitigate non-random selection into classes. We further investigate the empirical importance of endogenous classroom assignment after grade 2 in Section 2.5.

The estimation results for the full sample of twins are shown in table 2.IV. Column (1) to (8) shows the estimated effects using the random assignment specifications that are also used in table 2.III. Column (1) to (4) use the full sample of twins, in column (5) to (8) we only use twins for which previous test scores are available. Column (9) to (12) show the estimation results for the value-added specifications. A disadvantage of including previous test scores is that we typically loose the first observation (pupils in grade 2) because the previous test score is not available. However, since the random assignment of pupils in grade 2 ensures that there are no initial differences within the twin pairs we can replace the previous test score with a constant, in order to keep the first year of the data.[12] For the full sample of twins we find that one additional year of teacher experience in the classroom increases performance in reading with 0.9 to 1.4 % of a standard deviation and performance in math with 0.6 to 0.9 %. A comparison of the results from the 'random assignment specification' with the results from 'the value-added specification' can be considered as an important test for the non-random assignment of twins because generally previous test scores are important control variables.

11 We did not use a specification with a limited number of experience categories because of the restrictive nature of this approach as pointed out by Wiswall (2013).

12 Krueger (1999) and Mueller (2013) use a similar approach.

Table 2.IV: Twin-fixed effect estimates of teacher quality effects on student test scores in grades 2 to 8

	Random assignment specification								Value-added specification			
	Total sample				Sample with previous test scores				Sample with previous test scores			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Independent variables:	**reading**	**math**	**reading**	**math**	**reading**	**math**	**reading**	**math**	**reading**	**math**	**reading**	**math**
teacher experience	0.009***	0.006**	0.011***	0.006**	0.013***	0.008***	0.014***	0.008***	0.013***	0.009***	0.014***	0.009***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
female teacher			0.164*	0.132*			0.075	0.122			0.072	0.201**
			(0.090)	(0.078)			(0.107)	(0.095)			(0.103)	(0.090)
two teachers			-0.034	-0.001			-0.002	0.070			-0.015	0.061
			(0.062)	(0.056)			(0.072)	(0.060)			(0.071)	(0.059)
class room that mixes grades			0.142	0.116			0.125	0.147			0.098	0.115
			(0.110)	(0.116)			(0.114)	(0.132)			(0.112)	(0.125)
girl (in opposite sex twin pair)			0.163*	-0.121			0.229**	-0.099			0.208**	-0.097
			(0.092)	(0.089)			(0.097)	(0.086)			(0.093)	(0.080)
class size			-0.005	0.008			-0.020	0.018			-0.021	0.017
			(0.016)	(0.016)			(0.020)	(0.020)			(0.019)	(0.019)
% girls in class			-0.001	0.002			-0.002	0.004*			-0.002	0.004
			(0.003)	(0.002)			(0.003)	(0.002)			(0.003)	(0.002)
% natives in class			0.006*	-0.003			0.005*	-0.003			0.005	-0.002
			(0.003)	(0.003)			(0.003)	(0.003)			(0.003)	(0.003)
previous test score (t-2)									0.267***	0.340***	0.246***	0.348***
									(0.070)	(0.061)	(0.071)	(0.059)
# twin pairs	623	611	623	611	451	447	451	447	451	447	451	447
R-squared	0.016	0.008	0.039	0.022	0.034	0.019	0.063	0.041	0.065	0.082	0.089	0.105
Controls:												
Teacher/Class characteristics	no	no	yes	yes	no	no	yes	yes	no	no	yes	yes
Previous test scores	no	no	no	no	no	no	no	no	yes	yes	yes	yes

Notes: Each column shows the results of an OLS-regression. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Covariates have been imputed, see footnote 10.

2

We observe in table 2.IV that the estimates based on the 'random assignment specification' are very similar to the estimates based on the 'value-added specification' which suggests that the bias from non-random assignment will be limited. Again, including higher order terms of experience does not change the estimated effects. The estimated effect of teacher experience is robust to the inclusion of the other teacher characteristics and other controls. For the other teacher characteristics we find no systematic effects on student performance. Hence, the estimates we have found for the sample of twins in grade 2 are consistent with the estimates based on the whole sample of twins.

Teacher experience and grade level

Previous studies have reported different returns to experience by grade level. Krueger (1999) and Chetty et al. (2011) find higher effect of teacher experience for kindergarten than for higher grades. We have also investigated whether teacher experience is more important for younger pupils. Table 2.V shows the estimation results for teacher experience by grade level;

Panel A shows the results for the random assignment specification based on the total sample, Panel B shows the results for the value-added specification based on the sample for which we observe previous test scores. The estimates show that the effect of teacher experience depends on the grade of the pupil. Teacher experience in class matters most in grade 2 (kindergarten): an additional year of experience in class raises test scores by approximately 1.5 % of a standard deviation. Teacher experience becomes less important in higher grades. In grade 8 we don't find an effect of teacher experience on performance. Hence, teacher experience raises test scores especially for younger pupils. This finding is very similar to the results based on data from the STAR-project. Our finding might also be explained by the fact that students in grade 2 have the same teacher for two years whereas in higher grades this is less likely.

Table 2.V: Twin fixed effect estimates of teacher experience on student test scores by grade

	grade 2		grade 4		grade 6		grade 8	
	reading	**math**	**reading**	**math**	**reading**	**math**	**reading**	**math**
Panel A: random assignment specification, total sample								
teacher experience	0.014***	0.014***	0.009**	0.005	0.011**	0.001	0.004	-0.004
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.009)	(0.007)
# twin pairs	235	236	175	173	132	128	81	74
R-squared	0.061	0.079	0.074	0.090	0.077	0.075	0.112	0.120
Panel B: value-added specification, sample with previous test scores								
teacher experience	0.014***	0.014***	0.013**	0.010	0.013**	0.002	0.007	-0.007
	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.012)	(0.009)
# twin pairs	235	236	82	86	81	77	53	48
R-squared	0.061	0.079	0.180	0.275	0.260	0.262	0.235	0.568

Notes: Each column within a panel shows the results of an OLS-regression. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Panel A includes all covariates as in column (3) and (4) in table 2.IV. Panel B also includes previous test scores.

2.5 Sensitivity tests about non random classroom assignment after grade 2

To further investigate the empirical importance of endogenous classroom assignment after grade 2 we perform several tests. First, we regress test scores in second grade on classroom characteristics in fourth grade. If assignment is random, we should not observe a relationship between test scores and classroom characteristics. Table 2.VI shows the results and provides no evidence for a nonrandom assignment of twins after grade 2. For the models in column (8) that include all variables simultaneously we find two statistically significant effects but the F-test shows that we cannot reject the hypothesis that there is no classroom-effect. Better performing twins in grade 2 are not assigned more often to other type of classes in grade 4 than their (worse performing) twin brothers or sisters on observed class inputs.

Table 2.VI: Twin-fixed effect estimates of grade 2 reading scores on grade 4 classroom characteristics

Classroom characteristics in fourth grade:	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
teacher experience	-0.009							-0.011*
	(0.006)							(0.006)
female teacher		-0.240						-0.174
		(0.213)						(0.238)
two teachers			0.015					0.053
			(0.155)					(0.182)
split level classroom				0.248				0.275
				(0.176)				(0.191)
class size					-0.015			-0.017
					(0.040)			(0.044)
% girls in class						-0.006		-0.007
						(0.004)		(0.006)
% natives in class							0.003	-0.003
							(0.008)	(0.010)
p-value F-test: no class room effect								0.307
# twin pairs	87	87	88	88	87	112	108	79
R-squared	0.018	0.016	0.000	0.012	0.002	0.009	0.001	0.052

Notes: Each column shows the results of an OLS-regression. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

As our second test we regress test scores obtained in second, fourth, sixth and eighth grade on class room characteristics that were measured in the second grade. If assignment is random at school entry, but assignment in later years is not, reduced form estimates (assuming that classroom characteristics are correlated across grades) are informative about class input effects. Panel A of table 2.VII shows the results related to this test. The first two columns show the effect of teacher experience for the full sample; these results have already been shown in table 2.IV. The next two columns show the results for the sample of twins for which we have test scores in grade 2 and in at least one higher grade. The estimates effects for this sample imply that one additional year of teacher experience in class increases performance with 1.3 percent of a standard deviation. The last two columns of panel A show the estimated effect of teacher experience in grade 2 on the test scores in grade 2 or in higher grades. Hence, these columns show the reduced form estimates. If assignment to classrooms in higher grades is not random but assignment to grade 2 is random the reduced form estimates are informative about the effect of experience in class.

Table 2.VI continued: Twin-fixed effect estimates of grade 2 math scores on grade 4 classroom characteristics

Classroom characteristics in fourth grade:	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
teacher experience	-0.008							-0.009
	(0.006)							(0.006)
female teacher		-0.193						-0.108
		(0.131)						(0.119)
two teachers			0.027					0.035
			(0.139)					(0.153)
class room that mixes grades				-0.165				-0.001
				(0.172)				(0.165)
class size					-0.014			-0.012
					(0.037)			(0.040)
% girls in class						0.005		-0.000
						(0.006)		(0.007)
% natives in class							-0.008	-0.018**
							(0.006)	(0.008)
p-value F-test: no class room effect								0.177
# twin pairs	92	93	93	93	92	117	113	85
R-squared	0.019	0.012	0.000	0.007	0.002	0.006	0.014	0.072

Notes: Each column shows the results of an OLS-regression. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

The reduced form estimates indicate that one year of teacher experience in class increases test scores with approximately 1 percent of a standard deviation. These estimates lie in the same ball park as the previous estimates. In sum, these tests don't provide evidence for a non random re-assignment after grade 2, which might threaten the identification of the estimates in table 2.IV.

Further robustness analyses are shown in panel B of table 2.VII. Students in split level classrooms have now been excluded from the estimation sample. For these students we only have class composition information from the students in the same grade but not from the students in the other grade. Although this strongly reduces the sample size the estimated effects of teacher experience remain quite similar to the previous estimates.

Table 2.VII: Various twin-fixed effect estimates of teacher experience on student test scores for grades 2, 4, 6, 8

Panel A: sample with split level classrooms	**Total sample**		**Sample for which teacher experience in grade 2 is available**		**Reduced form**	
	(1)		**(2)**		**(3)**	
	reading	**math**	**reading**	**math**	**reading**	**math**
teacher experience	0.011***	0.006**	0.013***	0.013***	0.011**	0.009**
	(0.003)	(0.003)	(0.004)	(0.004)	(0.005)	(0.005)
# twin pairs	623	611	301	299	301	299
R-squared	0.039	0.022	0.038	0.079	0.030	0.060
Panel B: sample without split level classrooms	**Total sample: Random assignment specification**		**Longitudinal sample: Random assignment specification**		**Longitudinal sample: Value added specification**	
	(4)		**(5)**		**(6)**	
	reading	**math**	**reading**	**math**	**reading**	**math**
teacher experience	0.013***	0.005	0.017***	0.005	0.017***	0.006
	(0.004)	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)
# twin pairs	299	289	186	184	186	184
R-squared	0.055	0.072	0.070	0.103	0.128	0.191

Notes: Each column shows the results of an OLS-regression. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. In columns (2) and (3) the sample is used for which teacher experience from grade 2 is available. In columns (4)-(6) the sample is used that excludes split level classrooms. In column (4) and (5) the full specification is used as in columns (3) and (4) in table 2.IV. In column (6) also previous test scores have been included.

2.6 Does the effect of teacher experience reflect on the job training?

The main finding from the previous sections is that students who are allocated to classes with more experienced teachers perform better than students who are allocated to classes with less experienced teachers. We have observed that the effects of teacher experience are not affected by other classroom factors, which suggest that these results come from teachers and their qualities. In addition, the teacher experience estimates do not change when other teacher characteristics are included, suggesting an important role for teacher experience and everything else that is correlated with it. Although we are quite confident that this finding is not driven by non-random selection of students to more experienced teachers, the interpretation of this result is not immediately clear because the randomization that we exploit is about classrooms and not about teachers with different qualities. Hence, it is not clear whether the experience effect reflects the effect of training on the job or whether the experience effect is the result of unobserved teacher qualities that are correlated with obtaining more experience in teaching. As noted in previous studies (Chetty et al., 2011; Rockoff, 2004; Kane et al., 2006; Harris and Sass, 2011), the estimated effect of teacher experience might be driven by different mechanisms. First, the estimated effects might be the result of training on the job, which we label as the causal effect of teacher experience. Second, the results might be driven by positive or negative selection of teachers in the education sector. For instance, teachers that are more (less) skilled or motivated and committed might be more (less) likely to stay in the education profession. In fact, Wiswall (2013) finds evidence for negative selection of teachers in American public schools. Third, more experienced teachers might have had a better teaching education and therefore might be more skilled (Corcoran, Evans, and Schwab, 2004; Hoxby and Leigh, 2004; Bacolod 2007). If the quality of teacher education has deteriorated over time teacher experience will be correlated with the quality of teacher education. Fourth, selection into the teaching profession might have changed over time due to a changing labor market. The increasing demand for higher educated workers will probably have increased the number of alternatives for working in the teaching profession. Over the years the teaching profession might have attracted weaker teachers.

The effect of on the job training (the causal effect of teacher experience) can be isolated from unobserved quality differences across teachers by using multiple years of information of teachers (Rivkin et al., 2005; Wiswall, 2013). Unfortunately, we cannot apply this approach because the panel character of our data only relates to students and not to teachers. However, we can empirically explore the plausibility of the various mechanisms by looking at changes in the estimated effects over time or changes in the effect of experience over the teaching career. We start by looking at changes in the estimated effects over time. These changes are informative about the last two mechanisms which both state that older cohorts of teachers had more quality than younger cohorts. If we assume that the causal effect of experience on student performance does not change over time we expect that the estimated effect of teacher experience will decline over time because of older cohorts of teachers leaving the teaching profession. We can put these mechanisms to the test by exploiting the panel character of our data. We have constructed three time periods from the six waves of our panel and included interaction variables between these time periods and teacher experience in our main model.[13] The estimated effects of the interaction

13 Using all six periods separately would strongly reduce the number of observations for several periods.

variables show whether the experience effect has changed over time. Table 2.VIII shows the estimation results. We do not observe that the estimated effects decline over time as might be expected from the last two mechanisms. Hence, the evidence is not consistent with these explanations of the teacher experience effect.

Table 2.VIII: Twin-fixed effect estimates of teacher experience on student test scores by PRIMA survey year (1994-2004)

	reading		math	
Independent variables:	**(1)**	**(2)**	**(3)**	**(4)**
teacher experience	0.012	0.008	0.009	0.004
	(0.008)	(0.012)	(0.007)	(0.008)
teacher experience*year	0.000		0.000	
	(0.001)		(0.001)	
teacher experience*dummy=1 if survey years are 1998 or 2000		0.003		0.005
		(0.012)		(0.009)
teacher experience*dummy=1 if survey years are 2002 or 2004		0.003		0.001
		(0.012)		(0.009)
# twin pairs	623	623	611	611
R-squared	0.043	0.045	0.031	0.030

Notes: Each column shows the results of an OLS-regression using the random assignment specification. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.In columns (1) and (3), the omitted category of teacher experience is survey year 1994 (i.e. year=0 if survey year=1994, year=2 if survey year=1996 and so on). In columns (2) and (4) the omitted category of teacher experience is the dummy that equals 1 for survey years 1994 or 1996. In all columns the full specification is used as in columns (3) and (4) of table 2.IV.

We further attempt to purge the effect of unobserved quality differences from the effect of on the job training by looking at the experience effect of teachers that differ in mobility. If the effect of teacher experience is driven by unobserved teacher quality that is correlated with obtaining experience, then we expect that the bias from unobserved teacher quality will be smaller for teachers that are less likely to leave the profession. Because teacher mobility is highest in the initial years in the profession we use a model specification that can pick up differences in the effect of teacher experience during the career. First, we have included an interaction variable between the minimum teacher experience of both teachers of the twin pair and the difference in teacher experience. Second, we have constructed three categories for minimum teacher experiences and interacted these variables with the difference in teacher experience. The interaction variables measure whether the estimated effect of teacher experience changes over the career. We expect that the estimates for

Table 2.IX: Twin-fixed effect estimates of teacher experience on student test scores by minimum experience of teachers of both twins

	reading		math	
Independent variables:	**(1)**	**(2)**	**(3)**	**(4)**
teacher experience	0.012***	0.011***	0.007**	0.005*
	(0.004)	(0.004)	(0.003)	(0.003)
teacher experience*minimum teacher experience of both teachers	-0.000		-0.000	
	(0.000)		(0.000)	
teacher experience*dummy=1 if minimum teacher experience between 5 and 16 years		-0.000		0.003
		(0.007)		(0.006)
teacher experience*dummy=1 if minimum teacher experience 17 years or more		0.000		-0.007
		(0.010)		(0.009)
# twin pairs	623	623	611	611
R-squared	0.046	0.046	0.023	0.026

Notes: Each column shows the results of an OLS-regression using the random assignment specification. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. In columns (2) and (4) the omitted category is the dummy that equals 1 if the minimum years of experience of both teachers lie between 0 and 4. In all columns the full specification is used as in columns (3) and (4) of table 2.IV.

later career stages will be less likely to be biased by positive or negative selection. Hence, these estimates should be a better approximation of the effect of training on the job than estimates for early career stages. Table 2.IX shows the estimated effects for these specifications. The main result for both subjects is that the estimated effects for these interaction variables are all very small and statistically insignificant. This suggests that the teaching experience effect is quite constant during the teaching career. Hence, we also find returns to teaching experience for career stages in which we expect less bias due to positive or negative selection. The estimates of teaching experience for later career stages are expected to be a better approximation of the effect of on the job training in education. This finding is consistent with recent studies that also find evidence for the importance of teacher experience beyond the initial years in the career (Wiswall, 2013; Harris and Sas, 2011).

2.7 Conclusion

In this chapter we have examined the causal link that runs from classroom quality to student achievement by applying a new identification strategy. This strategy is based on an exogenous assignment of twins to different classrooms. Teacher quality is expected to be the main factor differing across classes. The main findings of this chapter are related to teacher quality. We find that teacher experience is the only observed teacher characteristic that matters for student performance. This finding is consistent with previous studies on teacher effects (Hanushek, 2010; Staiger and Rockoff, 2010, Chetty et al., 2011). Twins who are assigned to classes with more experienced teachers perform better and the effects are most pronounced in kindergarten and early grades. Krueger (1999) and Chetty et al. (2011) report similar results from their analysis using data from the STAR-experiment on class size reduction. Our estimates also suggest a linear effect of experience on student achievements. Until recently there was a consensus in the literature that teacher experience only matters in the initial years in the career (e.g. Rivkin et al., 2005; Rockoff, 2004; Staiger and Rockoff, 2010). However, recent studies also find gains from teacher experience in later years of the career (Harris and Sass, 2011; Wiswall, 2013). Moreover, previous studies based on the data from the STAR-project report linear effects of teacher experience (Krueger, 1999; Chetty et al., 2012). Hence, our estimates corroborate the recent findings about later returns to experience.

From our analysis we learn that teacher experience is very important but it is not clear how we should interpret this finding because our estimates only show that students do better in classes with more experienced teachers. It remains unclear whether this effect is caused by training on the job or reflects the effects of unobserved teacher quality correlated with attaining more experience in education. We have explored the plausibility of various mechanisms that might explain the robust finding that students in classrooms with more experienced teachers perform better. We do not find evidence consistent with mechanisms that stress the importance of changes over time such as changes in the quality of teacher education or changes in outside opportunities in the labor market. However, we find that teacher experience also matters for career stages with less labor market mobility. As it is less likely that these estimates will be biased by selection into or out of the teaching profession this suggests positive returns to on the job training of teachers. This finding is consistent with recent studies that also find positive returns to teacher experience for later career stages (Wiswall, 2013; Harris and Sas, 2011).

The main finding of this chapter is that experienced teachers are very important for student performance. Although we are unable to isolate the effect of on the job training from the effect of unobserved teacher quality this finding has important policy implications. More focused policies that maintain experienced teachers in the classroom appear beneficial, especially for younger students.

3

How much do children learn in school? International evidence from school entry rules[14]

14 This is joint work with Dinand Webbink.

Abstract

This study provides the first estimates of the causal effect of time in school on cognitive skills for many countries around the world, for multiple age groups and for multiple subjects. These estimates enable a comparison of the performance of education systems based on gain scores instead of level scores. We use data from international cognitive tests and exploit variation induced by school entry rules within a regression discontinuity framework. The effect of time in school on cognitive skills strongly differs between countries. Remarkably, we find no association between the level of test scores and the estimated gains in cognitive skills. As such, a country's ranking in international cognitive tests might misguide its educational policy. Across countries we find that a year of school time increases performance in cognitive tests with 0.2 to 0.3 standard deviations for 9-year-olds and with 0.1 to 0.2 standard deviations for 13-year-olds. Estimation of gains in cognitive skills also yields new opportunities for investigating the determinants of international differences in educational achievements.

3.1 Introduction

Many studies have found a strong association between the economic outcomes of nations and their cognitive skills (e.g. Hanushek and Woessman, 2008). It is therefore important to study international differences in the production of cognitive skills, and to examine how much children learn in school and whether this differs between countries. International tests, such as PISA, TIMSS or PIRLS, measure differences in cognitive skills of students between countries. The outcomes of these tests are increasingly used for the benchmarking of education systems and for designing educational policies.[15] However, it is difficult to investigate how much children learn in school because of the complex nature of the production of human capital. In the economic literature that investigates the so-called educational production function, student achievement at any point in time is typically seen as a cumulative result of the entire history of all inputs, for instance from family, peers, teachers and school, and the individual's ability (Hanushek and Rivkin, 2006). The multitude of observed and unobserved factors that might be important pose challenges for identifying the effect of time in school on cognitive skills and for assessing the performance of a country's education system. Previous studies in economics have addressed these challenges by applying quasi-experimental designs for estimating the effect of completed schooling (Cascio and Lewis, 2006; Hansen et al., 2004), pre-primary education (Berlinski et al., 2009; Gormley and Gayer, 2005; Leuven et al., 2010) or grade retention (Jacob and Lefgren, 2009) on cognitive skills for specific countries and specific age groups. To our knowledge, however, previous studies in the economic literature have not attempted to identify the effect of spending one additional year in school on cognitive skills across countries, age groups and subjects enabling comparisons between countries. Moreover, the recent literature that investigates the determinants of international differences in educational achievement has mainly focused on identifying cross-country associations (Hanushek and Woessmann, 2011).[16]

15 For instance, Germany, Denmark and Japan have experienced a 'PISA-shock' that resulted in a range of educational reforms. Lower-than-expected results triggered intense public and political debate on educational performance (Breakspear, 2012). TIMSS and PIRLS results have been used to inform policy considerations in for example Hong Kong, Norway, New Zealand, The Russian Federation and The Republic of South Africa. Participating countries use TIMSS and PIRLS for establishing achievement goals and standards for educational improvement, stimulating curriculum reform, and improving teaching (IEA, 2011).

16 Some recent studies apply a quasi-experimental approach for investigating specific factors such as the effects of class size (Woessmann and West, 2006), central exams (Jürges et al., 2005),

This study provides the first estimates of the effect of time in school on cognitive skills for many countries around the world, multiple age groups and multiple subjects which enable a comparison of the performance of education systems based on gain scores instead of level scores. We use data from international cognitive tests and exploit variation in time in school induced by school entry rules.[17] Students born in adjacent months are assigned to different grades due to these school entry rules. As a result, students who are almost the same age differ in their time spent in school. This provides the opportunity to isolate the effect of time spent in school from the effect of time spent outside of school.[18] We apply this framework for estimating the effect of spending one year in school for samples of countries that participated in international cognitive tests. Within this framework we also address issues, such as sampling bias and violations of the exclusion restriction, that have been neglected in previous studies that exploit variation induced by school entry rules (see Section 3.2).

This framework enables us to perform four types of empirical analyses. First, we estimate the average effect of one year of school time on student performance in math and science. This yields estimates across countries, for two age groups and for two subjects. Second, we are able to estimate the gains in cognitive skills for each separate country. These estimates capture the gain in student achievement from the last year in school before the test was taken which can be interpreted as a measure of the performance of the education system. Third, we rank countries based on this measure of performance and compare this ranking with the ranking based on the level of the scores in international cognitive tests that is currently used for benchmarking of education systems. Fourth, this framework provides new opportunities for investigating the determinants of international differences in student achievement. We illustrate this by examining the effect of external exit-exams on student achievement using a specification that yields estimates of gains in

relative age (Bedard and Dhuey, 2006) or private school competition (West and Woessmann, 2010) using data from international cognitive tests.

17 School entry cut-off dates have also been used for investigating the effects of relative age (Bedard and Dhuey, 2006), or the effects of school starting age on student performance (Black et al., 2011; Fredriksson and Öckert, forthcoming) and the effects of education on earnings (e.g. Angrist and Krueger, 1991).

18 This approach was introduced by development psychologists for separating schooling and age effects on test scores in a regression discontinuity framework (e.g. Baltes and Reinert, 1969; Cahan and Davis, 1987; Cahan and Cohen, 1989) and was recently applied in the economic literature for estimating the effect of completed schooling (Cascio and Lewis, 2006) or the effect of early childhood education (Gormley and Gayer, 2005) on cognitive skills.

achievement, and compare these results with previous studies that used a control strategy for estimating cross-country associations.

For applying this framework data are needed that include students in adjacent grades that took the same test in the same period. The data collected in the 1995 TIMSS study offer the opportunity to apply this framework.[19] In the TIMSS study 9-year-olds and 13-year-olds were tested in math and science. The achievement tests were based on a curriculum framework developed through an international consensus-building process by all participating countries. For the analysis we only use data from countries that apply clear nationwide school entry rules; 21 countries for the 9-year-olds and 34 countries for the 13-year-olds.

Our empirical results can be summarized in three main findings. First, we find large differences in the effect of time in school on student learning between countries for both subjects and age groups. Some countries produce high gains in cognitive skills whereas in other countries additional time in school does not increase cognitive skills. Countries that achieve higher gains in cognitive skills for math also achieve higher gains in science. Moreover, countries with higher gains for 9-year-olds also have higher gains for 13-years-olds. Across countries we find that time in school on average matters for student performance in international cognitive tests. A year of school time increases performance in cognitive tests with 0.2 to 0.3 standard deviations for 9-year olds and with 0.1 to 0.2 standard deviations for 13-year olds. Hence, the effect of time in schools seems to reduce with age. This might indicate that later grades add less to the knowledge base or that the tests do a poorer job at measuring the full range of skill differences.

Second, and most remarkable, we find no association between the estimated gains in achievement and the level of test scores of countries. At all levels of test scores we observe countries with high achievement gains and countries with low gains in achievement. The lack of association has been found for both tests (math, science) and for both age groups (9-year olds and 13-year olds). This implies that assessments of the performance of education systems based on the estimated gains in achievement often are inconsistent with performance assessments based on level scores, and raises concerns about the current use of the outcomes of international cognitive tests in educational policy. A mere focus on test score levels is likely to yield misleading information about the performance of the education system. Low levels of test scores, or declining trends in test scores, could not be the result of low performing education systems. High levels of test scores could mask low performing education systems. Our estimated gains in achievement tell a different story about

19 More recent TIMSS studies only sample students in one grade.

the performance of education systems. Using these gain scores as an additional instrument for the assessment of the performance of education systems is likely to reduce the risk of providing misleading policy information. The estimation of gain scores, which becomes possible when the current collection of international data is extended towards samples of students in adjacent grades, is likely to improve decisions on educational policies.

Third, the estimation of gain scores can be important for investigating international differences in educational achievement. For instance, studies that use control strategies have consistently found that students perform better in countries with external exit exams (Hanushek and Woessmann, 2011; Bishop, 1997; Woessmann, 2001, 2003). However, we do not find higher gains in achievement in countries with an external exit exam for 9-year-olds and 13-year-olds compared to countries that do not have external exit exams.

This chapter makes several contributions to the current economic literature. First, we contribute to the literature on the educational production function by applying a method for measuring gains in achievements across countries. To our knowledge no previous study has estimated causal effects of time in school for different countries using a quasi-experimental approach. This method produces estimates of gains in achievement by different education systems, which enable a comparison of the performance of education systems based on gain scores instead of level scores. We compare the assessment of the performance of education systems based on the estimated gains with the performance assessments based on level scores. This comparison reveals that educational policies solely based on test score levels are potentially misguided because they ignore the gains that have been achieved.

Second, we contribute to the literature that investigates the effect of time in school on student performance (e.g. Gormley and Gayer, 2005; Hansen et al., 2004; Cascio and Lewis, 2006; Berlinski et al., 2009; Leuven et al., 2010). We add to this literature by investigating the effect of time in school across countries, age groups and subjects. Third, previous studies that exploited variation induced by school entry rules have neglected various issues that might bias the estimates, such as sampling bias, relative age effects or violations of the exclusion restriction. In this chapter we explicitly address these problems. In particular, we interpret our main estimates as lower bounds and also generate estimates that are adjusted for sampling bias. Fourth, we contribute to the literature that uses international cognitive tests for investigating the determinants of international differences in educational achievements. The typical features of human capital production pose major challenges for the identification of the effect of characteristics of education system and the cross-sectional structure of the international tests hinders value-added or panel estimations. Therefore, it has been argued that 'further exploration of quasi-experimental settings in the international data should be high on the agenda' (Hanushek and Woessman, 2011). This is exactly what this chapter does. We apply a quasi-experimental approach using international data and this approach might yield new opportunities for identifying the effects of characteristics of education systems. We illustrate this by comparing estimates of the effect of external exit exams using level scores with the estimates based on gain scores.

This study is organized as follows. Section 3.2 explains the empirical strategy used for estimating the effect of one year of school time on test performance. The data used in the analyses are described in Section 3.3. Section 3.4 shows the estimates of the effect of one year of time in school for pooled samples of countries. In Section 3.5 differences between countries are investigated. Section 3.6 compares country rankings of level scores with country rankings of gain scores. Section 3.7 illustrates the opportunities of the quasi-experimental approach for investigating international differences in student achievements and Section 3.8 concludes.

3.2 Previous studies and empirical strategy

The basic framework in the economic literature that studies the effects of educational inputs models student achievement as a function of family, peer, community, teacher and school inputs and student ability (Hanushek and Rivkin, 2006). Student achievement at any point in time is seen as a cumulative result of the entire history of all inputs and the individual's initial endowment (e.g. innate ability). A common approach for modeling this so-called educational production function is to assume that the cumulative achievement function is additively separable and linear (e.g. Boardman and Murnane, 1979; Todd and Wolpin, 2003). Estimating the effect of input factors, such as time in school, is complicated because in any actual application we will generally not be able to control for all relevant school, family or student characteristics. If some omitted variables are correlated with time in school, then the estimated parameters will be biased. Hence, the cumulative character of the production of human capital poses challenges for identifying the effect of time in school.

Previous studies in economics have addressed these challenges by applying quasi-experimental designs for estimating the effect of schooling on cognitive skills.[20] The effect of schooling has been analyzed from different perspectives. A first strand of the literature

20 For surveys of the development psychology literature on the estimation of schooling effects, see Ceci (1991) and Stipek (2002). In addition, there are studies in the education literature about the effects of schooling, see for instance Luyten (2006).

focuses on the effect of completed schooling on cognitive skills. Several studies have used quarter of birth as an instrument for completed schooling (Neal and Johnson, 1996; Hansen et al., 2004) as in the seminal paper by Angrist and Krueger (1991). These studies find that one additional year of completed schooling increase cognitive skills with approximately 0.2 standard deviations. A recent study investigates the effect of an increase of compulsory schooling by one or two years on cognition (Meghir et al., 2013). They find that the reform increased cognitive skills on average, with 7 to 10 percent of a standard deviation. Cascio and Lewis (2005) exploit variation induced by school entry rules for estimating the effect of completed schooling on cognitive skills measured by the Armed Forces Qualification Test (AFQT). They find that an additional year of high school raised scores of minorities with 0.3 standard deviations.

A second strand of the literature focuses on variation in schooling from pre-primary education.[21] For instance, Gormley and Gayer (2005) and Gormley et al. (2005) estimate the impact of Oklahoma's pre-K program for 4-year-olds in Tulsa on cognitive/knowledge test scores, motor skills and language scores by exploiting cutoff requirements for enrolment in pre-K. Attendance increases test scores by approximately 0.4 standard deviations.[22] A third strand of the literature focuses on grade retention. For instance, Jacob and Lefgren (2009) estimate the effect of grade retention on high school completion by exploiting a nonlinear relationship between current achievement and the probability of being retained. They find that retention among sixth-grade students does not affect the likelihood of high school completion, but retention of eight-grade students increases high school dropout.[23] Our study is also related to a fourth strand of the literature which uses so-called value-added models for estimating gains in student achievement or the rate of learning over specific time periods. These models include measures of prior achievement to eliminate confounding by past unobserved parental and school inputs, for instance for estimating teacher fixed effects which can be linked to teacher characteristics (e.g. Rivkin et al., 2005; Hanushek et al., 2005). Dynamic sorting of teachers and students might bias the estimated effects in these models (Rothstein, 2010). Our approach also focuses on the estimation of gains in cognitive skills but uses a quasi-experimental approach instead of controlling for prior achievements.

21 Early childhood interventions like Head Start or the Perry Preschool Project have been studied intensively. For surveys, see Currie (2001) and Almond and Currie (2010).

22 For other recent studies, see Berlinsky et al. (2009) and Leuven et al. (2010).

23 For other recent studies, see Manacorda (2012) or Schwerdt and West (2012).

Empirical strategy

In this chapter we focus on estimating the effect of time in school on cognitive skills. For identifying the effect of time in school we use a quasi-experimental design that was first applied by development psychologists (e.g. Balter and Reinert, 1969; Cahan and Davis 1987; Cahan and Cohen, 1989) and recently also applied in economic studies (Cascio and Lewis, 2006; Gormley and Gayer, 2005).[24] The key idea for identification is that school entry rules create variation in time in school for children born close to the cut-off date. Students who are almost the same age differ in their time spent in school. A comparison of the test scores of students around this cut-off date yields estimates of the effect of one school year. In this chapter we apply this approach to samples of countries that participated in international cognitive tests. Figure 3.1 illustrates the approach using scores from the math and science tests of the 1995 TIMSS study for 9-year-olds in two adjacent grades.[25] The top panel shows results for Singapore, the bottom panel shows results for England. The left figure shows the assignment of students to grades on both sides of the cut-off date of the school entry rule; the middle (right) figures show the scores on the math (science) test. The horizontal axis shows the age of the student relative to the cut-off date. Each dot represents a monthly average of the grade-level or the test score.

The left figures show that both countries quite strictly apply the school entry rule for assigning students to grades. Nearly all students to the left of the cut-off date are in the higher of the two adjacent grades and nearly all students to the right of the cut-off date are in the lower of the two grades.[26] In both countries we also observe that scores in math and science decline with age which confirms previous findings about the importance of age at entry for test performance (Bedard and Dhuey, 2006). The cut-off date divides students of very similar age into groups that differ in the number of years they have spent in school. Students on the left hand side of the cut-off date have spent one more year in school than students on the right hand side. For students from Singapore we observe a discontinuity in the math and science scores around the cut-off date. This discontinuity can be interpreted

24 Also Luyten (2006) exploits a quasi-experimental design to estimate the effects of extra schooling with the same data used in this study. We differ from this study in two important ways: 1) we estimate the effects of additional schooling for more countries and 2) we use reduced form estimates of the effect of being born left of the cutoff on student achievement instead of OLS-estimates of the effect of grade level on student achievement.

25 It should be noted that these grades also include 8-year-olds and 10-year-olds.

26 The first stage estimates (equation (3.2)) for Singapore and England respectively are 0.96 and 0.93. Section 3.5 presents the first stage estimates of all countries used in the estimations.

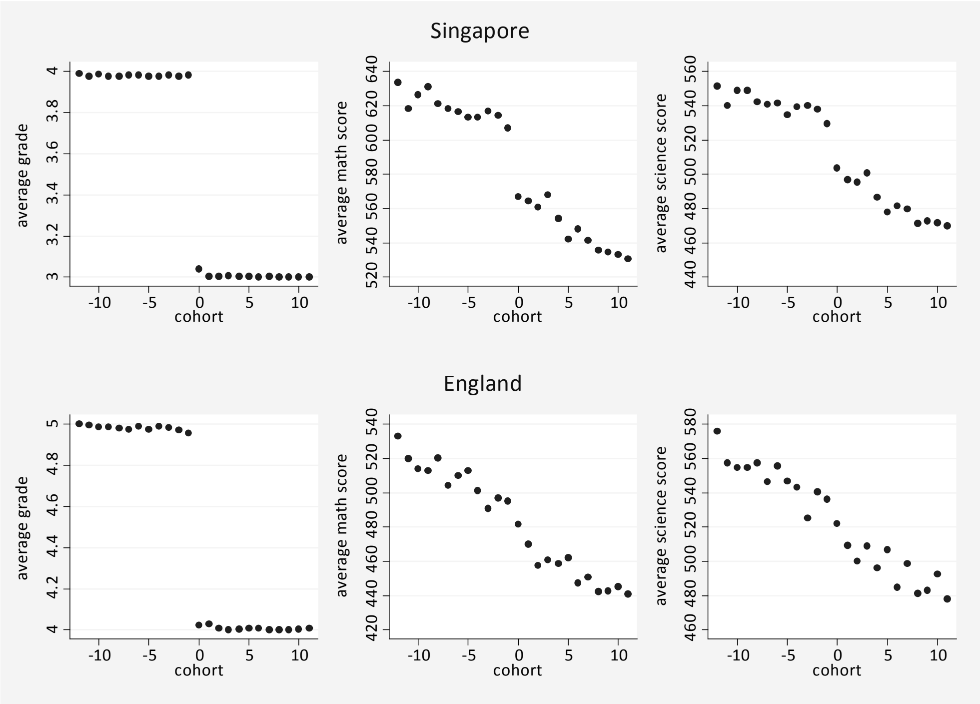


Figure 3.1: Grade level and math/science scores around the cut-off date for 9-year-olds from Singapore and England (TIMSS 1995)

Notes: Each dot represents a monthly average of the grade level or the test score. Students born in month 0 are born in the first month after the cutoff. See table 3A.1 for the cutoff dates per country.

as the effect of one year spent in school in Singapore. For students in England we do not observe a discontinuity in test scores. This suggests that one year spent in school in England does not add more to the performance of students in math and science measured in the 1995 TIMSS study than one year spent out of school.

Estimating the effect of time in school by exploiting school entry rules

In a situation of full compliance with the school entry rules, the effect of one year in school on student performance can be estimated by using a regression discontinuity model that exploits the discontinuities created by the entry rule. The basic assumption in this model is that students on both sides of the discontinuity are very similar and that the relationship between date of birth and student performance is smooth around the discontinuity.[27] For each country, the effect of one year in school can be estimated using the following specification:

(3.1) $$Y_i = \alpha_0 + \alpha_1 G_i + f(birthmonth - C) + \alpha_2 X_i + \varepsilon_i$$

where Y_i is the student performance of student i, G_i is a dummy variable for being in the higher grade, $birthmonth_i$ is the month of birth of the student, C is the cut-off date of the country, X_i is a vector of control variables and ε_i are unobserved factors. In this specification $f(.)$ is a smooth function of age which is allowed to be different at either side of the cut-off (f_l and f_r), as suggested by Lee and Lemieux (2010):

$$f(birthmonth_i - C) = f_l(birthmonth_i - C) + S_i[f_r(birthmonth_i - C) - f_l(birthmonth_i - C)]$$

The main parameter to be estimated is α_1 which can be interpreted as the effect of one year of school time on the test performance. Identification of α_1 is based on the non-linear relationship between age and time in school around the cut-off date.

A concern with this approach is non-compliance with the school entry rules. The grade level of a student (G) might differ from the time spent in school because of retention or acceleration, or because of schools that do not comply with the country's school entry rule. In that case, equation (3.1) would probably yield biased estimates of the effect of time in school because it is likely that students who deviate from the regular path are not a random draw from the population. This problem has been recognized in development psychology and in economics. Studies in development psychology have dealt with this problem by excluding non-compliers (e.g. Cahen and Cohen, 1989). This creates, however, a non-random sample that might induce biased estimates. Studies in the economic literature on schooling or starting age have often dealt with non-compliance by using an instrumental variable approach in which the school entry rule is used as an instrumental variable for the grade level (e.g. Cascio and Lewis 2006; Bedard and Dhuey 2006). In this approach the variation in time in grade that is induced by the school entry rule is used for estimating the

27 Cascio & Lewis (2006) exploit variation in school-entry dates across states in the USA and use individuals in other states as controls. With this approach they don't need to assume that relationship between date of birth and student performance is smooth.

causal effect of time in a specific grade on cognitive skills. The first stage and second stage equations can then be estimated using Two Stage Least Squares (2SLS):

(3.2) $$G_i = \beta_0 + \beta_1 S_i + f(birthmonth_i - C) + \beta_2 X_i + \eta_i$$
$$S_i = 1[birthmonth_i < C]$$

(3.3) $$Y_i = \gamma_0 + \gamma_1 \hat{G}_i + f(birthmonth_i - C) + \gamma_2 X_i + \vartheta_i$$

where S_i is a dummy variable for being born on the left side of the cutoff date, which is equivalent to being eligible for one extra year of time in school. Estimation of γ_1 will yield the causal effect of time in grade if the usual IV-assumptions hold (see below). This estimate can then be interpreted as the effect of time in grade for those students who move to the next grade if their expected time in school, due to the school entry rule increases by one year. Hence, for students who follow the regular path through education without deviations such as retention or acceleration. The estimate of the effect of time in school on grade level (β_1) in the first stage equation indicates the proportion of the students of a specific country that stays on the regular track of the education system of that country. For applying this IV-approach three assumptions should hold. First, the school entry rule should have an effect on the grade level of students. Hence, there should be no weak instrument problem. The empirical analysis in the next sections shows that in all selected countries the school entry rule is an important determinant of the observed grade level. The second assumption is that the cut-off date should not be correlated with unobserved determinants of cognitive skills. We will address this assumption below. The third assumption, which is neglected in previous studies on schooling or starting age, is the exclusion restriction; the instrument should only have an effect on cognitive skills through the endogenous variable. In our application this means that the difference in cognitive skills between students born on either side of the cutoff date should only be the result of the time spent in the highest grade by students who are on track. However, all students on the left side of the cutoff have been treated with an additional year in school; a year in a higher grade or a year of being retained. Hence, it is assumed that grade retention or acceleration of students has no effect on cognitive skills. Given the recent studies on grade retention (see above) it seems not likely that this assumption holds. We, therefore, focus our analysis on estimating the reduced form of this IV-approach:

(3.4) $$Y_i = \delta_0 + \delta_1 S_i + f(birthmonth_i - C) + \delta_2 X_i + \varepsilon_i$$
$$S_i = 1[birthmonth_i < C]$$

For the identification of the effect of time in school on cognitive skills in equation (3.4) several further issues are important. First, school entry rules not only induce a difference in the time spent in school for students of nearly similar age but also induce a difference in relative age in class (school starting age). Students on the left of the cut-off not only receive an additional year of education but are also assigned to be the youngest in their grade. Students on the right of the cut-off are assigned to be the oldest in their grade. Differences in relative age have been shown to be important for short-term and long-term cognitive outcomes (Bedard and Dhuey, 2006). We address this identification issue by using a model specification that allows the effect of the assignment variable age to be different at either side of the cutoff. Age and relative age are perfectly correlated because both are measured from the cut-off date. This means that in our specification the age effect on both sides of the cut-off not only controls for maturity but also for relative age in grade.[28]

Second, some countries apply a clear school entry rule but also use a rolling admission of students. For these countries the school entry rule does not create a one-year difference in time in school, but leads to a different timing of grade promotion. Hence, for these countries students on the right side of the cut-off date have spent more time in lower grades than students on the left side of the cut-off date. We address this issue in the sensitivity analysis in which we exclude countries with rolling admission from the estimation sample.

Third, equation (3.4) yields the causal effect of one year of school time on cognitive skills if the conditional independence assumption holds. Hence, the critical assumption is that students near the cut-off date are very similar on observed and unobserved characteristics. This assumption seems plausible since parents are unlikely to plan the exact date of birth of their child. However, there is evidence that parents in the U.S. schedule births in order to avoid taxes (Dickert-Conlin and Chandra, 1999). Several recent studies have investigated whether birth around the school entry cut-off dates is random.[29] For the US (Dickert-Conlin and Elder, 2010; McCrary and Royer, 2010), Chile (McEwan and Shapiro, 2008) and Argentina (Berlinski et al., 2011) no evidence has been found for the non-randomness of births around cut-off dates. However, the timing of births in Japan seems to be related with school entry cut-off dates (Shitgeoka, 2013). The number of births sharply increases in the first

28 Bedard and Dhuey (2006) estimated the effects of age relative to the cut-off data and frame the estimates in terms of relative age. These estimates are the combined effect of maturity and age at entry. Black et al. (2011) isolate the effect of these two variables.

29 Several studies have raised concerns about the randomness of season of birth (Bound and Jaeger, 2000; Cascio and Lewis, 2006; Dobkin and Ferreira, 2010; and Buckles and Hungerman, 2012).

days after the cut-off date. Hence, some Japanese parents seem to have a preference for their children to belong to the oldest in class. This might induce a bias for the estimated effect because it is not clear which parents try to postpone the birth of their child. To address this issue we exploit the fact that our data contains information about the exact date of birth. We will perform sensitivity tests by using estimation samples in which we exclude students born on the first days around the cut-off date (see Section 3.4).

Fourth, a further and related issue, which is not addressed in previous studies that use school entry rules, is sampling bias. Our sample consists of students in two adjacent grades that contained the largest proportion of students from a specific age group; 9-year-olds or 13-year-olds (see next section). The disadvantage of this sampling strategy is that that we do not observe students from these age groups that are not in these grades. If we imagine a country in which the 9-year-olds are evenly distributed over the two adjacent grades, then the higher grade will contain the oldest 9-year-olds (group B) together with the youngest 10-year-olds (group A), and the lower grade contains the youngest 9-year-olds (group C) together with the oldest 8-year-olds (group D). Groups A and B are on the left side of the cutoff in figure 3.1 and, groups C and D are on the right side of the cutoff. For our main estimation sample we use students from group B and group C, and we compare the difference in performance of these two groups at the cut-off. However, in group B we do not observe students who have been accelerated, and in group C we do not observe students who have been retained. It might be expected that this will induce a downward bias for the estimated effects because students who have been accelerated will probably have a relatively high ability, and students who have been retained will probably have a relatively low ability.[30] This implies that the estimated effects should be interpreted as lower bound estimates. It should be noted that the 'missing students' in our estimation sample are the students who, because of their relative age in grade, are the least likely to be accelerated or retained. To further address this issue we will perform two types of sensitivity analysis. First, we will estimate the main models for samples of countries in which most students are on track; countries with a first stage estimate (equation (3.2)) of at least 0.75.[31] In these countries only a very small proportion of students will not be observed. Second, the advantage of

30 In table 3A.4 it can be observed that retained students born at the left side of the cut-off on average score lower than students that are on track and, that accelerated students born at the right side of the cut-off on average score higher than students that are on track.

31 This first stage estimate should not be directly interpreted as the proportion of missing students. The missing students can be observed in groups A and D, the first stage estimate is based on group B and C.

the sampling strategy is that we also have data of students in groups A and D which we can exploit to approximate the sampling bias of the 9-year-olds in groups B and C. In group A, which contains the youngest 10-year-olds, we can observe students who have been retained. We use these students to adjust for sampling bias in group C, which contains the youngest students of the 9-year-olds. Moreover, in group D, the oldest 8-year-olds, we can observe students who have been accelerated. We use these students to adjust for sampling bias in group B, which contains the oldest students of the 9-year-olds. By assuming that the proportion of retained and accelerated students, and the relative score of these retained or accelerated students compared to students who are on track does not change between grades, we can adjust their scores and include them in the main estimation sample. Hence, for our approximation of the sampling bias we adjust the scores of some students from groups A and D, and include them in the main estimation sample consisting of students in groups B and C. We use these samples to obtain estimates that are adjusted for sampling bias (the Appendix provides further details about this procedure). For the main models we will show the lower bound estimates and the estimates that are adjusted for sampling bias.

3.3 Data

The data used in this study come from the 1995 TIMSS study.[32] The 1995 TIMSS study collected mathematics and science achievement results from third and fourth graders in 26 countries and from seventh and eighth graders in 41 countries.[33] These achievement tests are based on a curriculum framework developed through an international consensus-building process by all participating countries. International experts in mathematics, science, and measurement contributed to the development of the achievement tests and the tests were endorsed by all participating countries. The sampling focused on the two adjacent grades that contain the largest proportion of 9-year-olds – third and fourth graders in most countries – or the largest proportion of 13-year-olds – seventh and eighth graders in most countries. These samples also include students who are one year younger or older than the age groups that were targeted. This sampling strategy enables us to apply the regression discontinuity framework that we discussed in Section 3.2. After the 1995 TIMSS

32 See, http://timss.bc.edu/timss1995i/Database.html for TIMSS data.

33 These data were also used in a quasi-experimental study on international differences in class size effects (Woessmann and West, 2006).

study the sampling strategy was changed and focuses only on one grade, which makes it impossible to apply our estimation framework.

From the 1995 TIMSS study we include all countries in the analysis that apply clear nationwide school entry rules. For the nine-year-olds we included 21 out of 26 participating countries. Australia, the USA and Ireland have been excluded because the rules regarding the school cutoff date vary across regions or are at the discretion of educators or parents. Kuwait and Israel have been excluded because in those countries only one grade was sampled or no information on test scores was available. For the thirteen-year-olds we included 34 out of 41 participating countries. Again Australia, USA, Ireland, Kuwait and Israel have been excluded. Columbia has been excluded because there is no clear cut-off date in average grade. The Republic of South Africa has not been included because the teacher and school data were not deemed internationally comparable. Bedard and Dhuey (2006) excluded more countries from their estimation sample because of concerns about the strict application of the school starting age rules or measurement error in the date of birth. They additionally excluded Germany, the Netherlands, Hungary, Switzerland and Korea. In our analysis, which focuses on differences between grades, it seems that these countries can be included because we observe sharp discontinuities in average grade around the cutoff date. We test the robustness of our findings by replicating our main estimations for the sample of countries used by Bedard and Dhuey (2006).

As dependent variables we use the TIMSS test scores in math and science. These scores have been standardized with a mean of 500 points and a standard deviation of 100 points which can be easily translated into the usual effect sizes from a standard normal distribution. TIMSS uses an incomplete or rotated-booklet design for testing children on the major outcome variables. For each student and each test TIMSS selects five plausible values. In the estimation we use all five plausible values and adjust standard errors as recommended when using the plausible values methodology (Von Davier et al., 2009). Our main control variable is the date of birth of the student measured by month. For many countries we also have the exact date of birth, which we will use in the sensitivity analysis. Other control variables that we use are gender, born in country of test, living with mother/father, language of test spoken at home, number of books at home, and mother's and father's educational level.

School entry rules are crucial in our analysis. We use information from Bedard and Dhuey (2006) and several online sources, and empirically checked this information in our data. For some countries we could not obtain information about the cutoff dates. In those cases we used the cutoff date from the data (see table 3A.1 in the Appendix).

3.4 The effect of one year of school time on cognitive skills across countries

This section presents the first part of our empirical analysis. We estimate the average effect of one year of school time for students in different age groups, on the performance in cognitive tests for the pooled samples of countries. This estimate can be interpreted as the effect of spending one year in school across countries, and might be considered as an international benchmark for gains in cognitive skills for specific age groups and subjects.

To obtain estimates of the average effect of one year of school time in all the selected countries from TIMSS we have pooled the data for each test (TIMSS 9, TIMSS 13) and estimated equation (3.4). In this model we have also included country dummies and interactions of these dummies with a linear function of age such that we allow the functional form of age to be different at either side of the cut-off for each country. The critical assumption in applying this model is that students near the cut-off date are very similar on observed and unobserved characteristics. To investigate this assumption we compared the covariates of students born in the months around the cut-off date (see table 3A.2 in the Appendix). In addition, we performed balancing tests, in which observed characteristics are regressed on a dummy for being born at the left side of the cut-off and a function of age (table 3A.3 in the Appendix). For 9-year-olds we find that students on both sides of the cut-off are very similar. For 13-year-olds, however, we find a difference with respect to the educational level of the mother. This difference might be the result of sampling bias which we will address below.

Table 3.I shows the estimation results based on equation (3.4) for both subjects and age groups. In panel A and B we show the estimation results using the TIMSS achievement tests in math and science respectively for 9-year-olds and 13-year-olds. Columns (1) to (4) show the reduced form estimates from equation (3.4). The odd columns only control for age, the even columns also control for gender, born in country of test, lives with mother/father, language of test spoken at home and number of books at home. These columns also report the first stage estimate from equation (3.2) which can be interpreted as the proportion of students that is on track. We use two discontinuity samples around the cut-off date: ± 3 months and ± 6 months. Columns (5) to (12) show the result from various sensitivity analyses. All sensitivity analyses use the sample of students born six month before or after the cutoff date, and include all controls like in column (4). Columns (5) and (6) respectively include a quadratic or cubic function of month of birth. Columns (7) to (9) focus on the

Table 3.I: Estimates of the effect of one year of school time by subject and age group based on pooled regression for all countries

	Reduced form estimates (OLS)				Sensitivity analysis using ± 6 months around the cutoff							
	± 3 months		± 6 months		Functional form		Day of birth sample [column (9) excludes 3 days around the cutoff]			BD-sample	Countries with first stage >0.75	Sampling bias adjustment
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: TIMMS 9 (21 countries)												
math	25.8	25.1	27.0	26.4	29.9	34.5	27.5	27.9	28.2	23.8	29.7	30.4
	(2.2)	(2.2)	(1.5)	(1.4)	(2.1)	(2.8)	(1.5)	(1.4)	(1.5)	(1.9)	(2.0)	(1.4)
N	33615	33615	66803	66803	66803	66803	47782	47782	47001	31874	29861	68869
Coefficient first stage	*0.66*	*0.65*	*0.70*	*0.70*								
science	18.9	18.1	20.0	19.4	21.3	23.2	19.1	19.5	19.4	18.3	21.8	23.5
	(2.1)	(2.0)	(1.4)	(1.4)	(2.0)	(2.6)	(1.6)	(1.5)	(1.6)	(2.1)	(2.1)	(1.3)
N	33615	33615	66803	66803	66803	66803	47782	47782	47001	31874	29861	68869
Coefficient first stage	*0.66*	*0.65*	*0.70*	*0.70*								
Panel B: TIMMS 13 (34 countries)												
math	6.0	6.4	7.0	7.2	7.1	5.7	6.6	6.7	6.5	7.8	10.3	12.3
	(1.7)	(1.6)	(1.1)	(1.0)	(1.6)	(2.4)	(1.2)	(1.5)	(1.2)	(1.3)	(1.7)	(1.0)
N	51908	51908	104316	104316	104316	104316	85161	85161	83811	55651	36760	109942
Coefficient first stage	*0.56*	*0.56*	*0.62*	*0.62*								
science	10.9	11.3	10.8	11.0	11.9	11.4	11.7	11.6	11.6	10.1	17.6	16.4
	(1.5)	(1.4)	(1.0)	(0.9)	(1.5)	(2.3)	(1.0)	(1.0)	(1.0)	(1.5)	(1.8)	(0.9)
N	51908	51908	104316	104316	104316	104316	85161	85161	83811	55651	36760	109942
Coefficient first stage	*0.56*	*0.56*	*0.62*	*0.62*								
Birth month controls	linear	linear	linear	linear	square	cubic	linear	no	no	linear	linear	linear
Birth day controls	no	no	no	no	no	no	no	linear	linear	no	no	no
Additional controls	no	yes	no	yes	yes	yes	yes	yes	yes	yes	yes	yes

Notes: All models include country dummies and interactions of these dummies with the polynomial in birth month/day that differs at either side of the cutoff. The models in columns (2), (4), and (5)-(12) include additional controls. For 9-year-olds we control for gender, born in country of test, lives with mother/father, language of test spoken at home and number of books at home. For 13-year-olds we additionally control for parental education. These variables were not available 9-year-olds. Standard errors in parentheses are adjusted for using plausible values.

sample of countries for which the exact date of birth is available; 17 countries for 9-year-olds, 30 countries for 13-year-olds.[34]

Column (7) uses a linear specification of month of birth like in column (4), column (8) uses the exact date of birth as assignment variable in a linear specification, and column (9) excludes students born three days before or after the cutoff date. Column (10) uses the same sample of countries as used by Bedard and Dhuey (2006). The last two columns address the issue of sampling bias. In column (11) the estimation sample only includes countries in which most students are on track; countries with a first stage estimate (equation (3.2)) of at least 0.75. For the 9-year-olds the sample includes 9 countries; for the 13-year-olds 12 countries are included. Column (12) shows estimates that have been adjusted for sampling bias. The sampling bias is approximated by using data of students who are one year younger or one year older in the two adjacent grades. The standard errors of the estimates are adjusted for using plausible values (Von Davier et al., 2009), which causes a slight increase.

The estimates in column (1) to (4) of panel A of table 3.I show that one year of time in school increases performance of 9-year olds between 25 and 27 points in math and between 18 and 20 points in science, which is between 0.2 and 0.3 standard deviations of test scores (a standard deviation of test scores is 100 points). The estimates are precise, and robust to the discontinuity sample. The inclusion of controls only slightly reduces the estimated effects, which confirms that students born around the cut-off date are quite similar in observed characteristics. Columns (5) to (12) show the results from various sensitivity analyses. Columns (5) and (6) show that including a quadratic or cubic function of age, measured by year and month of birth, slightly increases the estimated effects. The estimates in columns (7) to (9) test the sensitivity of the results for using the exact date of birth as assignment variable and for a potential non-randomness of births around the cutoff. The estimates are very similar if we include the exact date of birth (column (8)) or exclude students born three days before or after the cutoff date (column (9)). Column (10) shows the estimation results for the sample of countries used by Bedard and Dhuey (2006). This sample is more restrictive and also excludes countries with a rolling admission. The estimated effect for this sample remains similar to the results in the other columns. Columns (11) and (12) aim to assess the sensitivity of the results to sampling bias. The estimated effects are 0.03 to 0.04 standard

34 The number of observations for Greece has reduced substantially due to missing values of the day of birth.

deviations higher than those in column (4), which suggests that for 9-year-olds sampling bias is small. For the sample of countries used for the estimations in column (11) the difference between the lower bound estimates and the unbiased estimates is 0.01 standard deviations (not shown in table 3.I). Hence, the lower bound estimates are probably not very different from the unbiased effects of one year of time in school.

The estimates in column (1) to (12) should be interpreted as the effect of one additional year in school at the age of 9. As in previous studies, we can also attempt to estimate the effect of time in grade on cognitive skills by using an IV-approach. The IV-estimates can be obtained as the ratio of the reduced form estimates and the first stage estimates shown in column (1) to (4). If the IV-assumptions hold we would find that one year of time in grade increases the scores in math and science by approximately 40 and 30 points respectively. As mentioned above, this approach assumes that grade retention or acceleration has no effect on cognitive skills.

Panel B shows the effects of time in school for 13-year-olds. For the lower bound estimates in columns (1) to (10) we find that one year of time in school increases performance by 6 to 8 points in math and by 11 to 12 points in science. The estimates of the lower bound effect are robust to the discontinuity sample and to various sensitivity tests. The estimated effects in columns (11) and (12), which are approximations of sampling bias, are larger than the lower bound estimates. The proportion of 13-year-olds that are not on track is larger than the proportion of 9-year-olds, which explains the increase in the difference between the lower bound estimates and the approximations of the unbiased effects. For the sample of countries used for the estimation in column (11) the difference between the lower bound estimates and the unbiased estimate is 0.02 standard deviations (not shown in table 3.I). Hence, for these 12 countries sampling bias is likely to be quite small.

The cross-country estimates of time in school yield three important findings. First, a year of school time matters for the performance of all age groups in math and science. Across countries a year of school time increases performance in cognitive tests with 0.2 to 0.3 standard deviations for 9-year olds and with 0.1 to 0.2 standard deviations for 13-year-olds. These effects are consistent with the results of previous studies based on credible research designs (Gormley and Gayer, 2005; Gormley et al., 2005; Hansen et al., 2004; Cascio and Lewis, 2006; and Berlinski et al., 2009). Second, an additional year of time in school matters more at the age of 9 than at the age of 13. Hence, the effect of time in school seems to reduce with age. This might indicate that later grades add less to the knowledge base or that the tests do a poorer job at measuring the full range of skill differences. Third, the difference between the lower bound estimates and the unbiased estimates seems to increase with the proportions of student that are not on track. For 9-year-olds the lower bounds estimates are likely to be quite close to the unbiased effects. For 13-year-olds this also holds for the sample of 12 countries with high proportions of students that are on track. The difference between the lower bound estimates and the approximations of the unbiased effects are larger for countries in which substantial proportions of students are not on track.

3.5 International differences in gains in cognitive skills

The second step in the empirical analysis is to investigate differences between countries. Which countries produce the largest effects of one year of school time on performance in cognitive tests?

Differences in achievements of 9-year-olds between countries

We start by analyzing the achievement of 9-year-olds. Column (1) of table 3.II shows the reduced form estimates (RF) of the gain in math skills caused by one year of additional school time. This estimate can be interpreted as the effect of spending one year in school in a specific country. The countries are ranked with respect to this estimate. We observe that the education systems of Norway and Singapore have produced the highest gain in achievement in math for 9-year-olds; the lowest gain in achievement in math has been produced by the education systems of New Zealand and Thailand. Column (2) shows estimates of the effect of one year of school time which are adjusted for sampling bias. Column (3) shows the mean level score of the highest of the adjacent grades for each country. Singapore and South-Korea have the highest scores, whereas Iceland and Iran have the lowest level scores in math in the upper grades. We call these average scores the country level scores and in the next section we will compare them with the estimates of the effect of one year of school time. Columns (4), (5) and (6) show the reduced form estimate, the estimate that is adjusted for sampling bias, and the mean level score of the upper grade for the science test. Column (7) shows the first stage estimates (FS) of the effect of being born on the left side of the cutoff date on the grade level. This estimate indicates to which extent a country keeps students on track. For instance, in Singapore, Iceland, Japan and England most students move through the education system in line with the prediction based on the school entry rule. All models control for age (in months) separately specified for both sides of the cut-off and use the sample of students born in the period between six month before

Table 3.II: Reduced form estimates of the gain in cognitive skills and mean upper grade by country for 9-year-olds

		Math			Science				
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ranking	Country	gain	corr. gain	mean	gain	corr. gain	mean	first stage	N
1	Norway	43.3 (7.4)	43.1 (7.4)	502	38.8 (8.5)	38.5 (8.5)	530	0.89 (0.02)	2133
2	Singapore	40.0 (5.2)	41.8 (5.2)	625	27.8 (4.9)	29.7 (4.9)	547	0.96 (0.01)	6986
3	Greece	36.3 (6.7)	38.0 (6.7)	492	21.7 (6.5)	23.2 (6.5)	497	0.89 (0.02)	2981
4	Iceland	34.6 (9.5)	35.9 (9.5)	474	30.9 (8.4)	32.4 (8.4)	505	0.96 (0.01)	1702
5	Iran	33.3 (7.1)	44.8 (6.6)	429	35.9 (6.8)	48.9 (6.4)	416	0.56 (0.04)	2716
6	Cyprus	30.5 (6.4)	33.9 (6.4)	502	20.4 (5.8)	23.2 (5.8)	475	0.79 (0.02)	3230
7	Japan	28.6 (5.4)	28.3 (5.4)	597	19.1 (5.3)	18.9 (5.3)	574	0.94 (0.01)	4343
8	Czech Republic	26.9 (6.9)	30.3 (6.9)	567	16.9 (6.5)	20.5 (6.5)	557	0.43 (0.03)	3108
9	Canada	24.4 (5.1)	29.9 (4.9)	532	21.7 (4.8)	28.7 (4.7)	549	0.70 (0.02)	7436
10	Netherlands	23.6 (6.8)	29.3 (6.9)	577	13.9 (6.7)	17.3 (6.6)	557	0.48 (0.04)	2241
11	Hungary	23.4 (7.8)	29.2 (7.8)	548	26.7 (7.0)	32.2 (7.0)	532	0.41 (0.04)	2743
12	Hong Kong	22.0 (5.6)	26.6 (5.5)	587	16.3 (5.0)	21.0 (4.9)	533	0.69 (0.02)	3851
13	Austria	21.9 (7.4)	27.9 (7.4)	559	11.4 (8.1)	17.6 (8.1)	565	0.55 (0.03)	2315
14	South-Korea	21.3 (6.5)	25.4 (6.2)	611	12.8 (6.4)	17.2 (6.1)	597	0.66 (0.03)	2636
15	Scotland	20.1 (6.6)	21.1 (6.6)	520	12.9 (7.2)	14.0 (7.2)	536	0.77 (0.03)	3089
16	Portugal	17.7 (7.3)	22.6 (7.3)	475	17.9 (8.9)	24.9 (8.9)	480	0.79 (0.03)	2310
17	England	13.5 (7.1)	13.6 (7.1)	513	12.9 (9.2)	13.4 (9.2)	551	0.93 (0.02)	3087
18	Slovenia	12.7 (7.3)	15.9 (7.3)	552	7.6 (7.1)	10.3 (7.1)	546	0.53 (0.03)	2484
19	Latvia	6.9 (7.8)	14.9 (7.4)	525	3.0 (9.0)	9.0 (8.5)	512	0.24 (0.05)	2116
20	New Zealand	2.7 (7.6)	2.3 (7.5)	499	3.8 (8.4)	3.6 (8.4)	531	0.40 (0.04)	2459
21	Thailand	-0.4 (6.2)	14.6 (5.7)	490	-5.1 (6.0)	6.9 (5.5)	473	0.41 (0.05)	2837

Notes: 21 out of 26 participating countries have been included. See section 3.3 for the exclusion of five countries. The estimation sample consists of students born in the period 6 months before and after the cut-off date. The countries in grey are from the sample used by Bedard and Dhuey (2006). Standard errors in parentheses are adjusted for using plausible values.

and six month after the cutoff date. The standard errors of the estimates are adjusted for using plausible values.

The estimates of the lower bound effect of one year of school time in columns (1) and (4) show that there are large differences in the gains in cognitive skills between countries. The estimated effects differ between 0 and 0.4 standard deviations of the test scores. High gains in achievement for both tests are found for Norway, Singapore and Iceland. On the other hand, we also find very low gains in cognitive skills; in five countries the estimated effects do not significantly differ from zero. Hence, one year of time in school does not yield a gain in cognitive skills in these countries. The country specific estimates remain quite similar when we use the exact birth date as assignment variable and exclude children born very close to the cutoff date (table 3A.5a in the Appendix). In general, countries with high gains in achievement in math also achieve high gains in science: the correlation between the estimates in columns (1) and (4) is 0.9, which is significant at the 1%-level. For most countries the lower bound estimates are very close to the estimates that are adjusted for sampling bias (columns (2) and (5)). However, for several individual countries, in particular Iran, Austria, Latvia and Thailand, the adjusted score is substantially larger than the lower bound estimates. In these countries the proportion of students that is not on track is relatively high, as indicated by the first stage equation in column (5). The ranking of countries based on the adjusted estimates is quite similar to the ranking based on the lower bound estimates (the correlation between these estimates for both math and science is 0.94, which is significant at the 1% level).

Differences in achievements of 13-year-olds between countries

The sample of countries that can be used for estimating the effect for 13-years-olds consists of 34 countries. Table 3.III shows the estimation results. Again we observe large differences between countries. The lower bound estimates of the gains in cognitive skills (columns (1) and (4)) are substantially smaller for 13-years-olds than for 9-year-olds, which is in line with the results from the previous section. Singapore achieves the highest gains in cognitive skills in both subjects; the results for science are remarkably far ahead of all other countries. Another remarkable finding for the 13-year-olds is the large number of countries for which the lower bound estimate of the effect of one year of school time is statistically insignificant. This suggests that in these countries one additional year of school time does not matter for the performance on the TIMSS math or science tests. More countries generate a statistically significant effect in science than in math. The estimates are quite similar when we use the exact birth date as assignment variable and exclude children born very close to the cutoff date (table 3A.5b in the Appendix). Again we observe that countries with high gains in achievement in math also achieve high gains in science: the correlation between the estimates in columns (1) and (4) is 0.73, which is significant at the 1%-level. For countries with high proportions of students that are on track, indicated by high first stage estimates, we observe that the lower bound estimates are quite similar to the estimates that are adjusted for sampling bias (columns (2) and (5)). However, for countries with a relatively low first stage estimate the adjusted scores can be substantially higher than the lower bound estimates.

Table 3.III: Reduced form estimates of the gain in cognitive skills and mean upper grade by country for 13-year-olds

		Math			Science				
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ranking	**Country**	**gain**	**corr. gain**	**mean**	**gain**	**corr. gain**	**mean**	**first stage**	**N**
1	Singapore	27.8 (4.6)	27.8 (4.6)	643	55.9 (5.9)	56.6 (5.9)	607	0.96 (0.01)	3567
2	Sweden	19.1 (5.4)	22.0 (5.4)	519	25.2 (5.5)	28.3 (5.5)	535	0.91 (0.01)	3403
3	Italy	18.5 (8.3)	29.4 (8.1)	493	15.0 (8.0)	24.7 (7.8)	498	0.71 (0.03)	2186
4	Norway	18.2 (6.2)	18.5 (6.1)	503	22.1 (6.7)	22.8 (6.7)	527	0.84 (0.02)	2751
5	Iran	15.5 (7.1)	24.9 (6.5)	428	7.7 (6.9)	14.1 (6.3)	470	0.36 (0.04)	2495
6	Czech Republic	15.5 (5.6)	20.1 (5.5)	564	9.1 (5.6)	13.2 (5.5)	574	0.52 (0.02)	3248
7	Spain	13.6 (5.7)	27.0 (5.5)	487	10.9 (5.9)	23.4 (5.6)	517	0.73 (0.03)	3157
8	Iceland	13.5 (7.2)	13.9 (7.2)	487	17.1 (7.4)	17.5 (7.4)	494	0.94 (0.02)	1834
9	South-Korea	13.1 (6.5)	14.6 (6.5)	607	10.7 (6.2)	12.0 (6.1)	565	0.82 (0.02)	2912
10	Denmark	12.1 (7.3)	13.8 (7.3)	502	8.3 (8.2)	10.5 (8.2)	478	0.51 (0.04)	2096
11	Canada	9.9 (3.6)	16.0 (3.6)	527	3.2 (4.4)	9.3 (4.2)	531	0.55 (0.03)	7733
12	Latvia	9.9 (7.4)	17.7 (7.2)	493	23.2 (7.6)	33.1 (7.4)	485	0.57 (0.04)	2334
13	Scotland	9.9 (6.6)	9.6 (6.5)	498	12.7 (7.1)	13.1 (7.1)	517	0.75 (0.03)	2824
14	Thailand	9.7 (4.9)	14.9 (4.3)	522	11.1 (3.8)	18.6 (3.4)	525	0.46 (0.04)	5232
15	Cyprus	9.5 (8.1)	16.2 (8.0)	474	14.7 (7.8)	21.2 (7.8)	463	0.80 (0.02)	2837
16	Slovak Republic	8.6 (5.5)	12.5 (5.5)	547	11.9 (5.8)	15.7 (5.8)	544	0.79 (0.02)	3475
17	Belgium (French)	7.1 (6.8)	20.2 (6.5)	526	11.5 (7.7)	26.9 (7.4)	471	0.55 (0.04)	1872
18	Switzerland	6.7 (4.3)	19.0 (4.0)	545	8.7 (5.5)	21.2 (5.0)	522	0.30 (0.04)	3727
19	Japan	4.8 (4.6)	5.0 (4.6)	605	14.1 (4.5)	14.1 (4.5)	571	0.98 (0.01)	5158
20	Greece	3.0 (6.3)	11.2 (6.2)	484	5.9 (6.8)	13.0 (6.7)	497	0.75 (0.02)	3543
21	Belgium (Flemish)	2.8 (4.8)	7.1 (4.8)	565	12.2 (5.4)	17.2 (5.3)	550	0.77 (0.03)	2622
22	Lithuania	2.6 (7.3)	7.5 (7.3)	477	17.6 (7.5)	23.5 (7.5)	476	0.55 (0.04)	2590
23	Romania	2.2 (6.3)	8.8 (6.0)	482	2.0 (6.8)	7.6 (6.5)	486	0.16 (0.03)	3504
24	Russia	1.9 (5.7)	4.8 (5.6)	535	14.9 (6.6)	19.2 (6.5)	538	0.47 (0.03)	3855
25	Hungary	0.5 (5.8)	3.2 (5.9)	537	4.0 (6.0)	6.3 (6.0)	554	0.30 (0.03)	2887
26	Portugal	0.4 (5.8)	18.8 (5.6)	454	13.4 (7.5)	34.5 (7.2)	480	0.62 (0.03)	2579
27	Germany	-0.1 (6.4)	9.8 (6.2)	509	-0.4 (7.4)	11.6 (7.1)	531	0.36 (0.04)	2447
28	Slovenia	-0.8 (6.3)	4.6 (6.2)	541	-10.3 (6.5)	-6.4 (6.4)	560	0.43 (0.03)	2688
29	France	-2.6 (6.7)	21.0 (6.3)	538	2.0 (6.6)	25.3 (6.2)	498	0.43 (0.04)	2159
30	Austria	-3.6 (6.0)	4.8 (5.9)	539	1.1 (6.8)	10.9 (6.7)	558	0.53 (0.03)	2569
31	New Zealand	-4.7 (6.0)	-4.6 (6.0)	508	-3.4 (6.7)	-3.2 (6.7)	525	0.29 (0.04)	3432
32	Netherlands	-5.7 (6.5)	4.1 (6.5)	541	5.5 (7.2)	14.0 (7.1)	560	0.39 (0.04)	1770
33	Hong Kong	-7.1 (6.3)	-2.5 (6.1)	588	-2.2 (6.7)	2.7 (6.5)	522	0.59 (0.03)	2996
34	England	-9.0 (9.4)	-8.3 (9.3)	506	1.8 (9.0)	2.6 (8.9)	552	0.94 (0.02)	1834

Notes: 34 out of 41 countries have been included (see Section 3.3). Estimation sample for estimating gains in achievement scores consists of students born in the period 6 months before and after the cut-off date. The countries in grey are from the sample used by Bedard and Dhuey (2006). Standard errors in parentheses are adjusted for using plausible values.

We have also investigated whether countries that have high achievement gains for 9-year-olds also have high achievement gains for 13-year-olds, and whether countries with low gains for 9-year-olds also have low gains for 13-year-olds. We find a correlation of 0.51 for the reduced form estimates for math and a correlation of 0.34 for the reduced form estimates for science. The correlations for the math tests are statistically significant. This implies that education systems that are more effective in producing cognitive skills for 9-year-olds are also more effective in producing cognitive skills for 13-year-olds.

3.6 Do gain scores and level scores yield a consistent assessment of education systems?

This section shows the results of the third part of our empirical analysis. We compare the estimates of the gains in cognitive skills with the levels of the cognitive skills as currently used for the benchmarking of education systems. Gain scores and level scores can both be considered as measures of the performance of an education system. An interesting question is whether the rankings of the estimates of gains in cognitive skills in tables 3.II and 3.III are consistent with the ranking based on the level of the test scores. On the one hand we would expect a positive correlation between gain and level scores because the level scores are the sum of all gains in cognitive skills caused by time in and out of school. On the other hand gain scores and level scores might differ because both measures have limitations. Level scores do not isolate the contribution of time in school from the contribution of time out of school. High level scores could mask a low performing education system if the conditions outside schools are favorable for learning, which means a high contribution of time out of school to student performance. Low level scores might also be misleading about the performance of the education systems if the conditions outside school are unfavorable for learning.[35] Gain scores isolate time in school from time out of school but only measure the effect of one year in school. It follows that low gain scores could be the result of a low quality of education but also the result of the timing of the curriculum. The latter, however, seems to contrast with the way the TIMSS tests have been developed (see Section 3.3).

For investigating whether the two measures show a consistent ranking of countries we compare the reduced form estimates of the gain scores with the mean upper grade

35 A similar concern arises when the performance of schools is compared. School with low level scores might actually have high 'value added' (Figlio and Loeb, 2011).

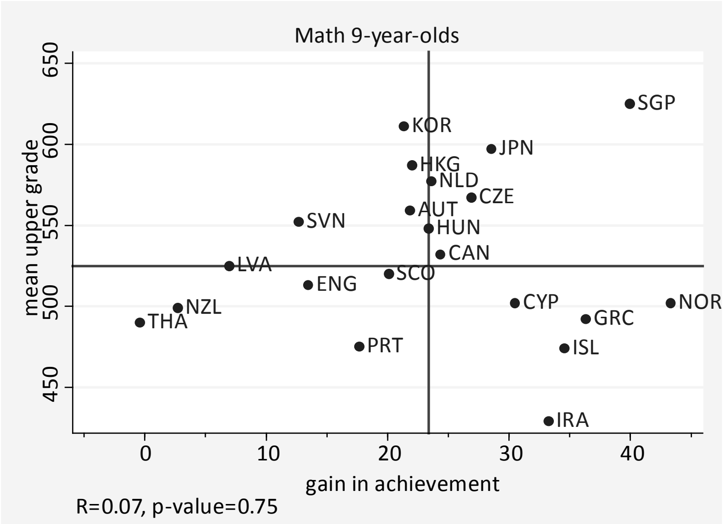
scores.[36] We use the mean of the upper grades scores, instead of the mean of the scores from both adjacent grades, because the gain scores measure the effect of time in school between the lowest and the highest grade and, therefore, are included in the mean upper grade scores. In figures 3.2, 3.3 and 3.4 we have plotted the mean upper grade scores for the different age groups and subjects on the vertical axis against the estimates of the gains in achievement on the horizontal axis. Figure 3.2 shows the results for the 9-years-olds in math and science. Figure 3.3 shows the results for the 13-years-olds. In addition, figure 3.4 plots the level score at age 13 against the gains in cognitive skills of 9-year-olds. We have included axes at the median level of gain scores and level scores in all figures which generate four quadrants of the performance of education systems: low level – low gain; low level – high gain; high level – low gain; high level – high gain.

In figure 3.2 we observe no association between the mean upper grade scores and the gains in cognitive skills for both subjects. For math we observe a large variation in gains in cognitive skills for countries below the median level scores. Hence, countries with a low level score are not only observed in the low gain quadrant but also in the high gain quadrant. For instance, Norway has the highest gain of all countries but also a level score below the median level. The gain scores for countries above the median level scores are less dispersed and more concentrated around the median gain scores. For science we observe a more even distribution of countries across the four quadrants of performance. A similar pattern is found for the 13-year-olds (figure 3.3).[37] We observe no association between mean level scores and gains in cognitive skills. Countries with high level scores are not consistently found in the top of the ranking based on the gain scores. Similarly, countries with low level test scores do not systematically have low gains in achievement. Singapore can be considered as a (positive) outlier for both subjects in figure 3.3. Figure 3.4 shows the association between the gain score at the age of 9 and the level score at the age of 13. We observe that the country rankings for the 13-year olds in TIMSS are not related with the gains in achievement for 9-year olds.

The results from figures 3.2 and 3.3 have been summarized in table 3.IV. This table shows correlations between the estimates of the gains in achievement and the country

36 See for the mean upper grade scores TIMMS 9 http://timssandpirls.bc.edu/timss1995i/TIMSSPDF/P1HiLite.pdf and for TIMMS 13 http://timssandpirls.bc.edu/timss1995i/TIMSSPDF/P2HiLite.pdf.

37 Figure 3.3 also distinguishes countries with and without an external exit exam which is relevant for the analysis in the next section.



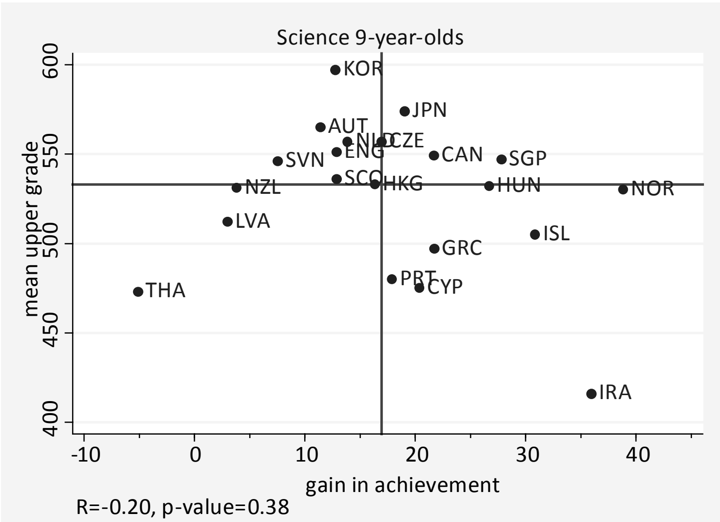
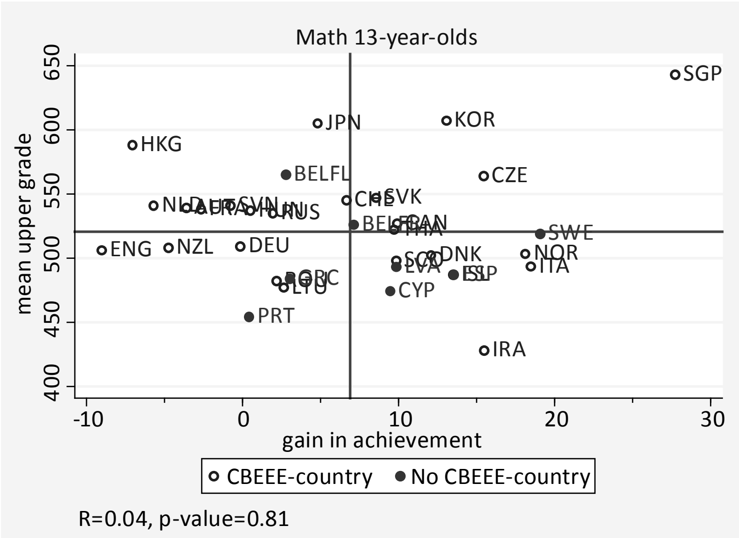


Figure 3.2: The association between country level scores and the effect of one year of school time on cognitive skills for 9-year-olds.



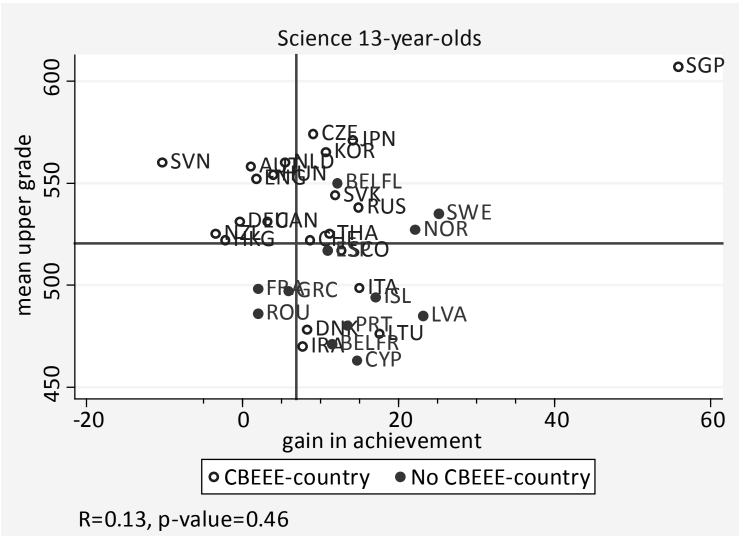
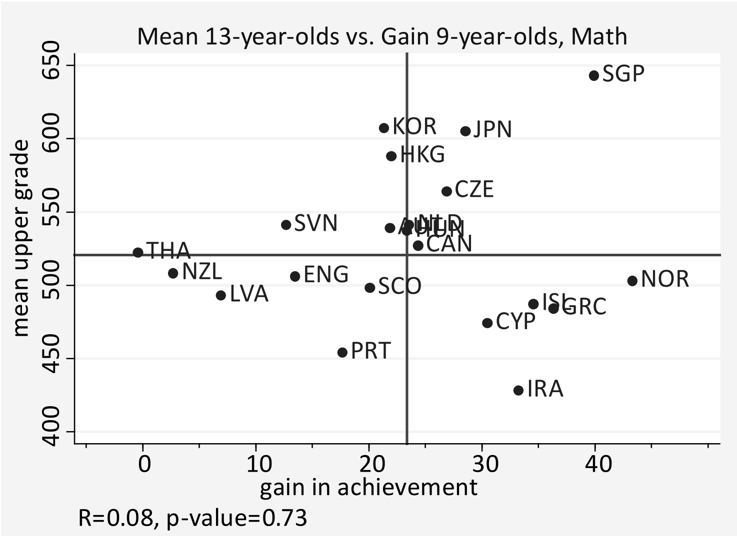


Figure 3.3: The association between country level scores and the effect of one year of school time on cognitive skills for 13-year olds.



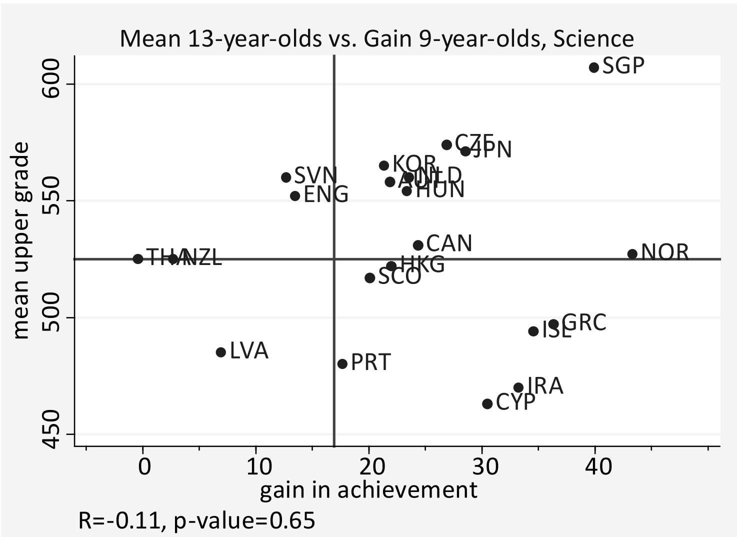


Figure 3.4: The association between country level scores of 13-year-olds and the effect of one year of school time on cognitive skills for 9-year-olds.

level scores by test and age group. The table also includes the results of sensitivity analyses with respect to the size of the discontinuity samples (panel A), the non-random timing of birth (panel B), and the inclusion of countries not used in the sample of Bedard and Dhuey (2006) and sampling bias (panel C). The main finding of table 3.IV is that the correlation between the gain scores and the country level scores is close to zero and statistically insignificant for all subjects and age groups. This result is found for the main results as shown in figure 3.2 and 3.3 (middle panel of panel A) and is robust to a series of sensitivity analyses as shown in the other panels of table 3.IV. In panel C we find higher correlations for 13-year-olds in the sample of countries with a first stage estimates above 0.75. However, this correlation is completely driven by the large gains of Singapore. The estimates that are adjusted for sampling bias suggest a negative correlation between gain scores and level scores.

The low correlations imply that country level scores and country gain scores often tell different stories about the performance of education systems. Hence, countries that are top ranked in the test are not necessarily characterized by high gains in achievement, and low ranked countries are not necessarily characterized by low gain scores. The current use of the outcomes of international cognitive tests in educational policy focuses on the ranking along the vertical axis. The figures in this section show that these rankings hide large variation in gains in cognitive skills between countries illustrated by the variation along the horizontal axis. As such, gain scores add a second dimension for assessing the performance of education systems. For educational policy it seems useful to focus not only on the ranking along the vertical axis but also take the horizontal axis into account, for instance by looking at the four quadrants of performance. For countries in the low level – low gain quadrant or in the high level – high gain quadrant the assessment of the performance seems clear. But for countries in the other two quadrants, the assessment of the performance of education system is less clear. For example, the below median level score of Norway can be interpreted as a signal of low quality education. However, the high gain scores tell a different story and suggest that other factors are likely to explain the low level scores.[38] For countries in these two quadrants a mere focus on the ranking along the vertical axis might yield misleading information for educational policy.

38 It might be speculated that the relatively late school starting age in Norway lowers the level scores.

Table 3.IV: Correlations between countries gain score and mean upper grade by age group and subject

Panel A		**Total sample using different discontinuity samples**								
		± 3 months			**± 6 months**			**± 9 months**		
test	**subject**	**correlation**	**p-value**	**N**	**correlation**	**p-value**	**N**	**correlation**	**p-value**	**N**
TIMSS 9	math	0.05	0.83	21	0.07	0.75	21	0.09	0.71	21
TIMSS 9	science	-0.23	0.32	21	-0.20	0.38	21	-0.18	0.44	21
TIMSS 13	math	0.08	0.65	34	0.04	0.81	34	0.04	0.84	34
TIMSS 13	science	0.05	0.80	34	0.13	0.46	34	0.14	0.43	34
Panel B		**Addressing birth selection around the cutoff using ± 6 months sample and different assignment variable:**								
		birth day			**birth day excluding 3 days before and after the cutoff**			**birth month**		
test	**subject**	**correlation**	**p-value**	**N**	**correlation**	**p-value**	**N**	**correlation**	**p-value**	**N**
TIMSS 9	math	0.30	0.24	17	0.28	0.28	17	0.29	0.25	17
TIMSS 9	science	-0.13	0.61	17	-0.08	0.77	17	0.06	0.82	17
TIMSS 13	math	0.14	0.46	30	0.13	0.49	30	0.15	0.44	30
TIMSS 13	science	0.14	0.45	30	0.13	0.51	30	0.14	0.47	30
Panel C		**Addressing rolling admissions & sample selection using ± 6 months sample:**								
		Bedard and Dhuey sample			**countries with first stage>0.75**			**gains corrected for sample selection**		
test	**subject**	**correlation**	**p-value**	**N**	**correlation**	**p-value**	**N**	**correlation**	**p-value**	**N**
TIMSS 9	math	0.01	0.99	10	0.22	0.56	9	-0.01	0.96	21
TIMSS 9	science	-0.28	0.44	10	-0.07	0.86	9	-0.38	0.09	21
TIMSS 13	math	-0.08	0.76	18	0.33	0.30	12	-0.14	0.43	34
TIMSS 13	science	-0.08	0.76	18	0.47	0.13	12	-0.08	0.66	34

3.7 Investigating the determinants of international differences in cognitive skills

The estimated effects of the effect of time in school are also of interest to the literature that uses international cognitive tests for investigating the determinants of international differences in educational achievement (Hanushek and Woessmann, 2011). This literature investigates whether differences in school inputs or institutions, such as school accountability and autonomy, central exams, competition between schools or tracking, can explain the large differences in achievements of students between countries. Most studies in this recent literature have focused on identifying cross-country associations.[39] However, due to the complex nature of the production of human capital it remains unclear whether these

39 Several studies have used a quasi-experimental design, see Section 3.1.

associations can be interpreted as causal effects. Therefore, it has been argued that 'further exploration of quasi-experimental settings in the international data should be high on the agenda' (Hanushek and Woessman, 2011). This chapter provides such a quasi-experimental setting and the previous section shows that assessments of the performance of education systems based on level scores might be different from assessments based on gain scores. The approach that is applied in this chapter might also yield new opportunities for investigating the determinants of international differences in education achievements between countries. Whereas the current literature tries to relate differences in student outcomes to differences in input factors or institutions, it is also possible to relate differences in gains in achievement to differences in input factor or institutions. The advantage of our approach is that it isolates the effects of time in school from the effects of time out of school.

To illustrate these opportunities, we re-examine the impact of curriculum-based external exit exam systems (CBEEE). Previous studies have investigated the effects of external exit-exams and provide a consistent picture about the beneficial effect of external exit-exams. The effects might be even larger than a whole-grade level equivalent, between 0.2 to 0.4 standard deviations of the respective tests (Hanushek and Woessmann, 2011). These results have also been found for the 1995 TIMSS math and science achievements of 13-year-olds in a study that uses country level data (Bishop, 1997) and in a study that uses micro-level data (Woessman, 2003). We conduct a similar analysis using gains in achievements. Table 3.V shows the estimations results; the left panel shows the results using country level-data, the right panel shows the results using micro-level data.

The first column of panel A of table 3.V replicates the estimates from Bishop (1997). The mean upper grade scores of 34 countries are regressed on a dummy for having a curriculum-based external exit exam. The estimates show that countries that have an external-exit exam score 29 points higher on math and 33 points higher on science tests. Bishop (1997) reports similar results (23 points for math and 34 points for science) in models with more controls. In columns (2) and (3) the estimated gains in achievement instead of the mean upper grade scores are used as dependent variable. Column (2) uses the lower bound estimate of the gains in cognitive skills; column (3) uses estimates that are adjusted for sampling bias. The estimates show that countries that have an external exit-exam do not produce higher gains in achievement than countries that do not have an external exit exam; the point estimates are negative and in column (3) we even find statistically significant negative effects. The right panel of table 3.V uses micro-level data; column (4) uses a specification as in Woessman (2003). The models in columns (5) and (6) are based on equation (3.4),

Table 3.V: The effect of curriculum-based external exams on cognitive skills of 13-year and 9-year olds

	Country level data			Micro-level data		
	(1)	(2)	(3)	(4)	(5)	(6)
	mean upper grade	gain	corr. gain	mean upper grade	gain	corr. gain
Panel A: 13-year-olds						
math	28.5*	-1.9	-6.8**	24.5***	-1.8	-3.1
	(15.6)	(3.0)	(3.0)	(0.6)	(2.5)	(2.5)
Observations	34	34	34	116235	104316	109942
science	33.0***	-2.5	-8.8**	29.5***	-0.9	-1.4
	(11.7)	(3.8)	(3.8)	(0.6)	(2.5)	(2.5)
Observations	34	34	34	116235	104316	109942
Panel B: 9-year-olds						
math	41.0**	-0.0	-0.7	51.4***	3.9	6.0
	(17.0)	(6.6)	(5.5)	(1.1)	(3.8)	(3.8)
Observations	21	21	21	71874	66803	68869
science	33.9**	-4.3	-3.9	35.9***	-2.3	0.86
	(15.2)	(6.0)	(5.3)	(0.8)	(3.6)	(3.6)
Observations	21	21	21	71874	66803	68869

Notes: In columns (1)-(3) the country's mean/gain has been regressed on a dummy for CBEEE-country. In column (4) individual test scores have been regressed on the CBEEE-dummy. The model in columns (5) and (6) also includes a dummy for CBEEE-countries and the interaction of CBEEE with the dummy for being born before the cutoff date (see equation (3.5)). For the country level data robust standard errors are used. For the micro-level data standard errors are adjusted for using plausible values.

and include dummies for having an external exit exam and an interaction of time in school (grade) with the external exit exam (CBEEE) like in equation (3.5):

$$Y_i = \lambda_0 + \lambda_1 S_i + \lambda_2 CBEEE + \lambda_3 S_i * CBEEE + f(birthmonth_i - C) + \lambda_4 X_i + \varepsilon_i \tag{3.5}$$

The estimates show that students in countries with external exit exams score 25 (30) points higher on math (science) and the estimated effects are statistically significant. Woessman (2003) also finds statistically significant positive effects, but these effects are smaller after including an extensive set of family background and school-input controls (11 (16) points for math (science)). However, we do not find a positive effect of external exit-exams on the gains in achievement in the last year before the test. This result can also be observed in figure 3.3 which distinguishes countries with and without an external exit exam. We do not observe that countries with external exams have higher gains in achievement than countries without an external exit exam.

A limitation of the gain score approach is that it only refers to the achievements in the last year before the test. Hence, it is possible that we fail to find an effect of external exit-exams because they only affect the results in earlier years in school. The TIMSS test scores of 9-year-olds provide an opportunity to observe what happened in one of the earlier years. The estimation results, based on the same models, are shown in panel B of table 3.V. Again we observe substantial positive effects of external exit exams in columns (1) and (4). The estimates indicate that central exit-exams increase test scores in math by 51 points and in science by 36 points. However, the estimated effects become statistically insignificant when we focus on gains in achievement. This means that we do not find higher gains in achievement during two school years for students in countries that have an external exit exam compared to students in countries that do not have external exit exams. Although this finding relates to two school years, which is one quarter of the total amount of time in school, it raises concerns about unobserved differences between countries that have external exit exams and countries that do not have external exit exams in the studies that previously used the 1995 TIMSS data for estimating the effect of external exit-exams.

In sum, our re-examination of previous results on the effect of external exit exams using the 1995 TIMSS data shows that results based on level scores might differ from the results based on gain scores. This illustrates that an approach that focuses on gains in achievement might offer new opportunities and insights for investigating the determinants of international differences in educational achievement.

3.8 Conclusions

This study applies a quasi-experimental approach for estimating the effect of one year of school time on the performance in international cognitive tests by exploiting the assignment of students to different grades based on school entry rules. This method produces estimates of gains in cognitive skills for students in different age groups in the year before the test for worldwide samples of countries and for individual countries. This method also enables a comparison of the performance of education systems based on gain scores instead of level scores.

We find that time in school on average matters for student performance in international cognitive tests. For the pooled sample of countries we find that a year of school time increases performance in cognitive tests with 0.2 to 0.3 standard deviations for 9-year olds and with 0.1 to 0.2 standard deviations for 13-year-olds. These effects are consistent with the results of previous studies based on credible research designs (Gormley and Gayer, 2005; Gormley et al., 2005; Hansen et al., 2004; Cascio and Lewis, 2006; and Berlinski et al., 2009). We also find large differences in the estimated gains in achievement between countries for both subjects and age groups. Countries that achieve higher gains in cognitive skills for math also achieve higher gains in science. Moreover, countries with higher gains for 9-year-olds also have higher gains for 13-years-olds.

The sampling strategy of the TIMSS-project, which focused on two adjacent grades, might induce a downward bias for our estimates. Therefore, the main estimates should be interpreted as lower bound estimates. For 9-year-olds the lower bound estimates are probably quite close to the unbiased effect. However, for 13-year-olds the lower bound estimates will probably underestimate the gains in cognitive tests for countries in which a large proportion of students is not on track.

Remarkably, we find no association between the estimated gains in achievement and the level scores of countries. At all levels of test scores we observe countries with high achievement gains and countries with low gains in achievement. The lack of association has been found for both tests (math, science) and for both age groups. Hence, assessments of the performance of education systems based on the estimated gains in achievement often are inconsistent with performance assessments based on level scores. This inconsistency might be explained by limitations of both measures. Level scores do not distinguish between the contribution of time in school and the contribution of time out of school. The gain scores only refer to the gain in achievement in the year before the test. The inconsistency of the two measures implies that the benchmarking of education systems based on level scores might yield misleading information about the performance of education systems. Low levels of test scores, or declining trends in test scores, might not be the result of low performing education systems. High levels of test scores could mask low performing education systems. Using gain scores as an additional instrument for the assessment of the performance of education systems is likely to reduce the risk of providing misleading policy information.

The quasi-experimental approach for estimating gains scores used in this chapter can also be important for investigating international differences in educational achievements. For instance, studies that use control strategies have consistently found that students perform better in countries with external exit exams. However, we do not find higher gains in achievements in countries with a central exit exam for 9-year-olds and 13-year-olds compared to countries that do not have central exit exams.

This study shows that time in school is important for acquiring cognitive skills and that there are large differences in the effects between countries. Estimates of the gains in achievement for separate countries provide a different assessment of the performance of education systems, and of the effect of specific elements of education systems, than level scores. The estimation of gain scores, which becomes possible when the current collection of international data is extended towards samples of students in adjacent grades, is likely to improve decisions on educational policies and could offer new opportunities for investigating the determinants of international differences in student achievement.

3.9 Appendix

Table 3A.1: Cutoff dates per country & source and data availability

Country	Cutoff date	Source	TIMSS 9	TIMSS 13
Austria[1]	September 1	Bedard and Dhuey	yes	yes
Belgium-Flemish	January 1	Bedard and Dhuey	no	yes
Belgium-French	January 1	Bedard and Dhuey	no	yes
Canada	January 1	Bedard and Dhuey	yes	yes
Czech Republic	September 1	Bedard and Dhuey	yes	yes
Denmark	January 1	Bedard and Dhuey	no	yes
England	September 1	Bedard and Dhuey	yes	yes
France	January 1	Bedard and Dhuey	no	yes
Greece	April 1	Bedard and Dhuey	yes	yes
Iceland	January 1	Bedard and Dhuey	yes	yes
Italy	January 1	Bedard and Dhuey	no	yes
Japan	April 1	Bedard and Dhuey	yes	yes
New Zealand	May 1	Bedard and Dhuey	yes	yes
Norway	January 1	Bedard and Dhuey	yes	yes
Portugal	January 1	Bedard and Dhuey	yes	yes
Slovak Republic	September 1	Bedard and Dhuey	no	yes
Spain	January 1	Bedard and Dhuey	no	yes
Sweden	January 1	Bedard and Dhuey	no	yes
Germany[2]	July 1	Internet/TIMMS Data	no	yes
Singapore[3]	January 1	Internet/TIMMS Data	yes	yes
South-Korea[4]	March 1	Internet/TIMMS Data	yes	yes
Latvia[5]	September 1	Internet/TIMMS Data	yes	yes
Scotland[6]	March 1	Internet/TIMMS Data	yes	yes
Lithuania[7]	September 1	Internet/TIMMS Data	no	yes
Romania[8]	September 1	Internet/TIMMS Data	no	yes
Hungary[9]	June 1	Internet/TIMMS Data	yes	yes
Slovenia[10]	January 1	Internet/TIMMS Data	yes	yes
Netherlands[11]	October 1	Internet/TIMMS Data	yes	yes
Iran	October 1	TIMSS Data	yes	yes

Country	Cutoff date	Source	TIMSS 9	TIMSS 13
Thailand	January 1	TIMSS Data	yes	yes
Cyprus	March 1	TIMSS Data	yes	yes
Switzerland	January 1	TIMSS Data	no	yes
Russia	October 1	TIMSS Data	no	yes
Hong Kong	January 1	TIMSS Data	yes	yes

Notes: All cutoff dates have been checked in our data and show a (sharp) discontinuity in average grade around the given cutoff. The column 'Source' shows whether the cutoff was also shown in other sources. Cutoff dates refer to the situation in 1995. The columns 'TIMSS 9' and 'TIMSS 13' show whether the country was included for these tests.

1 Bedard and Dhuey use January 1 as the cutoff date, we deviate based on the data and: http://virtuelleschule.bmukk.gv.at/fileadmin/folder/Folder_Basisinformationen/school_system_Austria_EN.pdf
2 http://de.wikipedia.org/wiki/Schulpflicht_(Deutschland)
3 http://www.moe.gov.sg/education/
4 http://en.wikipedia.org/wiki/Education_in_South_Korea#Elementary_school
5 http://www.viaa.gov.lv/files/news/1808/educ_in_latvia.pdf
6 http://en.wikipedia.org/wiki/Education_in_Scotland
7 http://en.wikipedia.org/wiki/Education_in_Lithuania
8 http://en.wikipedia.org/wiki/Romanian_educational_system
9 http://en.wikipedia.org/wiki/Education_in_Hungary
10 http://eacea.ec.europa.eu/education/eurydice/documents/eurybase/national_summary_sheets/047_SI_EN.pdf
11 http://www.onderwijsinspectie.nl/actueel/vraagantwoord#Wie_bepaalt_of_een_kind_overgaat_naar_groep_3_

Table 3A.2a: Means of test scores and covariates by age relative to the cutoff for the pooled sample of 9-year-olds

Relative age	Math	Science	Female	Born in country	Speaks language of test at home	Living with mother	Living with father	Number of books at home	N
-12	549.21	538.68	0.51	0.93	0.75	0.96	0.85	95.94	4999
-11	547.29	536.93	0.51	0.93	0.76	0.95	0.84	96.77	5137
-10	545.45	535.48	0.52	0.93	0.74	0.96	0.84	97.78	5613
-9	543.34	532.30	0.50	0.92	0.76	0.95	0.84	97.69	5664
-8	542.74	532.14	0.51	0.93	0.76	0.95	0.84	97.41	5763
-7	538.04	528.83	0.50	0.93	0.75	0.96	0.84	95.76	5676
-6	538.46	528.05	0.50	0.93	0.76	0.96	0.84	97.62	5707
-5	533.71	522.10	0.49	0.92	0.74	0.96	0.84	95.01	5812
-4	533.02	521.58	0.51	0.93	0.74	0.96	0.84	97.58	5535
-3	529.74	516.15	0.50	0.93	0.72	0.96	0.85	92.75	6099
-2	526.90	515.43	0.50	0.92	0.73	0.95	0.84	95.14	5678
-1	517.61	506.14	0.50	0.93	0.73	0.96	0.85	93.07	5911
0	498.54	494.00	0.51	0.92	0.75	0.96	0.85	97.37	5400
1	489.24	487.23	0.49	0.92	0.74	0.96	0.84	97.33	5266
2	488.91	484.65	0.51	0.92	0.75	0.96	0.84	96.59	5261
3	485.70	483.74	0.49	0.93	0.75	0.96	0.85	98.58	5348
4	481.92	479.15	0.52	0.93	0.74	0.95	0.85	93.91	5431
5	479.01	476.73	0.51	0.92	0.75	0.95	0.84	96.18	5355
6	479.60	475.20	0.49	0.92	0.74	0.95	0.84	96.72	5372
7	474.85	471.19	0.50	0.93	0.72	0.94	0.84	92.73	5200
8	474.75	470.37	0.48	0.92	0.72	0.95	0.84	92.43	5099
9	475.07	466.54	0.49	0.92	0.71	0.95	0.85	93.02	4957
10	473.42	466.33	0.49	0.92	0.70	0.95	0.85	91.21	4436
11	468.58	460.85	0.49	0.93	0.70	0.95	0.84	89.33	4623

Notes: The relative age of the oldest students is -12; relative age 0 means born in the first month at the right side of the cut-off data.

Table 3A.2b: Means of test scores and covariates by age relative to the cutoff for the pooled sample of 13-year-olds

Relative age	Math	Science	Female	Born in country	Speaks language of test at home	Living with mother	Living with father	Number of books at home	High educated mother	High educated father	N
-12	518.765	513.228	0.502	0.940	0.827	0.952	0.826	102.198	0.261	0.315	7752
-11	517.974	514.059	0.498	0.939	0.838	0.952	0.832	105.008	0.286	0.342	7701
-10	517.791	513.232	0.501	0.944	0.842	0.955	0.829	106.183	0.282	0.337	8666
-9	518.192	513.094	0.501	0.945	0.849	0.954	0.838	106.671	0.297	0.350	9126
-8	518.042	512.474	0.502	0.947	0.841	0.958	0.827	106.556	0.284	0.334	9329
-7	513.727	508.215	0.505	0.946	0.834	0.959	0.833	104.163	0.285	0.341	9222
-6	515.361	509.274	0.499	0.943	0.837	0.955	0.837	104.889	0.294	0.340	9168
-5	514.537	508.856	0.494	0.946	0.837	0.959	0.841	106.206	0.300	0.354	9256
-4	512.667	505.842	0.497	0.944	0.832	0.954	0.840	106.218	0.293	0.340	9233
-3	511.510	504.443	0.492	0.949	0.831	0.959	0.838	105.772	0.295	0.345	9160
-2	509.452	503.850	0.491	0.948	0.835	0.954	0.838	104.219	0.285	0.333	8747
-1	505.535	500.905	0.498	0.950	0.822	0.957	0.843	102.909	0.289	0.354	9191
0	500.315	489.859	0.490	0.949	0.839	0.954	0.837	104.547	0.307	0.357	8563
1	499.595	488.652	0.499	0.952	0.838	0.957	0.844	106.647	0.307	0.351	7883
2	496.908	487.421	0.486	0.954	0.845	0.958	0.843	107.520	0.316	0.368	8364
3	495.877	484.989	0.490	0.955	0.841	0.958	0.843	107.507	0.313	0.362	8422
4	493.663	482.046	0.490	0.949	0.836	0.962	0.844	106.720	0.297	0.358	8403
5	493.269	482.782	0.494	0.959	0.842	0.962	0.846	108.773	0.316	0.364	7926
6	493.782	480.746	0.479	0.956	0.838	0.961	0.846	107.214	0.313	0.362	8156
7	492.918	481.231	0.483	0.954	0.836	0.964	0.843	106.992	0.325	0.383	7962
8	492.970	480.117	0.488	0.954	0.824	0.963	0.847	108.596	0.324	0.377	7906
9	493.018	478.652	0.481	0.954	0.823	0.967	0.851	107.079	0.312	0.361	7316
10	491.793	479.242	0.477	0.955	0.831	0.959	0.852	106.122	0.309	0.361	7052
11	488.953	478.188	0.467	0.959	0.815	0.967	0.858	105.302	0.308	0.368	6634

Notes: The relative age of the oldest students is -12; relative age 0 means born in the first month at the right side of the cut-off data. High educational level is defined as having some vocational education or more.

Table 3A.3a: Balancing tests for 9-year-olds

	Effects on variable		Effects on dummy=1 if variable is missing	
	± 3 months	± 6 months	± 3 months	± 6 months
Effect of being born left of the cutoff date on:	(1)	(2)	(3)	(4)
Female	-0.00644	0.00189	0.00106	0.000600
	(0.0117)	(0.00757)	(0.00123)	(0.000932)
N	33502	66560	33615	66803
Born in country of test	0.00681	0.00311	-0.00254	-0.000786
	(0.00603)	(0.00432)	(0.00289)	(0.00221)
N	30798	61120	33615	66803
Language at home is language of test	0.0121	0.00229	0.00277	0.00351
	(0.00994)	(0.00687)	(0.00664)	(0.00513)
N	24666	49050	33615	66803
Living with father	-0.00613	0.000481	0.00213	-0.000342
	(0.00855)	(0.00576)	(0.00326)	(0.00240)
N	30575	60559	33615	66803
Living with mother	0.000436	-0.00503	0.000197	-0.000594
	(0.00464)	(0.00309)	(0.00277)	(0.00219)
N	30702	60793	33615	66803
Number of books at home	-0.525	-0.914	-0.00340	-0.00851**
	(1.733)	(1.254)	(0.00492)	(0.00340)
N	29706	58884	33615	66803

Notes: Each cell shows the estimation results of a separate regression of the covariate on a dummy for being born at the left side of the cut-off and a linear function of age. All models include country dummies and interactions of these dummies with the age function which differs at either side of the cutoff. The dependent variable in columns (3) and (4) is a dummy for having a missing value on the relevant covariate.

Table 3A.3b: Balancing tests for 13-year-olds

	Effects on variable		Effects on dummy=1 if variable is missing	
	± 3 months	± 6 months	± 3 months	± 6 months
Effect of being born left of the cutoff date on:	(1)	(2)	(3)	(4)
Female	0.00222	0.00194	-0.00193***	-0.00125***
	(0.00996)	(0.00636)	(0.000690)	(0.000452)
N	51839	104162	51908	104316
Born in country of test	-0.00279	-0.00257	0.00118	0.000752
	(0.00414)	(0.00273)	(0.00185)	(0.00127)
N	47727	95783	51908	104316
Language at home is language of test	-0.000758	-0.00226	-0.00927*	-0.00736**
	(0.00613)	(0.00424)	(0.00485)	(0.00370)
N	45561	91554	51908	104316
Living with father	-0.00368	-0.00240	0.00180	-0.00121
	(0.00693)	(0.00467)	(0.00231)	(0.00157)
N	48507	97408	51908	104316
Living with mother	-0.000945	0.000630	-0.000108	-0.00186
	(0.00365)	(0.00255)	(0.00187)	(0.00128)
N	48725	97833	51908	104316
Number of books at home	-1.159	-0.871	-0.00105	-0.00117
	(1.362)	(0.913)	(0.00219)	(0.00148)
N	48536	97463	51908	104316
High educated mother	-0.0102	-0.0150***	-0.00740	-0.0173***
	(0.00872)	(0.00576)	(0.00786)	(0.00498)
N	35683	71869	51908	104316
High educated father	0.00549	-0.00569	-0.00229	-0.0102**
	(0.0103)	(0.00694)	(0.00745)	(0.00505)
N	34732	69875	51908	104316

Notes: Each cell shows the estimation results of a separate regression of the covariate on a dummy for being born at the left side of the cut-off and a linear function of age. All models include country dummies and interactions of these dummies with the age function which differs at either side of the cutoff. The dependent variable in columns (3) and (4) is a dummy for having a missing value on the relevant covariate.

Appendix: sampling bias adjustment

Table 3A.4 illustrates the sampling bias adjustment. For instance, we do not observe retained students born in the first month at the right side of the cut-off date (relative age=0). However, we do observe retained students born one year earlier (relative age = -12). We adjust the scores of these students and include them in the main estimation sample. The adjusted score is obtained by:

$$Y^{i}_{delayed(missing)} = Y^{i}_{ontrack} * \frac{Y^{i-12}_{delayed}}{Y^{i-12}_{ontrack}}$$

Hence, for the missing students with relative age=0 we get: 492.40*439.79/553.60=492.4 0*0.79=391.17.

Adjusted Scores for missing accelerated students are similarly obtained:

$$Y^{i}_{accelerated(missing)} = Y^{i}_{ontrack} * \frac{Y^{i+12}_{accelerated}}{Y^{i+12}_{ontrack}}$$

To obtain estimates that are adjusted for sampling bias we perform this adjustment for each separate month $i \in [-6,-1]$ for missing delayed students, and $i \in [-6,-1]$ for missing accelerated students.

Table 3A.4: Fraction on track and average test scores of those on track and not on track (delayed or accelerated) for TIMMS 9. l=left of the cutoff, r=right of the cutoff

rel. age (l)	on track (l)	math on track (l)	math delayed (l)	science on track (l)	science delayed (l)	N
-12	0.96	553.60	439.79	543.00	431.11	4999
-11	0.95	552.02	447.80	541.93	431.64	5137
-10	0.96	550.33	435.40	540.32	426.11	5613
-9	0.95	548.72	450.66	537.28	446.69	5664
-8	0.94	548.59	449.64	537.74	443.10	5763
-7	0.94	543.89	449.54	534.36	445.10	5676
-6	0.93	546.16	442.95	535.30	438.11	5707
-5	0.91	540.77	461.40	528.43	457.27	5812
-4	0.90	541.75	453.36	529.38	450.50	5535
-3	0.86	542.77	450.16	527.61	446.10	6099
-2	0.85	538.12	464.04	524.60	464.07	5678
-1	0.79	532.37	462.32	518.05	461.56	5911
rel. age (r)	**on track (r)**	**math on track (r)**	**math accelerated (r)**	**science on track (r)**	**science accelerated (r)**	**N**
0	0.88	492.40	544.92	488.86	532.84	5400
1	0.94	485.58	541.89	484.26	530.05	5266
2	0.96	487.53	525.59	483.54	514.22	5261
3	0.98	484.87	518.83	483.14	507.47	5348
4	0.98	481.18	520.73	478.69	503.52	5431
5	0.98	478.50	507.47	476.18	507.22	5355
6	0.99	479.26	507.59	474.97	494.04	5372
7	0.99	474.46	502.53	470.97	486.76	5200
8	0.99	474.63	483.30	470.19	483.32	5099
9	0.98	474.50	507.13	465.99	497.18	4957
10	0.99	473.45	471.19	466.38	461.11	4436
11	0.99	468.39	483.51	460.52	487.32	4623

Table 3A.5a: Estimated gains in achievement by assignment variable (birth month/birth day) for 9-year olds. 3 days around the cutoff excluded when using birth day as assignment variable

		Math				Science			
		birth month		birth day		birth month		birth day	
Ranking	**Country**	**gain**	**N**	**gain**	**N**	**gain**	**N**	**gain**	**N**
1	Norway	43.3 (7.4)	2133	-	-	38.8 (8.5)	2133	-	-
2	Singapore	40.0 (5.2)	6986	38.3 (5.2)	6855	27.8 (4.9)	6986	26.3 (4.9)	6855
3	Greece	36.3 (6.7)	2981	33.4 (11.2)	1055	21.7 (6.5)	2981	23.0 (11.0)	1055
4	Iceland	34.6 (9.5)	1702	34.2 (9.2)	1474	30.9 (8.4)	1702	26.6 (9.0)	1474
5	Iran	33.3 (7.1)	2716	-	-	35.9 (6.8)	2716	-	-
6	Cyprus	30.5 (6.4)	3230	34.3 (6.5)	3167	20.4 (5.8)	3230	23.7 (5.6)	3167
7	Japan	28.6 (5.4)	4343	31.6 (5.3)	4268	19.1 (5.3)	4343	21.3 (5.4)	4268
8	Czech Republic	26.9 (6.9)	3108	28.2 (7.4)	2777	16.9 (6.5)	3108	14.3 (7.4)	2777
9	Canada	24.4 (5.1)	7436	-	-	21.7 (4.8)	7436	-	-
10	Netherlands	23.6 (6.8)	2241	21.5 (6.6)	2195	13.9 (6.7)	2241	11.0 (6.9)	2195
11	Hungary	23.4 (7.8)	2743	28.8 (8.0)	2590	26.7 (7.0)	2743	30.1 (7.6)	2590
12	Hong Kong	22.0 (5.6)	3851	22.0 (5.5)	3724	16.3 (5.0)	3851	17.0 (5.1)	3724
13	Austria	21.9 (7.4)	2315	22.0 (7.8)	2263	11.4 (8.1)	2315	11.2 (8.1)	2263
14	South-Korea	21.3 (6.5)	2636	24.2 (6.5)	2570	12.8 (6.4)	2636	16.0 (6.8)	2570
15	Scotland	20.1 (6.6)	3089	18.5 (7.1)	2593	12.9 (7.2)	3089	7.3 (7.8)	2593
16	Portugal	17.7 (7.3)	2310	16.6 (7.5)	2185	17.9 (8.9)	2310	17.9 (8.8)	2185
17	England	13.5 (7.1)	3087	-	-	12.9 (9.2)	3087	-	-
18	Slovenia	12.7 (7.3)	2484	14.5 (7.3)	2416	7.6 (7.1)	2484	9.4 (7.1)	2416
19	Latvia	6.9 (7.8)	2116	5.4 (8.5)	1906	3.0 (9.0)	2116	3.0 (9.1)	1906
20	New Zealand	2.7 (7.6)	2459	4.4 (7.6)	2405	3.8 (8.4)	2459	5.9 (8.2)	2405
21	Thailand	-0.4 (6.2)	2837	13.5 (6.3)	2558	-5.1 (6.0)	2837	7.1 (6.1)	2558

Table 3A.5b: Estimated gains in achievement by assignment variable (birth month/birth day) for 13-year-olds. 3 days around the cutoff excluded when using birth day as assignment variable

		Math				Science			
		birth month		birth day		birth month		birth day	
Ranking	Country	gain	N	gain	N	gain	N	gain	N
1	Singapore	27.8 (4.6)	3567	26.8 (4.9)	3502	55.9 (5.9)	3567	55.0 (6.2)	3502
2	Sweden	19.1 (5.4)	3403	16.6 (5.5)	3235	25.2 (5.5)	3403	23.3 (5.7)	3235
3	Italy	18.5 (8.3)	2186	16.1 (8.4)	2142	15.0 (8.0)	2186	12.1 (8.0)	2142
4	Norway	18.2 (6.2)	2751	-	-	22.1 (6.7)	2751	-	-
5	Iran	15.5 (7.1)	2495	-	-	7.7 (6.9)	2495	-	-
6	Czech Republic	15.5 (5.6)	3248	14.4 (5.8)	3162	9.1 (5.6)	3248	8.0 (5.8)	3162
7	Spain	13.6 (5.7)	3157	14.6 (5.9)	3084	10.9 (5.9)	3157	11.6 (6.0)	3084
8	Iceland	13.5 (7.2)	1834	11.8 (7.6)	1641	17.1 (7.4)	1834	14.0 (7.9)	1641
9	South-Korea	13.1 (6.5)	2912	13.8 (6.8)	2857	10.7 (6.2)	2912	11.8 (6.3)	2857
10	Denmark	12.1 (7.3)	2096	15.2 (7.7)	2000	8.3 (8.2)	2096	10.8 (8.5)	2000
11	Canada	9.9 (3.6)	7733	-	-	3.2 (4.4)	7733	-	-
12	Latvia	9.9 (7.4)	2334	9.1 (7.6)	2263	23.2 (7.6)	2334	24.2 (8.0)	2263
13	Scotland	9.9 (6.6)	2824	8.5 (7.1)	2457	12.7 (7.1)	2824	9.4 (7.7)	2457
14	Thailand	9.7 (4.9)	5232	8.3 (4.8)	5120	11.1 (3.8)	5232	10.1 (4.2)	5120
15	Cyprus	9.5 (8.1)	2837	6.5 (8.4)	2772	14.7 (7.8)	2837	12.5 (8.1)	2772
16	Slovak Republic	8.6 (5.5)	3475	9.3 (5.5)	3385	11.9 (5.8)	3475	10.5 (5.9)	3385
17	Belgium (French)	7.1 (6.8)	1872	6.8 (7.2)	1778	11.5 (7.7)	1872	11.9 (7.9)	1778
18	Switzerland	6.7 (4.3)	3727	5.4 (4.4)	3595	8.7 (5.5)	3727	8.9 (6.0)	3595
19	Japan	4.8 (4.6)	5158	3.7 (4.7)	5075	14.1 (4.5)	5158	13.7 (4.8)	5075
20	Greece	3.0 (6.3)	3543	14.4 (10.2)	1484	5.9 (6.8)	3543	12.6 (10.4)	1484
21	Belgium (Flemish)	2.8 (4.8)	2622	2.8 (5.0)	2479	12.2 (5.4)	2622	13.0 (5.5)	2479
22	Lithuania	2.6 (7.3)	2590	2.2 (7.7)	2531	17.6 (7.5)	2590	19.2 (7.8)	2531
23	Romania	2.2 (6.3)	3504	-1.8 (6.4)	3305	2.0 (6.8)	3504	-3.8 (7.3)	3305
24	Russia	1.9 (5.7)	3855	1.1 (5.9)	3780	14.9 (6.6)	3855	14.5 (6.7)	3780
25	Hungary	0.5 (5.8)	2887	2.4 (6.0)	2827	4.0 (6.0)	2887	6.0 (6.1)	2827
26	Portugal	0.4 (5.8)	2579	1.5 (5.9)	2532	13.4 (7.5)	2579	15.2 (7.6)	2532
27	Germany	-0.1 (6.4)	2447	-	-	-0.4 (7.4)	2447	-	-
28	Slovenia	-0.8 (6.3)	2688	1.0 (6.4)	2646	-10.3 (6.5)	2688	-9.3 (6.6)	2646
29	France	-2.6 (6.7)	2159	-1.0 (7.0)	2121	2.0 (6.6)	2159	3.7 (6.7)	2121
30	Austria	-3.6 (6.0)	2569	-6.1 (6.2)	2498	1.1 (6.8)	2569	-0.4 (7.0)	2498
31	New Zealand	-4.7 (6.0)	3432	-3.4 (6.3)	3333	-3.4 (6.7)	3432	0.2 (7.1)	3333
32	Netherlands	-5.7 (6.5)	1770	-4.1 (6.8)	1677	5.5 (7.2)	1770	6.6 (7.6)	1677
33	Hong Kong	-7.1 (6.3)	2996	-6.7 (6.4)	2895	-2.2 (6.7)	2996	-0.6 (7.0)	2895
34	England	-9.0 (9.4)	1834	-11.3 (9.9)	1635	1.8 (9.0)	1834	-1.7 (9.7)	1635

4

Zero returns to compulsory schooling: is it certification or skills that matters?

Abstract

This chapter evaluates the effects of the raising of the minimum school leaving age (ROSLA) from 14 to 15 in the Netherlands in 1971. The policy goal was to increase the number of high school graduates. The analysis shows that the change led to a decrease in the high school dropout rate of approximately 20%. However, there were no benefits in terms of employment or higher wages. I investigate several explanations for this finding and present suggestive evidence in support of a skill-based explanation that no more labor-market relevant skills were learned during this extra year of school compared to those skills previously learned out of school.

4.1 Introduction

Changes in compulsory schooling laws have often been used by policymakers to increase the educational level of the population, in particular for those in the left tail of the ability distribution. Typically, these changes come in the form of rises in the minimum school leaving age, such as the rise from 14 to 15 in the UK in 1947 or from 14 to 16 in France in 1967. The most prominent argument for implementing these changes is that this would improve the labor market prospects for those considered to underinvest in schooling. However, the literature that evaluates the effects of changes in compulsory schooling laws shows that these changes do not always lead to better labor-market outcomes. Whereas studies that use Anglo-Saxon data find positive returns to increases in compulsory schooling, studies that use data from continental European countries find no returns.[40] This puzzling result has led researchers to speculate about what explains the differences between these countries. One explanation is that it matters whether labor-market relevant skills are learned in the additional year, as these skills are important for later labour market outcomes (Pischke and von Wachter, 2008; Oosterbeek and Webbink, 2007). Another explanation is that it matters whether the additional year is accompanied by certification (Grenet, 2013).

The aim of this chapter is to explore this puzzling result further by evaluating a change in a compulsory schooling law that was specifically targeted at increasing the number of high school graduates. If certification is the leading explanation, an increase in the number of high school graduates should result in higher earnings and a better labor market position. If it matters what type of skills are learned in school, this has not necessarily to occur: when no relevant labor-market skills are learned but only a certificate is obtained (for example for the acquisition of other type of skills), an additional year in school may not result in higher earnings or a better labor market position.

To this end, I exploit the raising of minimum school leaving age from 14 to 15 in the Netherlands. The policy goal of this reform was to enable low ability students to obtain a high school diploma. The change in the compulsory schooling law was implemented in 1971 to make the duration of compulsory education equivalent to the duration necessary to complete the lowest track of secondary school. It extended the duration of compulsory schooling from 8 to 9 years without changing the curriculum. Because at that time the Netherlands applied a school entry rule with a cutoff date set at the first of October, I am able to estimate the impact of this change in a regression discontinuity design. Individuals

40 See Section 4.2 for a short review of this literature.

born on or after 1 October 1956 were the first to be affected by the change in the law. By using data on the educational level, and labor market position I can compare the outcomes of those born around this date. I exploit a unique dataset that includes information on earnings and employment over the period 2000-2009 and on social welfare benefits since the fifteenth birth day of the respondents. This enables me to estimate effects on these outcomes over a period that spans the respondents working life.

I find that the change led to a significant decrease in the high school dropout rate of approximately 20%. However, I do not find benefits in terms of employment or higher wages when the population is aged 43-53. I explore several explanations for this finding, and give some suggestive evidence in support of the skill-based explanation similar to that of Pischke and von Wachter (2008) that no more labor-market relevant skills were learned during this extra year of school compared to those skills previously learned out of school. I show that the change also did not lead to gains in literacy skills or labor market outcomes other than earnings.

I contribute to the current literature in two ways. First, I give a new piece of evidence that increases in compulsory education do not lead to better labor market outcomes. This is important as they are often considered as an effective policy to improve the labor market position of students of low ability. Second, I show that the differences in returns between countries are not necessarily explained by certification. In both the UK and the Netherlands the raising of the minimum school leaving led to more certificates, but only for the UK the change also led to higher earnings. Moreover, the UK- reform did also lead to better (literacy) skills whereas the Dutch reform did not. This suggests that differences between countries are more likely to be explained by differences in the type of skills learned in an additional year in school than by differences in certification.

The remainder of this chapter proceeds as follows. In the next section I discuss the previous studies. Section 4.3 provides additional information on the reform. In Section 4.4 I give a description of the data. Section 4.5 illustrates the empirical strategy and Section 4.6 shows results. In Section 4.7 I explore several explanations for the main findings, and Section 4.8 concludes.

4.2 Related literature

This section discusses the literature that estimates the (financial) returns to changes in compulsory schooling laws.[41] In particular I focus on the returns to educations literature that uses increases in the minimum school leaving age as instrument for schooling in a Mincerian wage equation. However, I will not focus on the returns to schooling, but rather on the direct impact of the change of the law on earnings. This impact is implicitly estimated in reduced form regressions where an outcome for earnings is regressed on the instrument (for example a dummy for a birth cohort that has been first affected by the change in the law). I do so because the aim of this chapter is to investigate what the effects are of changes in compulsory schooling laws, not what the effects are of extra schooling. Moreover, I focus on studies that use credible exogenous variation in the timing of implementation of these laws to estimate this impact. As pointed out by Grenet (2013) these studies use 'this type of variation across states or regions within a difference-in-difference framework or studies in which smoothly is controlled for the evolution of education and earnings across cohorts in a regression-discontinuity design'. This means, for example, that they adequately control for a smooth function in birth cohort when using the change of the compulsory schooling law as instrument for schooling. Hence, the reported effects below are the reduced form estimates from the returns to education literature with credible research designs. I will distinguish between Anglo-Saxon countries and continental European countries because the findings differ between these countries.

Anglo-Saxon countries

For the UK, USA and Canada positive returns have been found to changes in compulsory schooling. One of the reforms that have been studied most is the raising of the minimum school leaving age from 14 to 15 in the United Kingdom in 1947. This reform affected a large part of the British population and increased the number of years of schooling on average by about 0.4 years. This reform has been exploited by Oreopoulos (2006) and Devereux and Hart (2010) to estimate the returns to schooling. Depending on the specification and outcome measure used (log hourly wages or annual/weekly earnings), reduced form estimates of the effects of this change range from 1 to 7%. In addition, Grenet (2013)

41 There is also a large and fast growing body of research that uses changes of compulsory schooling laws to estimate its impact on outcomes of cognitive and non-cognitive skills, subjective well-being, health, crime, teenage childbearing and political involvement, see for example Grenet (2013) for a recent overview of this literature.

evaluates the effects of the raising of the minimum school leaving age from 15 to 16 in 1972 for the same country. He finds that this reform increased the average number of schooling years by about 0.3 years and that estimated returns are in the order of 2% using the log hourly wage as outcome measure. An important difference between the 1947-reform and the 1972-reform is that the latter was accompanied by certification,[42] whereas the former was not.

For the United States and Canada changes in the minimum school leaving age from 14 to 15 have been exploited by Oreopoulos (2006). He finds that the impact of these changes affected a smaller part of the population than in the UK as they increased the number of years of schooling on average by about 0.1 years. Estimated returns are also somewhat smaller and range between 1-2%.

Continental European countries

For Sweden, Germany, France and the Netherlands no returns are found to similar changes in compulsory schooling. Meghir and Palme (2005) evaluate a social experiment in Sweden where compulsory schooling was increased from seven or eight to nine years. Overall the change increased the years of schooling by about 0.3 years, but did not have an impact on earnings. Pischke and von Wachter (2008) use the extension of the duration of the basic vocational track by one additional year in Germany just after the Second World War. They find that the change increased the years of schooling by about 0.2 years but document zero returns. Grenet (2013) finds no returns for France, exploiting the Bedoin reform that was adopted in 1959, which raised the minimum school leaving age from 14 to 16 in 1967 and increased the years of schooling by about 0.3 years. Oosterbeek and Webbink (2007) find no returns to an extra year of basic vocational education in the Netherlands in which the extra year was accompanied by an increase of the minimum school leaving age from 15 to 16 in 1975.

These European studies show that it is not obvious that increases in duration of compulsory schooling would improve labour market outcomes. Moreover, the studies show a remarkable difference between the Anglo-Saxon countries and the countries from continental Europe. In the literature two different explanations are given for these differences. A first explanation, given by Grenet, is that it matters whether the additional year in school is accompanied by certification. Another explanation, given by Piscke and von Wachter, is

42 Especially the number of junior secondary schooling certificates increased due to the 1972-reform, see Grenet (2013).

that it matters whether relevant labor-market skills are learned. The aim of this chapter is to explore this issue further by exploiting a reform in the Netherlands that was specifically targeted at increasing the number of high school certificates. If certification is the leading explanation, an increase in the number of high school certificates should result in higher earnings and a better labor market position. If it matters what type of skills have been learned in school, this has not necessarily to occur: when no relevant (labor-market) skills are learned but only a certificate is obtained, an additional year in school may not result in higher earnings or a better labor market position.

It is important to note that the reform studied in this chapter is different from the one studied by Oosterbeek and Webbink. Whereas they exploit the extension of the duration of the basic vocational track by one year in 1975 which was accompanied by the rise of the minimum school leaving age from 15 to 16, I exploit the 1971-rise of the minimum school leaving age from 14 to 15 which was not accompanied by a change in the length or content of the curriculum. Hence, the curriculum remained the same in the 1971-reform.

4.3 Reform & institutional background

The compulsory schooling law was changed in 1971, with the argument that the years of compulsory schooling needed to be adapted to the 'current requirements that society imposes upon those who will participate in the labor market' (*Memorie van Toelichting, 1968*). The duration of compulsory education was extended from 8 to 9 years. This was equivalent to raising the minimum school leaving age from 14 to 15 as the starting age of compulsory schooling remained unchanged at six.[43] The reason for the change was that the old law, with 8 years of compulsory schooling, did not cover the complete duration of any track that would lead to a secondary school certificate. Primary school took 6 years, and finishing the lowest track of secondary school took at least another 3 years. Hence, a minimum of 9 years was required to obtain a high school diploma. The policy document that addressed the preparation of the law stated this in the following way: "One of the basic ideas [...] is that after six years of primary school every pupil should have the possibility to receive both general and vocational education that suits his abilities and capacities the best as possible. To realize this idea, it is therefore necessary to extent the duration of

43 More formally, the rule stated that a student had to go to school in the school year in which he became 7 years old.

compulsory schooling, as receiving a certificate of the lowest vocational track in secondary school (LBO) takes at least 3 years'' (*Memorie van Toelichting 1968*).

At that time, the Dutch secondary education system was a highly differentiated system. After six years of primary school, students (aged 12) were tracked into one of the four tracks of the secondary school system. These tracks can be considered as different levels of education. The two lowest tracks, LBO and MAVO, offered students a basic vocational program. Finishing these tracks took at least 3 years.[44] The third track (HAVO) prepared students for high vocational training and took 5 years. The highest track (VWO) gave access to university and took 6 years.

The main purpose of the policy change was to enable students, who otherwise would have dropped out of school, to obtain a high school diploma. These students were in particular (low ability) students who after eight years in school were either in one of the two lowest tracks of the secondary school system (LBO or MAVO) or still in primary school because of (multiple) retention (*Memorie van Toelichting 1968*). The individuals who were first affected by the new law were born in October 1956: they were 14 years old at the end of the school year 1970/1971 and had to stay one year longer in school than individuals born before that date. This October cutoff was based on the rights and obligations for students to enroll in primary school. The Appendix will provide more information on this rule.

The result of the increase in the minimum school leaving age was an increase in the high school completion rate. Figure 4.1 shows the relationship between high school completion and birth month. Each dot represents the fraction that completed high school (y-axis) for a cohort of individuals born in a given month (x-axis). The cohort of students that was first affected by the law, the cohort born in October 1956 has been rescaled to 0. The discontinuity in the high school completion rate is visible at that point; the rate suddenly rises by 2 percentage points, from 90% to 92%. This suggests that the change in compulsory schooling law resulted in a decrease in the high school dropout rate from 10% to 8%, which means a reduction of school dropout by 20%. In the remainder of this chapter, I will use this discontinuity to estimate the returns to the change in the compulsory schooling law in a regressions discontinuity framework.

44 Until 1975 around half of all graduates from Dutch basic vocational schools finished a 3-years program, the other half finished a 4 years program. In 1975 all 3 years programs were extended to four years. This change is studied by Oosterbeek & Webbink (2007).

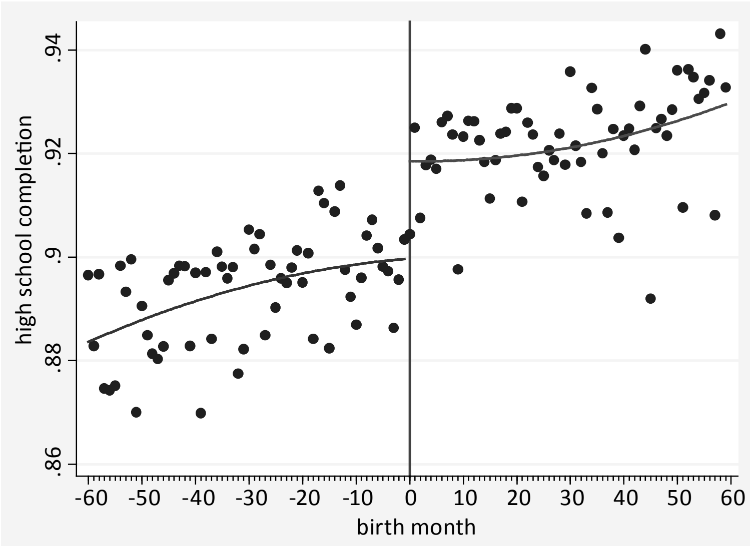


Figure 4.1: High school completion by birth month

Each dot represents the fraction that completed high school for a cohort of individuals born in a given month. The x-axis ranges from October 1951 (-60) to September 1960 (59). The cohort born in October 1956 is represented by 0 on the x-axis. Quadratic polynomial fitted at either side of the cutoff.

4.4 Empirical strategy

The aim of the empirical strategy is to estimate the causal effect of the change in the compulsory schooling law on education and labor market outcomes. I exploit the cutoff date of 1 October 1956 to estimate this impact. By using data on date of birth, education and labor-market position I can compare the outcomes of those born around this date. That is, I am able to evaluate the effect of the change in the compulsory schooling law in a regression discontinuity (RD)-framework (see Oreoupoulos, 2006; or Grenet, 2013, for examples).

The basic assumptions in RD-models are that individuals on both sides of the cutoff are very similar and that the relationship between date of birth and the outcome is smooth around the discontinuity. I follow the standard approach as proposed by Lee and Lemieux (2010) by estimating the following equation for each outcome Y_i,

(4.1) $$Y_i = \alpha_0 + \alpha_1 OCT1956_i + f_g(Z_i) + f_g(Z_i) * OCT1956_i + \alpha_2 \mathbf{X}_i + \varepsilon_i$$

4

Herein is $OCT1956_i$ a dummy for being born on or after 1 October 1956 and $f_g(Z_i)$ a smoothing polynomial of order g in birth month Z_i, with Z_i=0 referring to October 1956. I include a linear (g=1), square (g=2), cubic (g=3), quartic (g=4), quintic (g=5) and sextic (g=6) function of Z_i in my models and interactions of $OCT1956_i$ with this function ($f_g\left(Z_i\right)* OCT1956_i$). By including the interactions I allow the polynomial to be different at either side of the cutoff. By increasing the order, the polynomial should become flexible enough to pick up non-linear age effects.[45] In the analysis I will use the Akaike Information Criterion to choose the optimal order of the polynomial (see Lee and Lemieux, 2010). $\boldsymbol{X}_i$ is a vector of dummies for gender and ethnicity. Equation (4.1) is estimated by OLS and standard errors are adjusted for heteroskedasticity and clustering at the birth month level. The main parameter of interest is α_1 which represents the effect of the change in the compulsory schooling law, i.e. the effect of being born after the cutoff date implying 9 years of compulsory schooling in stead of 8 years. I can only interpret this coefficient as such if I may assume that the cohorts born on or after 1 October 1956 are not confronted with other changes in laws or regulations compared to those born before this date. To my knowledge there were no other laws implemented that affected this birth cohort.

4.5 Data

The data for the analysis come from the Dutch Labour Market Panel (*Arbeidsmarktpanel, Centraal Bureau voor de Statistiek*), which is a large panel dataset constructed by Statistics Netherlands. The sample is representative for the Dutch population. It contains information about the labor-market position of about 1.2 million Dutch inhabitants over the years 1999-2009. The data contain information on completed education, date of birth, income, employment, and social security dependency. I restrict the sample to those individuals who are born between October 1946 and September 1966. This means I have 10 year cohorts born before October 1956 and 10 after. This yields a sample of approximately 390,000 observations. In the analysis I will exploit all years except 1999 because no information on gross hourly wage is available for this year. Also, I focus the analysis on the period 2006-

45 Assuming that $E[\varepsilon_i|Z_i]$, the conditional expectation of the unobserved determinants of Y_i given the birth month, is continuous, we can approximate it by a polynomial of order g, and the approximation will become arbitrarily accurate as g goes to infinity ($g \to \infty$) (Cellini et al., 2010).

2009 because for the earlier years there are more missing values on this variable.[46] Other outcomes (see below) have been consistently measured over the period 2000-2009.

In the main analysis, I use as dependent variables three measures: (1) high school completion (2) a dummy for whether the respondent is employed in a given year and (3) log gross hourly wage in a given year. High school completion is a dummy that equals 1 if the respondent completed at least the lowest track of secondary school, LBO (equivalent to 9.5 years of formal schooling)[47] and 0 if the respondent only finished primary school (equivalent to 6 years of formal schooling). I use this variable rather than completed years of schooling because the aim of the law was to enable students to obtain a high school certificate in the lowest tracks of secondary school. The second measure, a dummy for working, is used because for low ability students one of the benefits of having a high school degree is that they are able to acquire a job rather than to earn more conditional on having a job.

Besides these main outcomes I will use dummies for being self-employed, and dummies for participation in social welfare, unemployment and disability benefits. In addition, I will use three labor-market related measures that cover a larger period of the respondents working life. A special feature of the Dutch Labor Market Panel is that it provides information about 1) the number of years that people had a paid job since age 15, 2) the number of years they were unemployed since age 15 and 3) the number of years they received disability benefits since age 15. These outcomes were obtained by surveys in which respondents were asked to retrospectively evaluate their employment status since their fifteenth birthday. These variables will be exploited in Section 4.7 when I look at the impact of the change on these outcomes.

As independent variable I use a dummy that equals 1 if the respondent is born on or after 1 October 1956 (hence aged 53 years or less in 2009) and 0 if he/she is born before that date.

Also, covariates are added to the models. In a regression discontinuity framework the most important covariate is the 'assignment' or 'running' variable. In my case, this is birth month. I rescale the birth month of the cohort that was first affected by the law to zero. That is, I assign a zero to the individuals born in October 1956. This means that those born in September 1956, August 1956 are assigned -1, -2, etcetera, and those born in November

46 I loose about half the number of observations for hourly wage over the period 2000-2005 which may raise concerns about the representativeness of this sample.

47 Based on the fact that about half of students finished a 3-year track and the other half a 4-year track in secondary education at that time. Hence, 6+(3+4)/2=6+3.5=9.5.

1956, December 1956 are assigned 1, 2 and so on (as in figure 4.1). Other covariates in the models are ethnicity and gender.

In table 4.I, I present descriptive statistics (i.e. averages) of the main outcomes and covariates for each birth cohort born between October of a given year and September next year. The discontinuity in high school completion is clearly visible: the rate rises from 0.90 to 0.92 if one switches from the year-cohort born before the cutoff date to the cohort born thereafter. However, there seems not to be a discontinuity in log gross hourly wage and employment; the values of these variables after the cutoff date are not much different from those before. This suggests that the change in the law did not have an impact on these outcomes.

4.6 The impact of the reform on education, employment and earnings

This section presents the estimates of the impact of the raising of the minimum school leaving age from 14 to 15 on high school completion, employment and earnings based on equation (4.1). It consists of two parts. In the first part the main estimates are shown for the outcomes measured in the last year of my dataset, 2009 (when the respondents born in October 1956 are about 53 years old). With the outcomes from this year various robustness analyses are performed with two estimation samples, the total sample and the low track sample, see below. In the second part estimates are given for each year over the period 2000-2009 for both estimation samples using the preferred specification. Hence, in this part I estimate the impact of the change in the law separately on the 2000-outcome, the 2001-outcome, 2002-outcome and so forth. These estimates will give the complete earnings and employment profiles for the population aged 43-53.

Part I: the effect of the reform on 2009-outcomes, total sample

Table 4.II presents estimates of the change of the compulsory schooling law on high school completion (panel A), employment (panel B), and earnings (panel C)[48] for the total sample. Columns (1)-(6) include a linear, square, cubic, quartic, quintic and sextic function of birth month respectively. At the bottom of each panel the Akaike Information Criterion (AIC) is reported which is a goodness-of-fit test on which we base the order of the polynomial that should be included. The lower the AIC, the better the fit is. In almost all analyses this

48 The estimation sample for earnings is smaller than for being employed, because information on earnings is only available for employed individuals.

Table 4.I: Descriptives for 2009-outcomes and covariates, total sample

Birth cohort	High school completion	Log hourly wage	Log annual earnings	Employed	Self-employed*	Disability benefits	Unemployment benefits	Welfare benefits	Years having paid job since age 15	Years unemployed since age 15	Years living on disability benefits since age 15	Dutch**	Female	N
Not affected by the compulsory schooling law														
October 1946 - September 1947	0.861	2.864	9.611	0.260	0.181	0.113	0.016	0.031	17.433	0.551	1.590	0.875	0.510	18856
October 1947 - September 1948	0.866	2.895	9.873	0.392	0.143	0.109	0.018	0.031	17.500	0.586	1.489	0.873	0.500	18358
October 1948 - September 1949	0.867	2.904	10.031	0.496	0.118	0.102	0.020	0.030	17.556	0.580	1.429	0.869	0.493	17680
October 1949 - September 1950	0.870	2.915	10.134	0.600	0.102	0.097	0.024	0.030	17.616	0.641	1.423	0.867	0.490	17184
October 1950 - September 1951	0.880	2.921	10.175	0.646	0.101	0.095	0.022	0.028	17.696	0.637	1.303	0.864	0.499	17439
October 1951 - September 1952	0.886	2.922	10.181	0.680	0.099	0.084	0.021	0.028	17.754	0.674	1.166	0.865	0.499	17913
October 1952 - September 1953	0.889	2.914	10.195	0.702	0.098	0.080	0.019	0.025	17.589	0.686	1.116	0.866	0.505	18198
October 1953 - September 1954	0.895	2.918	10.218	0.724	0.098	0.074	0.020	0.027	17.632	0.716	1.081	0.866	0.503	18424
October 1954 - September 1955	0.900	2.917	10.233	0.752	0.099	0.062	0.017	0.025	17.628	0.700	0.948	0.868	0.505	18816
October 1955 - September 1956	0.897	2.916	10.234	0.759	0.095	0.061	0.020	0.025	17.542	0.757	0.910	0.861	0.500	19180
Affected by the compulsory schooling law														
October 1956 - September 1957	0.918	2.917	10.241	0.779	0.097	0.056	0.018	0.024	17.486	0.748	0.846	0.860	0.506	19547
October 1957 - September 1958	0.922	2.907	10.240	0.793	0.102	0.051	0.016	0.023	17.372	0.768	0.825	0.861	0.506	19444
October 1958 - September 1959	0.922	2.913	10.245	0.794	0.101	0.047	0.019	0.025	17.218	0.758	0.737	0.855	0.510	20600
October 1959 - September 1960	0.920	2.897	10.238	0.801	0.101	0.043	0.018	0.024	16.966	0.764	0.657	0.844	0.510	20440
October 1960 - September 1961	0.929	2.895	10.239	0.809	0.100	0.041	0.019	0.023	16.704	0.776	0.679	0.856	0.513	20692
October 1961 - September 1962	0.928	2.890	10.219	0.811	0.105	0.037	0.018	0.025	16.467	0.744	0.640	0.848	0.512	20924
October 1962 - September 1963	0.929	2.885	10.213	0.821	0.107	0.036	0.017	0.022	16.164	0.739	0.572	0.842	0.508	21089
October 1963 - September 1964	0.936	2.888	10.218	0.824	0.108	0.034	0.018	0.022	15.791	0.736	0.536	0.837	0.506	20980
October 1964 - September 1965	0.939	2.880	10.203	0.820	0.109	0.034	0.018	0.023	15.307	0.707	0.501	0.832	0.507	21107
October 1965 - September 1966	0.943	2.880	10.192	0.821	0.111	0.030	0.018	0.021	14.879	0.663	0.480	0.832	0.508	20661

* Conditional on being employed. **The Dutch definition *autochtoon* is used: both parents were born in the Netherlands. In our regressions we also distinguish between Moroccan, Turkish, Surinam and Antilles

4

statistic is smallest for third degree polynomials or higher. This indicates that at least a cubic should be included to let the polynomial to be flexible enough for our RD-design.[49] Hence, results based on linear and squares in birth month should not be considered reliable.

Panel A shows that the rise of the minimum school leaving age from 14 to 15 significantly increased the level of schooling. The estimates are robust to the inclusion of different orders of the birth month polynomial. Based on the most flexible specification in column (6), I find that the fraction of individuals that obtained a high school certificate has increased with about 1.8 percentage points. This is equivalent to an increase of about 0.1 years of schooling, which is about the same size as the impact of the changes in compulsory schooling laws on education in Canada and the USA (see Section 4.2). However, in contrast to these Anglo-Saxon countries the change in the compulsory schooling law did not lead to an increase in earnings or a higher probability of being employed in 2009. The estimated coefficients in panels B and C are precise, close to zero and insignificant. For earnings the estimated coefficients are even negative in all specifications; the point estimate in column (5) indicates that the change in the law decreased income by 1%. Hence, the change in the compulsory schooling law caused an increase in the level of schooling but did not lead to higher wages.

Part I: the effect of the reform on 2009-outcomes, low track sample

The purpose of the change in the law was to enable low ability students to obtain a high school certificate. These students were in particular students who after eight years of schooling were in one of the two lowest tracks of the secondary school system, LBO or MAVO. Hence, one may expect that this group would benefit most from the new law. One way of investigating this is to look at the instrumental variables estimates of the returns to high school completion. They would give the returns to high school completion for those who comply with the change, such that one could obtain an idea of whether those affected would have (financially) benefited from the change of the law. These estimates are easily obtained by dividing the reduced form coefficients in panels B or C by the first stage

49 The estimated coefficients of the polynomial are highly significant when using a linear or square function of birth month. The estimated coefficients turn insignificant when including a third order or higher.

Table 4.II: Estimated effects on 2009-outcomes, total sample

Panel A	**Dependent variable: high school completion**					
Independent variable	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Dummy for individual born in or after October 1956	0.0105***	0.0142***	0.0178***	0.0156***	0.0154***	0.0183***
	(0.00180)	(0.00271)	(0.00346)	(0.00412)	(0.00497)	(0.00578)
Birth month controls	linear	square	cubic	quartic	quintic	sextic
Observations	387532	387532	387532	387532	387532	387532
R-squared	0.055	0.055	0.055	0.055	0.055	0.055
Akaike Information Criterion (AIC)	122084.5	122084.1	122081.7	122084.3	122088.3	122089.6
Panel B	**Dependent variable: employed 2009**					
Dummy for individual born in or after October 1956	-0.0742***	0.0427***	0.00395	-0.00467	0.00524	0.00599
	(0.0112)	(0.00712)	(0.00666)	(0.00874)	(0.0115)	(0.0138)
Birth month controls	linear	square	cubic	quartic	quintic	sextic
Observations	387532	387532	387532	387532	387532	387532
R-squared	0.152	0.159	0.159	0.159	0.160	0.160
Akaike Information Criterion (AIC)	423237.9	419869	419640.7	419640.1	419610.3	419613.3
Panel C	**Dependent variable: earnings 2009**					
Dummy for individual born in or after October 1956	-0.00919**	0.0121**	-0.00264	-0.00349	-0.0114	-0.00239
	(0.00379)	(0.00547)	(0.00671)	(0.00832)	(0.00967)	(0.0117)
Birth month controls	linear	square	cubic	quartic	quintic	sextic
Observations	248081	248081	248081	248081	248081	248081
R-squared	0.103	0.104	0.104	0.104	0.104	0.104
Akaike Information Criterion (AIC)	287383.3	287339.3	287332.3	287335.9	287337.2	287339.2

Notes: Each cell is an OLS-regression. All regressions include dummies for gender and ethnicity. Standard errors (in parentheses) are adjusted for heteroskedasticity and clustering at the birth month level .*** p<0.01, ** p<0.05, * p<0.1

4

coefficients from panel A.[50] However, as the reduced form estimates are close to zero and insignificant (and for earnings even negative), these instrumental variables estimates will be little informative.

To investigate whether low ability students may have benefited from the law, I focus on a subsample of the total sample, consisting of individuals who reported that their highest level of completed education is at most a certificate of one of the two lowest tracks of secondary school (LBO or MAVO). The individuals in this sample either obtained a LBO/MAVO high school diploma or only finished primary school. Hence, individuals with a diploma higher than MAVO are excluded. I define this sample as the low track sample (see table 4A.1 and figure 4A.1 in the Appendix for descriptive statistics and a discontinuity graph).

A possible concern with this approach is that this sample is selected on the outcome. Such a selective sample may give biased results because I exclude individuals that could have obtained certificates higher than MAVO through the change in the law. However, this problem might be mitigated by the fact that at that time low ability students in the lowest tracks of the secondary school system (LBO, MAVO) often did not end up in higher tracks.

Table 4.III presents estimates of the change in the compulsory schooling law on the main outcomes for the low track sample. As expected, the estimated coefficient for high school completion is larger in this sample than in the total sample (panel A). Based on the most flexible specification in column (6), high school completion among low ability students increased 5 percentage points through the change in the compulsory schooling law (equivalent to an increase of about 0.2 years of schooling). Again, however, I do not find evidence that the change resulted in higher earnings or higher probability to be employed; the estimates in panels B and C are not significantly different from zero. Hence, the change in the compulsory schooling law increased the level of schooling for students of low ability, but the change has not been financially beneficial for them.

Robustness

Estimates from a regression discontinuity design could be sensitive to the window of the estimation sample chosen and the degree of the smoothing polynomial included (see Lee and Lemieux, 2010). In this section I investigate to what extent my estimates change when

50 Estimates and standard errors are obtained by an instrumental variables approach to a Mincerian wage equation in which high school completion is instrumented by the dummy for being born on or after 1 October 1956. Estimates in panel A in tables 4.II and 4.III should then be considered first stage estimates. The F-statistic of the first stages are above 10 (they range between 12 and 16), indicating that there is no weak instrument problem.

Table 4.III: Impact of 1971-ROSLA on 2009-outcomes, low track sample

Panel A	**Dependent variable: high school completion**					
Independent variable:	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Dummy for individual born in or after October 1956	0.0279***	0.0356***	0.0441***	0.0490***	0.0495***	0.0507***
	(0.00489)	(0.00752)	(0.00953)	(0.0110)	(0.0122)	(0.0129)
Birth month controls	linear	square	cubic	quartic	quintic	sextic
Observations	125767	125767	125767	125767	125767	125767
R-squared	0.057	0.057	0.057	0.057	0.057	0.057
Akaike Information Criterion (AIC)	150421.3	150423.2	150423.4	150425.7	150429.1	150429.7
Panel B	**Dependent variable: working 2009**					
Dummy for individual born in or after October 1956	-0.0603***	0.0399***	0.00917	-0.00209	0.0161	0.0119
	(0.0106)	(0.00905)	(0.0105)	(0.0135)	(0.0164)	(0.0199)
Birth month controls	linear	square	cubic	quartic	quintic	sextic
Observations	125767	125767	125767	125767	125767	125767
R-squared	0.159	0.164	0.164	0.164	0.164	0.164
Akaike Information Criterion (AIC)	157332.3	156618	156587.2	156585.7	156573.9	156577.6
Panel C	**Dependent variable: earnings 2009**					
Dummy for individual born in or after October 1956	-0.0136**	0.0162**	0.00156	0.0140	0.0223	0.0165
	(0.00585)	(0.00755)	(0.0101)	(0.0124)	(0.0152)	(0.0182)
Birth month controls	linear	square	cubic	quartic	quintic	sextic
Observations	66443	66443	66443	66443	66443	66443
R-squared	0.117	0.118	0.118	0.118	0.118	0.118
Akaike Information Criterion (AIC)	45810.34	45763.7	45763.1	45764.54	45765.8	45769.5

Notes: Each cell is an OLS-regression. All regressions include dummies for gender and ethnicity. Standard errors (in parentheses) are adjusted for heteroskedasticity and clustering at the birth month level .*** p<0.01, ** p<0.05, * p<0.1

I restrict the sample to 5 and 3 years distance to the cutoff and use a square and linear function of birth month as smoothing polynomial respectively.

The results of my robustness analysis are shown in table 4.IV. In panel A I report the results of regressions where I use the sample of individuals that are born between October 1951 and September 1960, which is a 5 year distance to the cutoff. In this estimation sample I include a square function of birth month. In panel B I restrict the sample further to individuals that are born between October 1953 and September 1959, which is a 3 year distance to the cutoff. Because of the smaller sample size, I only include a linear term in birth month. By using this estimation window, I exclude individuals that also were affected by the 1975-reform studied by Oosterbeek and Webbink. The estimates in panel A and B are given for the total sample and the low track sample.

Table 4.IV shows that the main results do not change: also when using different estimation windows and smoothing polynomials the estimates are about the same size for high school completion and insignificant for outcomes employed and earnings. In the next part I will give the estimates for each year over the period 2000-2009, thereby showing employment and income profiles.

Part II: employment and income profiles over 2000-2009

The outcomes used in tables 4.II and 4.III come from the year 2009 when the respondents who were first affected by the law (born in October 1956) are about 53 years old. Bhuller, Mogstad and Salvanes (2012) point out that estimates may differ at different ages because of life cycle bias. In addition, the effect on lifetime income may be different from the effect on income measured at a particular age. Perhaps respondents gained from the change in the law earlier in life, and these gains are not reflected in what is measured in 2009.[51] Therefore, I will investigate returns at other ages by estimating the impact of the change of the law on employment and earnings for each year (2000, 2001, 2002, etc) in the dataset. Hence, I estimate the effect of the change in the law on earnings and employment when those born in October 1956 are 43 years old in 2000, 44 years old in 2001 and so on. The results of these analyses are summarized in figures 4.2 and 4.3. I call these figures the employment and income profiles. The y-axis represents the estimated effect (solid line)

51 Because 2009 is a recession year, one could argue that in particular in recessions persons would benefit from extra schooling such that we should have seen returns to extra schooling especially in this year.

Table 4.IV: Impact of 1971-ROSLA on 2009-outcomes for various estimation windows

	Dependent variable:		
	high school completion	employed 2009	earnings 2009
Independent variable:	**(1)**	**(2)**	**(3)**
Panel A: 5 year distance to cutoff			
Total sample			
Dummy for individual born in or after October 1956	0.0161***	0.00665	-0.000
	(0.00378)	(0.00701)	(0.00699)
Observations	193254	193254	132705
Low track sample			
Dummy for individual born in or after October 1956	0.0472***	0.0102	0.00794
	(0.0104)	(0.0112)	(0.0106)
Observations	61245	61245	35767
Panel B: 3 year distance to cutoff			
Total sample			
Dummy for individual born in or after October 1956	0.0154***	0.00518	-0.00232
	(0.00316)	(0.00594)	(0.00604)
Observations	116011	116011	80285
Low track sample			
Dummy for individual born in or after October 1956	0.0441***	0.0106	0.00893
	(0.00876)	(0.00988)	(0.00893)
Observations	36442	36442	21437

Notes: Each cell is an OLS-regression. The regressions in panel A include a square in (rescaled) birth month and interactions of the dummy for individual born in or after October 1956 with this square. In panel B it is a linear term. All regressions include dummies for gender and ethnicity. Standard errors (in parentheses) are adjusted for heteroskedasticity and clustering at the birth month level .*** p<0.01, ** p<0.05, * p<0.1

and the 95% confidence interval (dotted lines). The x-axis represents the year used in the estimation, which ranges from 2000 to 2009. Estimates are based on the sample with a 5 year distance to the cutoff and a square in birth month that differs at either side of the cutoff. The order of the polynomial is based on the Akaike Information Criterion.[52] The estimates are robust to the various robustness analyses presented in the first part.

The figures show that the estimates for employment and earnings (solid line) are not significantly different from zero in all years 2000-2009: over this period the zero is covered by the 95% confidence interval (dotted lines). This holds for both the total and the low track sample.[53] For the total sample, the estimates for income are even negative in most years. The 95% confidence interval shows that positive effects of 1% can be excluded. In comparison, for Anglo-Saxon countries returns have been found between 1 and 7% (see Section 4.2). Hence, the results for the Netherlands are different from those for the Anglo-Saxon countries and do not depend on the year the outcomes have been measured, nor on the estimation window used or the order of the birth month polynomial included.

4.7 Why zero returns?

Why did the raising of the minimum school leaving age from 14 to 15 in the Netherlands not lead to higher wages while researchers do find returns in the Anglo-Saxon countries when exploiting similar changes? In this section, I investigate three possible explanations.

First I investigate whether the zero returns are a result of possible wage rigidity differences between the Netherlands and the Anglo-Saxon countries. Second, I investigate two other explanations that have previously been given in the literature: whereas Grenet postulates that it matters whether the additional year in school is accompanied by certification, Pischke and von Wachter postulate that it matters whether labor-market relevant skills are learned in the extra year.

52 In many estimations with this sample (i.e. the 5-year distance to the cutoff sample), the AIC was smallest when using a square function of birth month.

53 Except for employed in the low track sample in 2002.

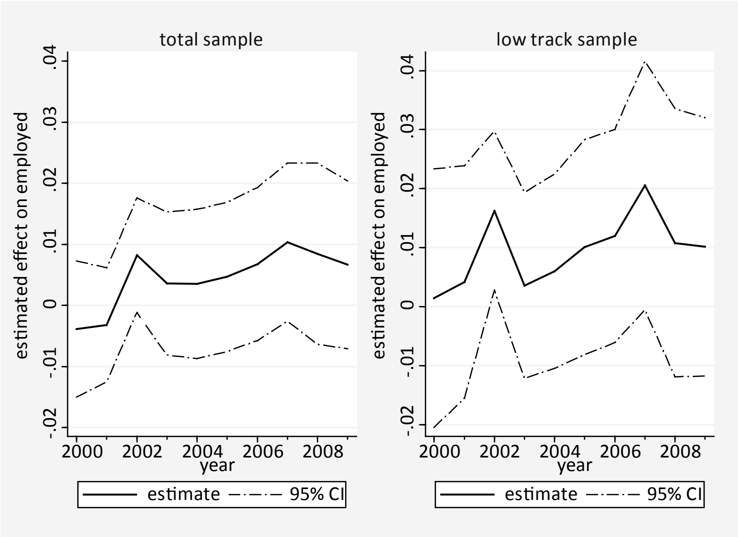


Figure 4.2: The impact of 1971-ROSLA on being employed in years 2000-2009

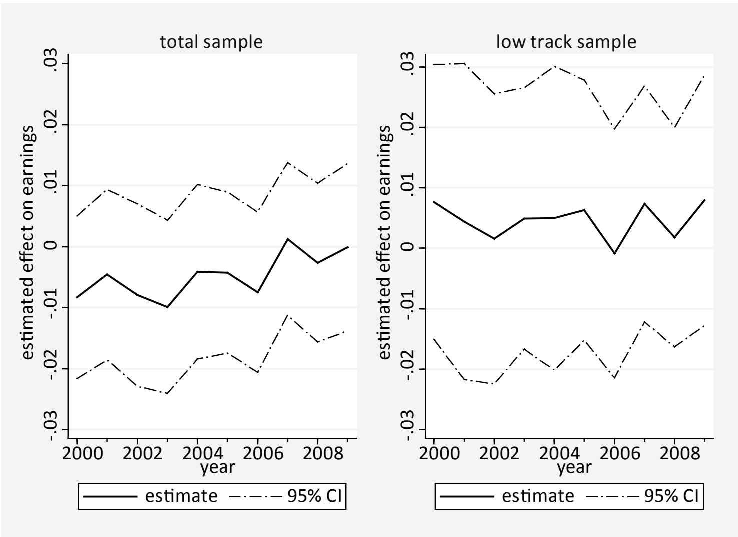


Figure 4.3: The impact of 1971-ROSLA on log hourly wage in years 2000-2009

Wage rigidity?

An explanation for not finding returns could lie in the wage setting institutions in the Netherlands, which, in contrast to those in the UK or USA, may prevent the adjustments necessary to reflect any returns. For example, the fraction of workers that is covered by collective bargaining agreements is much higher in the Netherlands than in the UK or USA: 82% versus 35 or 12% respectively (OECD, 2009).[54] This could mean that productivity differences between individuals are not reflected in earnings because negotiated wages are kept higher than the market-clearing level wages. In that case the change in the compulsory schooling law may have contributed to the productivity of the individuals who were affected, but this higher productivity is not reflected in their wages. I give some suggestive evidence that may rule out this possibility.

First, because employers should be more willing to employ the more productive individuals with the extra schooling than the equally expensive individuals without the extra schooling, it might be expected that employers hire them more often. But as shown in the previous section, the estimated effect of the change in the compulsory schooling law on being employed is not significantly different from zero.

Second, if individuals notice that their extra skills are not recognized or rewarded by employers, they may decide to start their own business. In that case, it might be expected that they would be self-employed more often. However, estimates of the impact of the reform on a dummy for self-employed are also not significantly different from zero, see figure 4.4. This suggests that, although the change in the compulsory schooling led to a higher educated population, it did not lead to a more productive one. Pischke and von Wachter (2007) and Grenet (2013) perform similar analyses for their countries and also find no effects on being (self-) employed, suggesting that wage rigidity is also not important for explaining the zero returns for Germany and France.

What matters most: certification or skills?

Another possible explanation for not finding returns is that the additional year in school was not accompanied by academic credentials (i.e. certifications). This explanation is given by Grenet. In his paper, he finds no returns for France when exploiting the Bedoin Reform which raised the minimum school leaving age by 2 years from 14 to 16 in 1967, whereas he does find returns for the UK in the 1972-reform that increased the minimum school leaving age by only one year (from 15 to 16). Because the latter was accompanied by a certificate,

54 http://www.oecd.org/els/emp/43116624.pdf.

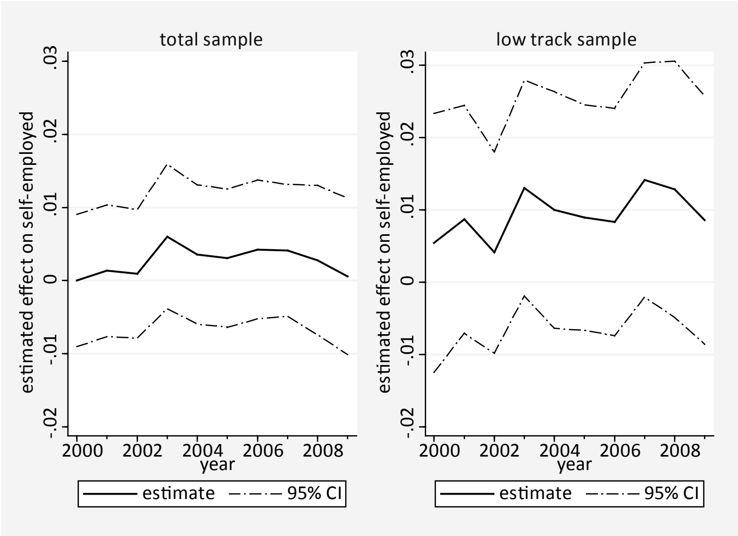


Figure 4.4: The impact of 1971-ROSLA on being self-employed in years 2000-2009 (conditional on being employed)

and the former not, he argues that the certification differences between the countries could explain the differences in returns.

However, the results in the previous sections suggest that it is not certification that explains the zero returns. In both the UK and the Netherlands the raising of the minimum school leaving age led to more certificates, but only for the UK the change also led to higher earnings. A reason for this difference might be that the group of people affected by the 1971-ROSLA from 14 to 15 in the Netherlands was smaller than the group of people affected by the 1972-ROSLA from 15 to 16 in the UK (+0.1 years of schooling versus +0.3 years of schooling). However, the impact of the reform on education in the Netherlands was about the same size as in the USA and Canada where positive effects of increases in compulsory schooling on earnings have been found. Hence, one might expect that estimates of the effect of the Dutch reform on earnings would lie in the same ballpark of the estimates found for these countries. But as shown in figure 4.3, effects larger than 1% can be excluded.

Further suggestive evidence that certification does not explain the zero results may come from a comparison between the two reforms in the UK: the 1947-ROSLA from 14 to 15 (without certificate) and the 1972-ROSLA from 15 to 16 (with certificate). If academic credentials were important then it would have been less likely to find returns from the

1947-reform as this reform was not accompanied by such credentials. However, as shown in Section 4.2, in both reforms positive returns have been found. This makes a story that heavily relies on certification less likely. Moreover, both reforms led to more cognitive skills such that it is impossible to disentangle which mechanism was responsible for the higher earnings.[55] In the 1972-reform, for example, the higher earnings could be due to the extra skills learned in the additional year, the certificate or the combination of both. Hence, a comparison of these two reforms might be more in favor of a skill-based explanation than of an explanation that solely or heavily relies on certification.

Such a skill based explanation is given by previous studies that find zero returns to increases in compulsory education. Pischke and von Wachter give some suggestive evidence for this explanation in their study for Germany. They show that German individuals perform much better on quantitative and mathematics test than people in a similar position in the UK, Canada or USA, thereby suggesting that most labor-market relevant skills were learned earlier in the education system in Germany than in the Anglo-Saxon countries. This would mean that children did not learn more relevant skills in the additional year in school than in the year out of school. Oosterbeek and Webbink follow the same line of reasoning for the Netherlands, which is a country very similar to Germany, and also come up with this skill based explanation for the zero returns to the Dutch 1975-reform.

In this chapter, which is also set in the Netherlands, I will add a few pieces of evidence that may support this skill based view. First, I investigate whether the raising of the minimum school leaving age from 14 to 15 led to higher cognitive skills as measured by the International Adult Literacy Survey (IALS) in 1994 (when those born in October 1956 are about 38 years old). Table 4.V presents reduced form regressions of the impact of the reform on three measures of literacy skills (see figures 4A.2-4A.4 for graphs). As can be seen, the estimates are not significantly different from zero. It should be noted, though, that the sample size is limited, which may make it difficult to detect any small effects of the reform on these skills. However, the estimates do rule out modest or large impacts.[56] This stands in contrast with the UK-reforms in which positive effects on cognitive skills have been found.

55 See Banks and Mazonna (2012) for the effect of the 1947-reform on cognitive skills and Grenet (2013) for the effect of the 1972-reform on literacy skills.

56 The raw scores have been used as outcome measure. If I translate these effect sizes in percentage of standard deviations I am able to rule out effects larger than 0.2σ using the 95% confidence interval.

Table 4.V: Impact of 1971-ROSLA on literacy skills from Adult Literacy Survey 1994

	Dependent variable:		
Independent variable:	**prose literacy score**	**document literacy score**	**quantitative literacy score**
Dummy for being born in or after 1957	-0.762	-0.943	-0.315
	(4.066)	(4.893)	(5.055)
Observations	1517	1517	1517
R-squared	0.018	0.018	0.010

Notes: Each cell is an OLS-regression. The regressions include a linear term of rescaled birth year (with 0 referring to birth year 1957) and the interaction of the dummy for being born in or after 1957 with this term. Birth year is constructed by subtracting age from the survey year, hence 1994-age. Sample is restricted to those born between 1946 and 1966. All regressions include a dummy for gender. Standard errors (in parentheses) are adjusted for heteroskedasticity and clustering at the birth year level .*** p<0.01, ** p<0.05, * p<0.1

4

Second, I investigate whether the reform led to gains in other labor-market outcomes than earnings. If there were any (small) effects of the reform on relevant labor-market skills, one might expect that the change would affect other labor-market outcomes such as the probability to live on disability benefits.[57] In figures 4.5-4.8 I present estimates of the impact of the reform on dummies for participation in disability, unemployment, and social welfare benefits. Estimates are based on the sample with a 5 year distance to the cutoff and a square in birth month that differs at either side of the cutoff. As can be seen, the estimates of the effects of the change in the compulsory schooling law on these three measures are insignificant in almost all years.[58]

57 The acquisition of those skills may lead to other type of jobs that may be less stressful. This in turn may reduce the risk to become disabled.

58 Except for unemployment benefits in 2007.

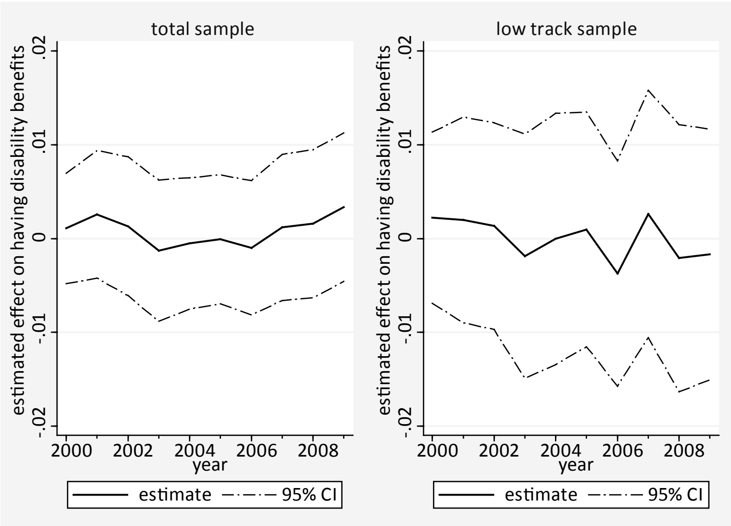


Figure 4.5: The impact of 1971-ROSLA on having disability benefits in years 2000-2009

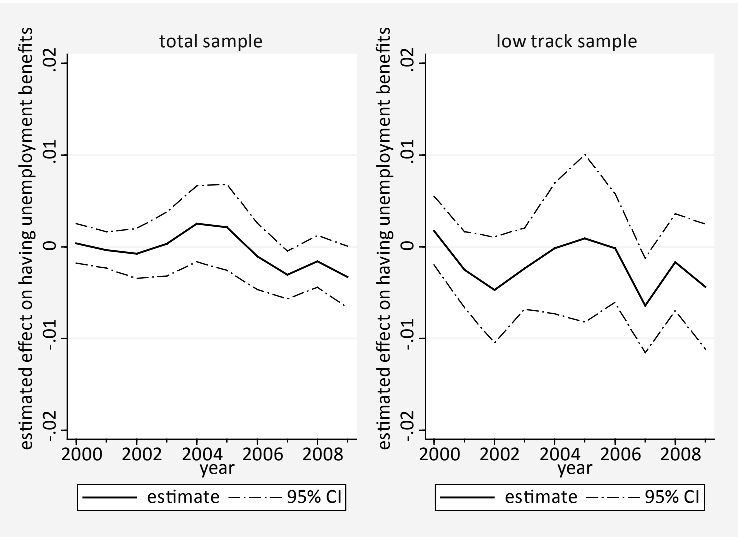


Figure 4.6: The impact of 1971-ROSLA on having unemployment benefits in years 2000-2009

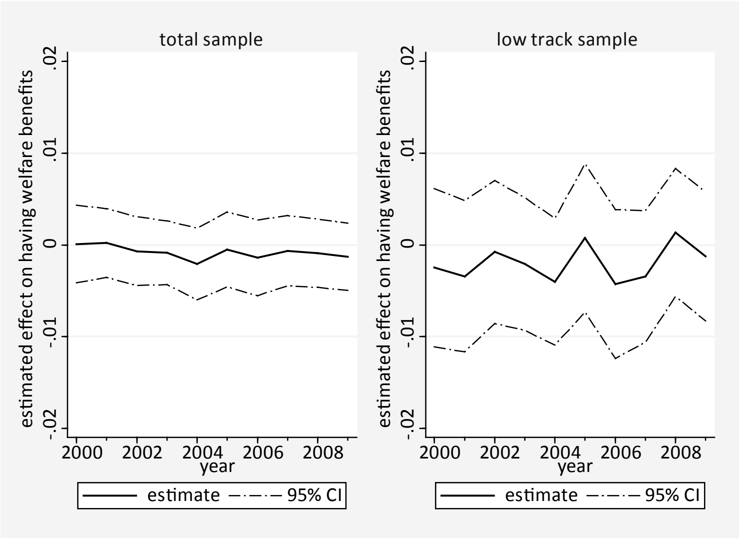
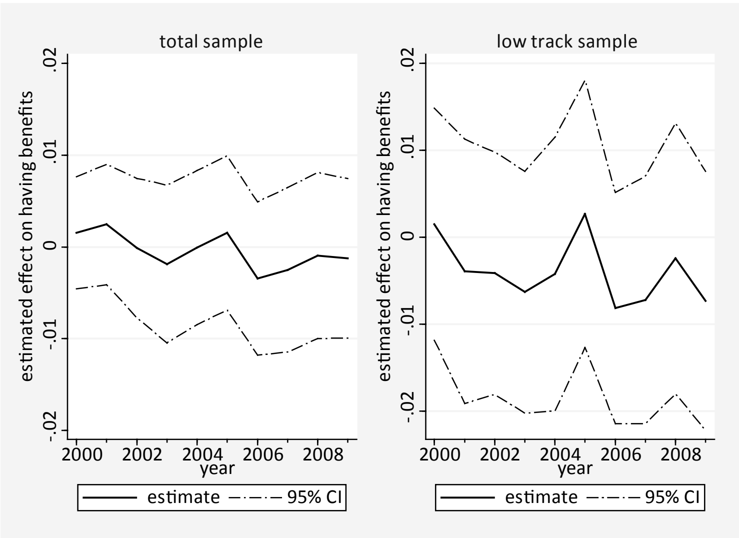


Figure 4.7: The impact of 1971-ROSLA on having welfare benefits in years 2000-2009



*Unemployment, welfare or disability benefits

Figure 4.8: The estimated effect of the change in the law of having benefits* in years 2000-2009

Finally, I investigate whether individuals may have gained from the change before age 43, as estimated effects on outcomes only go back as far as the year 2000 (when those born in October 1956 are about 43 years old). One might argue that the gains from the change in the law are internalized in their adolescent years, meaning that the outcomes measured in 2000-2009 do not capture the total number of years being (un)employed or having a paid job since the respondents left school. As pointed out in the data section, a nice feature of the Dutch Labour Market Panel is that it provides information on the number of years that people had a paid job, were unemployed and living on disability benefits since their fifteenth birthday. Regressions results with these three outcomes are presented in columns (1)-(3) of table 4.VI. They show that the change in the compulsory schooling law did not result in more years of having a paid job, less years of being unemployed or less years of living on disability benefits since age 15. Hence, the evidence presented in this section suggests that the change in the compulsory schooling law did not lead to higher literacy skills or to better labor-market outcomes in general. This can be interpreted as evidence supporting a skill-based explanation for the zero returns found in this chapter.

4.8 Conclusions

In this chapter I evaluated the raising of the minimum school leaving age from 14 to 15 in the Netherlands in 1971. The policy goal of this change was to increase the number of high school certificates. The analysis showed that the change reduced the high school dropout rate by approximately 20%. However, there were no benefits in terms of employment or higher wages when the population was aged 43-53. These results are consistent with previous work from continental Europe that documents zero returns to compulsory schooling, but stands in contrast with that from Anglo-Saxon countries in which positive returns are found. As a reason for not finding returns I give some suggestive evidence in support of the skill-based explanation that no more labor-market relevant skills were learned during this extra year of school compared to those skills previously learned out of school. I show that the change also did not lead to gains in literacy skills, nor in labor market outcomes other than earnings.

The contributions of these results to the current literature will be twofold. First, I give an additional piece of evidence that changes in compulsory schooling might not be effective in improving the productivity of the population. Second, I show that differences in returns between countries are not necessarily explained by differences in certification. In both the UK and the Netherlands the raising of the minimum school leaving led to more certificates, but only for the UK the change also led to higher (literacy) skills and earnings.

Table 4.VI: Impact of 1971-ROSLA on employment and living on disability benefits since age 15 of the respondent

	Years having job since age 15	Years unemployed since age 15	Years disabled since age 15
Independent variable:	**(4)**	**(5)**	**(6)**
Panel A: 5 year distance to cutoff			
Total sample			
Dummy for individual born in or after October 1956	-0.0448	0.00498	0.0665
	(0.0553)	(0.0323)	(0.0406)
Observations	186105	190841	190958
Low track sample			
Dummy for individual born in or after October 1956	0.0542	-0.0910	0.106
	(0.0975)	(0.0641)	(0.0755)
Observations	56858	59861	59901
Panel B: 3 year distance to cutoff			
Total sample			
Dummy for individual born in or after October 1956	-0.00388	0.0106	0.0327
	(0.0471)	(0.0275)	(0.0342)
Observations	111845	114581	114622
Low track sample			
Dummy for individual born in or after October 1956	0.0637	-0.0682	0.0902
	(0.0811)	(0.0565)	(0.0644)
Observations	33910	35627	35626

Notes: Each cell is an OLS-regression. The regressions in panel A include a square in (rescaled) birth month and interactions of the dummy for individual born in or after October 1956 with this square. In panel B it is a linear term. All regressions include dummies for gender and ethnicity. Standard errors (in parentheses) are adjusted for heteroskedasticity and clustering at the birth month level .*** p<0.01, ** p<0.05, * p<0.1

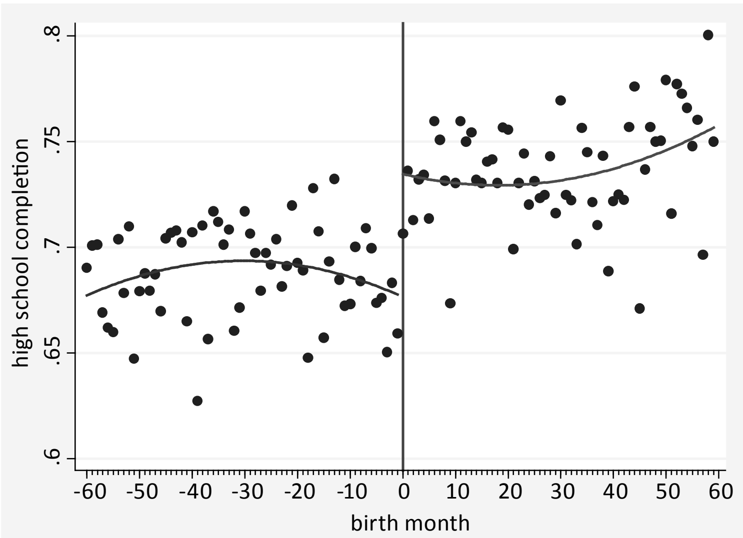
4

Taken together, this suggests that (further) increasing compulsory schooling may be ineffective for improving the productivity of the population if no labor market relevant skills are learned in the additional schooling year(s).

4.9 Appendix

Derivation of the October cutoff date

A Dutch school year runs from August 1 of a given year to July 31 next year. In the period under investigation, students had to go to school in the school year in which they became 7 years old. There was an exception, however, for students born in August or September: they had the right to enroll in the current school year (starting in August) or in the school year prior to this year (starting in August one year earlier). Hence, students born from October onwards had to start primary school in the current school year. This means that the students born in October 1956 enrolled primary school in the school year 1963/1964 (i.e. started school on the first of August 1963). They completed 8 years of compulsory schooling on 31 July 1971 (6 years of primary school, and 2 years of secondary school if they did not retain). They had to complete a ninth compulsory year in the school year 1971/1972.



Each dot represents the fraction that completed high school for a cohort of individuals born in a given month. The x-axis ranges from October 1951 (-60) to September 1960 (59). The cohort born in October 1956 is represented by 0 on the x-as. Quadratic polynomial fitted at either side of the cutoff.

Figure 4A.1: Discontinuity in low track sample

Table 4A.1: Descriptives for 2009-outcomes and covariates for the low ability sample

Birth cohort	High school completion	Log hourly wage	Log annual earnings	Employed	Self-employed*	Disability benefits	Unemployment benefits	Welfare benefits	Years having paid job since age 15	Years unemployed since age 15	Years living on disability benefits since age 15	Dutch**	Female	N
not affected by the compulsory schooling law														
October 1946 - September 1947	0.668	2.594	9.230	0.195	0.174	0.140	0.016	0.047	15.978	0.597	2.081	0.883	0.626	7870
October 1947 - September 1948	0.673	2.609	9.477	0.302	0.133	0.136	0.019	0.048	16.106	0.647	1.938	0.873	0.612	7510
October 1948 - September 1949	0.666	2.643	9.651	0.394	0.113	0.133	0.019	0.048	16.169	0.615	1.941	0.875	0.595	7041
October 1949 - September 1950	0.670	2.667	9.781	0.482	0.094	0.126	0.022	0.047	16.272	0.728	1.960	0.865	0.592	6782
October 1950 - September 1951	0.679	2.645	9.793	0.528	0.086	0.122	0.024	0.047	16.375	0.715	1.794	0.861	0.594	6499
October 1951 - September 1952	0.681	2.651	9.795	0.565	0.078	0.117	0.022	0.048	16.411	0.786	1.693	0.855	0.592	6398
October 1952 - September 1953	0.686	2.641	9.795	0.576	0.080	0.109	0.021	0.046	16.222	0.792	1.516	0.863	0.603	6434
October 1953 - September 1954	0.696	2.648	9.846	0.602	0.089	0.108	0.021	0.049	16.396	0.794	1.577	0.854	0.587	6378
October 1954 - September 1955	0.695	2.648	9.862	0.634	0.090	0.089	0.019	0.048	16.449	0.828	1.362	0.851	0.583	6180
October 1955 - September 1956	0.681	2.626	9.835	0.637	0.082	0.095	0.026	0.049	16.378	0.939	1.378	0.843	0.570	6173
Affected by the compulsory schooling law														
October 1956 - September 1957	0.728	2.630	9.838	0.662	0.088	0.086	0.024	0.048	16.392	0.903	1.319	0.839	0.569	5900
October 1957 - September 1958	0.739	2.634	9.877	0.685	0.096	0.081	0.019	0.047	16.482	0.952	1.358	0.836	0.554	5810
October 1958 - September 1959	0.731	2.629	9.868	0.680	0.088	0.082	0.020	0.055	16.266	1.061	1.245	0.820	0.555	6001
October 1959 - September 1960	0.726	2.627	9.882	0.688	0.087	0.074	0.022	0.055	15.995	1.070	1.069	0.803	0.556	5983
October 1960 - September 1961	0.755	2.614	9.894	0.712	0.094	0.065	0.020	0.051	16.034	1.058	1.153	0.820	0.540	5988
October 1961 - September 1962	0.755	2.634	9.896	0.709	0.091	0.065	0.023	0.054	15.885	1.040	1.084	0.807	0.535	6162
October 1962 - September 1963	0.749	2.613	9.848	0.716	0.094	0.064	0.022	0.053	15.747	1.112	1.018	0.797	0.530	5983
October 1963 - September 1964	0.763	2.612	9.870	0.719	0.104	0.067	0.024	0.049	15.446	1.115	0.951	0.786	0.505	5698
October 1964 - September 1965	0.772	2.615	9.877	0.715	0.100	0.062	0.025	0.054	15.093	1.147	0.869	0.782	0.501	5644
October 1965 - September 1966	0.778	2.613	9.863	0.727	0.106	0.058	0.026	0.048	14.661	1.096	0.894	0.771	0.509	5333

* Conditional on being employed. **The Dutch definition *autochtoon* is used: both parents were born in the Netherlands. In our regressions we also distinguish between Moroccan, Turkish, Surinam and Antilles

4

Figure 4A.2-4A.4: Birth year 1957 is represented by zero on the x-axis; at each dot the number of observations is given that was used for the calculation of the averages.

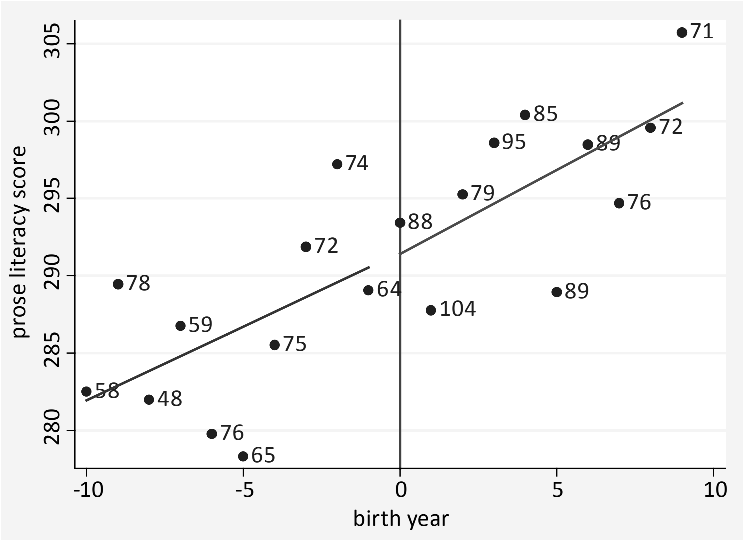


Figure 4A.2: No impact of 1971-ROSLA on prose literacy

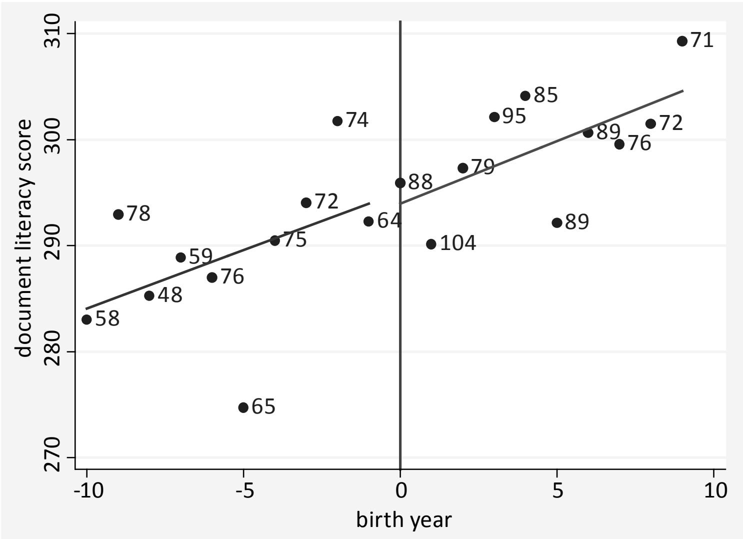


Figure 4A.3: No impact of 1971-ROSLA on document literacy

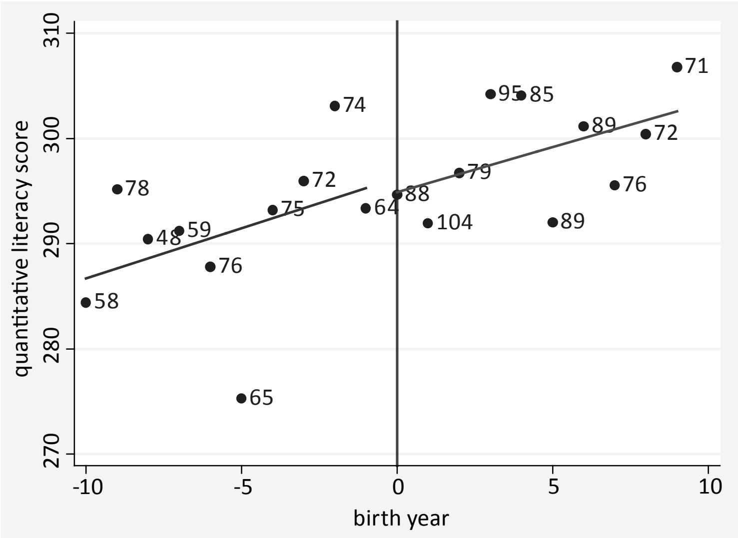


Figure A.4: No impact of 1971-ROSLA on quantative literacy

5

Do better school facilities yield more science and engineering students?[59]

59 This is joint work with Dinand Webbink.

Abstract

This chapter evaluates the effects of a subsidy program targeted at improving the school facilities for biology, physics and chemistry in secondary schools. The goal of this policy was to increase the enrollment rate in science and engineering (S&E) related courses at secondary and subsequent education institutions. The subsidy was assigned to schools based on a priority score reflecting the ambition level of schools to improve student achievement. Schools with scores below a threshold value did not receive subsidy. We exploit the assignment procedure in a regression discontinuity framework to estimate the impact of the subsidy on student outcomes. We find that the subsidy increased the enrollment rate in S&E-related courses in secondary school by 3 percentage points (equivalent to a rise of approximately 7.5%). In addition, we find that the enrollment rate in S&E-related courses in *tertiary* education increased by 2.5 percentage points (equivalent to a rise of approximately 11%). We do not find that the increased enrollment led to a deterioration in student achievement as measured by students' biology, physics, and chemistry grades. This suggests that supply side policies that make S&E-related courses more attractive are capable of increasing the number of S&E students, while keeping the quality of the supply of S&E students constant.

5.1 Introduction

Shortages of science and engineering (S&E) workers are seen as a threat to society's welfare as these workers are important for research and development (R&D), which is generally considered to be an important driver of economic growth (Freeman, 2006; Goolsbee, 1998). In many countries the government intervenes in the labour market for S&E workers by subsidizing R&D activities. The economic rationale for government intervention is that, when left to the market, private firms will underinvest in R&D because they cannot fully appropriate the returns on their investments. Hence, government interventions that increase the R&D activities of private firms can raise domestic welfare.

There are two main ways by which a government can try to increase the number of S&E workers: by supply side policies or by demand side policies. Supply side policies focus on increased enrollment and graduation in S&E studies. This could result in more S&E graduates, which might lead to more R&D-activity. Typical supply-side instruments are financial incentives (such as lower tuition fees) or projects aimed at increasing interest in technology or promoting the graduation rate of S&E students. Demand side policies focus on the demand for R&D by private firms. Typical instruments are R&D subsidies that directly increase R&D spending, which may induce a higher supply of S&E workers.

Although both types of policies are widely used, little is known about their effectiveness. There is some evidence that demand side policies (for example increased spending on R&D subsidies) do not lead to more S&E workers but only increase wages because workers supply labor inelastically (Goolsbee, 1998). With respect to supply side policies, even less seems to be known. We know of no convincing evidence of the impact of such policies on the supply of S&E-graduates.

This chapter looks at an intervention that was specifically targeted at increasing the enrollment rate in S&E-related courses. In 2004 the Dutch Inspectorate of Education raised concerns about the poor state of school facilities for biology, chemistry, and physics in many Dutch secondary schools (Annual Report 2003/2004). They argued that this would discourage students from pursuing a science and engineering career. More specifically, it would negatively affect a student's choice for S&E related courses in secondary education and subsequent courses in higher education. These concerns were raised in a broader context in which shortages of Dutch engineers were considered to be a serious threat for future welfare.

To address this problem, the Minister of Education launched a subsidy in which schools could apply for extra funds to improve their school facilities (*Regeling Bètalokalen*). This arrangement was implemented in 2007. The assignment of the subsidy was based on a priority score related to the school's ambition level for improving student achievement. Schools with levels above a certain cutoff value were assigned the subsidy; schools with levels lower than this value were not. Because the cutoff score was (and remained) unknown to the schools, the assignment procedure gives an opportunity to estimate the causal impact of the subsidy in a regression discontinuity framework. Using the ambition level as 'running variable', assignment to treatment status can be considered 'as good as random.'

The features of the Dutch education system allow us to estimate the impact of the subsidy on enrollment rates in S&E-related courses and academic outcomes such as grades. After the third grade in secondary school, students (aged 15) have to choose between 4 subjects ('profiles') for the continuation of their school career. These subjects differ in the courses students are obliged to follow. Students who choose 'nature' profiles have to follow courses in biology, chemistry, and physics. Students who choose 'society' profiles typically follow courses in economics, literature, and history. At the end of secondary education students are tested in each course by nationwide exams and, after graduation, they choose which course to follow in tertiary education. By using data on grades and subject choice of multiple cohorts of students from 2008-2012, we estimate the short and long-term impact of the subsidy on three important outcomes: (1) student's choice with respect to their profiles in secondary education, (2) students' biology, chemistry, and physics grades and (3) students' study choice in tertiary education. By focusing on students' profile choices and their chosen subject in tertiary education we investigate whether the additional investment in school facilities yield more S&E students. By focusing on students' biology, chemistry, and physics grades, we are able to investigate whether increased enrollment in S&E related studies implies a deterioration of the quality of the supply of S&E students. Typically biology, chemistry, and physics are taught in labs, which is not the case for other courses such as languages and economics. If better school facilities attract new students of lower ability, we might expect lower achievement levels in these S&E related courses.

We find that the subsidy raised the fraction of students that choose nature profiles by 3 percentage points (a rise of approximately 7%). Four years after the intervention this translates into an almost equivalent rise in the enrollment of students in S&E related courses in tertiary education. We do not find that the additional investment in school facilities led to lower biology, physics or chemistry grades. This suggests that supply side policies that make S&E-related courses more attractive are capable of increasing the number of S&E students, while keeping the quality of the new supply of S&E students constant.

The remainder of this chapter proceeds as follows. In Section 5.2 we discuss the related literature. In Section 5.3 we give additional information on the Dutch secondary school system. In Section 5.4 we discuss the features of the subsidy program and give additional information on the school's proposed projects. In Section 5.5 we discuss the assignment procedure and our empirical strategy. Section 5.6 shows the estimation results. Finally, Section 5.7 offers some concluding comments.

5.2 Related literature

To our knowledge, no previous studies have investigated the relationship between better school facilities and the supply of S&E graduates. There is, however, a large body of research that investigates the relationship between the quality of the learning environment and student achievement. Although the majority of this literature focuses on associations (see Hanushek, 1997; Earthman, 2002; and Mendell and Heath, 2004, for reviews), there are some studies that aim to estimate the causal impact of a better learning environment on student achievement. Obtaining credible estimates of this impact is difficult because exogenous variation in the quality of school facilities is rare. Investment in these facilities is typically endogenous as schools are free to choose how much to invest and often align this amount to the needs of their students. Because students sort into schools based on these investments, students in schools with high or low quality facilities are not likely to be comparable. Unobserved characteristics may differ between the students of these schools, which would confound the causal relationship between the quality of school facilities and student outcomes. In this section, we therefore focus on the literature that addresses these identification issues.

A first body of literature is related to additional spending on schools in general (see Guryan, 2000; Papke, 2005; Card and Payne, 2002; Chay et al., 2005; Leuven et al., 2007). In these studies exogenous variation in expenditure is produced by funding schemes or by regulation (and changes therein) through which some schools obtain more resources than others. These studies find mixed results of the impact of extra resources for schools on student outputs. Because schools are free to choose what to do with the extra resources, this literature finds it difficult to disentangle the mechanisms that led to the observed student outcomes. Higher achievement, for instance, could be due to extra spending on personnel, school facilities, or on a combination of both.

A second strand of literature is related to additional investment in information and communication technology (ICT) that aims to promote the use of computers in the classroom.

The exogenous variation used in these studies is often produced by the assignment procedure of subsidy programs that are specifically earmarked for the purchase of computers (see Angrist and Lavy, 2002; Goolsbee and Guryan, 2006; Leuven et al., 2007; Machin et al., 2007). Except for the study of Machin, the studies find no effect of increased ICT-spending (or use of computers) on student achievement.

One study that focuses on school facilities in general is Cellini et al. (2010). In this study the broad economic impact of investment in school facilities is estimated by using house prices as main outcome. The authors exploit exogenous variation in referenda outcomes for bond proposals in California which determined whether school districts could issue bonds for the improvement of their school facilities. They use a dynamic regression discontinuity framework in which they compare school districts where a bond proposal passed by one vote with school districts where a proposal failed by the same margin. They find that house prices rise in districts where the proposal passed. Moreover, they find that the willingness to pay for houses in these districts is larger than the cost of extra capital spending. In addition, they find positive effects on student achievement. Their results indicate that investment in school facilities matters and that Californian school districts underinvest in these facilities. As the authors note, the advantage of working with house prices is that they capture more than academic benefits alone: non-academic benefits such as students' health and the aesthetic appeal for well featured school facilities are captured too. When they relate house prices to student achievement, they find that the increased student achievement explains only a small part of the increased house prices, suggesting that the benefits of better school facilities are typically non-academic.

5.3 Dutch secondary education

The Dutch secondary education system is publicly funded[60] and is a highly differentiated system. After 8 years of primary school, students (aged 12) are tracked into one of the three tracks of the secondary school system. These tracks can be considered to have successive levels of education. The lowest track, VMBO, offers students a basic vocational program at four different levels. Finishing this track takes 4 years. The middle track (HAVO) prepares students for higher level vocational training (HBO) and takes 5 years. Finishing the highest

60 Since 1995 the system is lump sum financed. That is, schools receive a fixed amount of money based on the number of students in the school.

track (VWO) gives access to university and takes 6 years. Our study focuses on the HAVO and VWO tracks.

After completing the third year in secondary school, students (aged 15) in HAVO and VWO tracks have to choose between four profiles which prepare them for education in specific fields of tertiary education (economics, literature, health, engineering etc.). For HAVO students this means spending an additional two years in high school doing courses in their chosen profile. For VWO students this means an additional three years. The profiles students can choose from are (1) culture & society, (2) economics & society, (3) nature & health, and (4) nature & engineering. Switching profiles during these years is difficult and uncommon. In figure 5.1 we sketch the Dutch education system.

Each profile contains a number of compulsory courses. Specifically, and most importantly for this study, is that HAVO and VWO students in nature-profiles (3) and (4) have to do courses in physics, chemistry, and biology. Typically, these profiles prepare students for S&E related courses in higher education. For example, VWO-students who graduate from high school with profile (4) have access to mechanical engineering at an institute of technology at university.

Graduation from high school is based on passing all (or a number of) compulsory courses in the chosen profile.[61] Passing a course is determined by the average of the grades of two exams: a school exam (*Schoolexamen, SE*) and a nationwide exam (*Centraal Schriftelijk Examen, CSE*). Exams are taken for each subject (biology, chemistry etc.) and level (HAVO, VWO). Dutch grades range from 1 (very bad) to 10 (excellent), with an average grade of 5.5 being sufficient to pass a course. The content of the school exam is determined by the school and thus not comparable across schools. The content of the nationwide exam is centrally determined by the CITO-agency, which has been commissioned by the Ministry of Education to develop exams. This exam is taken in May or June, near the end of the school year.[62] All students are obliged to participate in these exams. The grades of these exams are comparable across schools because the exams are the same within each subject and level. We will use the grades of this exam to investigate the effects of the policy on student achievement in physics, chemistry, and biology.

61 The rules with respect to graduation changed over the period under investigation. This doesn't matter for our this chapter as we compare the outcomes of treated and non-treated schools in a given year.

62 The Dutch school year runs from August 1 of a given year to July 31 next year.

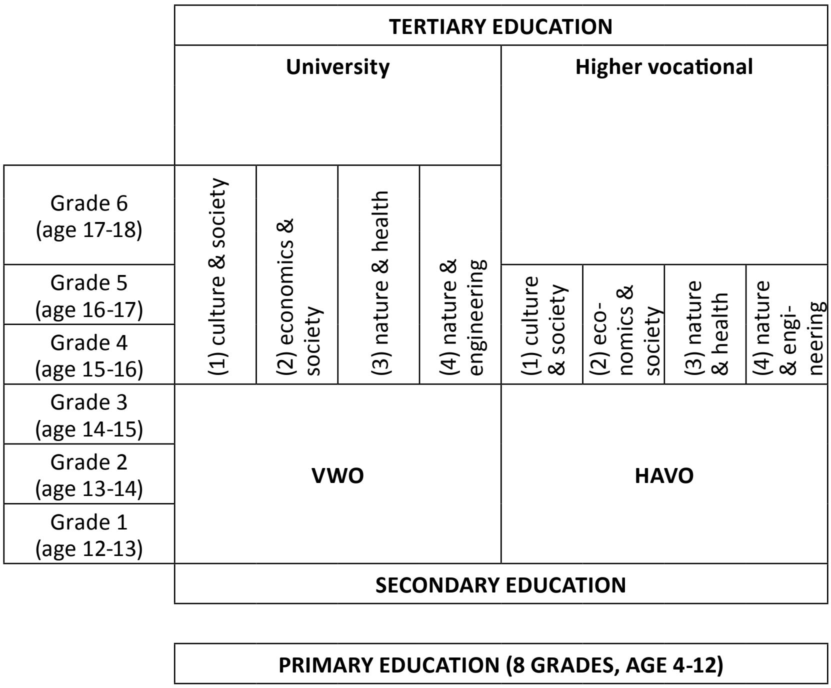


Figure 5.1: Sketch of Dutch secondary education system (excluding VMBO)

5.4 The subsidy program

The policy

In 2004 the Dutch Inspectorate of Education raised concerns about the poor state of the school facilities for biology, chemistry, and physics in many Dutch secondary schools (Annual Report, 2003/2004). This would negatively affect the choice of HAVO and VWO students for pursuing an engineering career. More specifically, it would negatively affect the choice for nature profiles in secondary education and subsequent decisions to follow a higher engineering education. These concerns were raised in a broader context in which shortages of Dutch engineers were seen a serious threat to future welfare.

To address this problem, the Minister of Education launched a subsidy arrangement in which schools could apply for extra funds to improve their school facilities (*Regeling Bètalokalen*). The arrangement was introduced in 2007. Between March 31, 2007 and May 15, 2007 all Dutch secondary schools could apply. The maximum amount of money that could be requested was €150,000, and schools had to co-finance their project for at least 50% of the total budget. This meant that if a project would cost €250,000 in total, the schools had to pay at least €125,000 themselves, and could only receive a subsidy up to €125,000. If a project would cost €450,000, then they had to pay at least €300,000 themselves.

In total 272 schools applied. From these 272 schools, 150 schools were assigned the subsidy (a total of €19.3 million). In the next section we describe the assignment procedure, which is important for our empirical strategy. First, we give some background information on the schools' proposed projects.

Background information on proposed projects for treated and non-treated schools

Table 5.I presents additional information on the schools' proposed projects. We do this both for schools that received the subsidy and for schools that did not. Throughout the chapter we refer to them as the treated and non-treated schools, respectively. The information is derived from a policy report (*Dialogic, 2007*).[63] It was based on surveys among subsamples of the treated (n=126) and non-treated schools (n=88).

In panel A we give descriptive statistics about the amount of money requested, and the amount co-financed by the schools. The treated schools requested approximately €130,000, the non-treated schools about €107,000. The table shows that getting no extra subsidy did not restrain the non-treated schools from additional spending on technology labs; 30% decided to fully carry out their project regardless of the fact their application was rejected, whilst 44% decided to carry out their project partially. This means that 74% of the non-treated schools decided to invest in new school facilities, even though they did not receive the extra subsidy. However, their spending was much less than originally planned as the amount requested exceeded the amount invested by almost a half (€107,238 versus €54,392). More importantly, the difference between the average amount invested in treated and non-treated schools was about €500,000, meaning that the investment of treated schools were about 10 times larger than that of non-treated schools. In addition, the treated schools invested about 3 times more than the extra subsidy they received (€418,581 versus €128,533). Apparently, receiving the subsidy triggered schools to invest more in their labs than they originally planned as the amount invested is much more than the co-finance requirement of 50%.

63 http://www.dialogic.nl/documents/2007.051-1013.pdf

Table 5.I: Background information on proposed projects from survey; treated schools (n=126) and non-treated schools (n=88)

	Treated schools	Non-treated schools
A. Finance		
Average amount requested for proposed project (subsidy)*	€ 128,533	€ 107,238
% of schools that fully implemented their proposed project	100%	30%
% of non-treated schools that implemented their proposed projects partially	-	44%
Average amount invested by school for proposed project	€ 547,114	€ 54,392
Average amount (co-) financed by school for proposed project	€ 418,581	€ 54,392
B. How was money spent by treated schools?		
Adjusting or restructuring old technology labs	41%	-
New equipment and materials for technology labs, i.e. oxygen cabinets, furniture, etc.	23%	-
Building new technology labs	15%	-
New ICT-equipment, i.e. computers, laptops, beamers, smart boards, calculators etc.	15%	-
Infrastructural adjustments with respect to ICT, i.e. new internet connections, fiber connections etc.	6%	-
C. Timing of implementation		
% schools that finished implementation of proposed project in 2007	17%	-
% schools that finished implementation of proposed project in 2008	70%	-
% schools that finished implementation of proposed project in 2009	91%	-
% schools that finished implementation of proposed project in 2010	100%	-
N	126	88

* The averages in this row are based on the total sample which consist of 150 treated and 122 non-treated schools

Panel B shows information about how the money was spent by the treated schools. In the survey the schools were asked how they distributed the money over 5 items. The percentages had to add up to 100%. 79% of the investment was related to technology labs: that is, adjusting or restructuring old labs (41%), building new labs (15%), and buying new equipment such as oxygen cabinets (23%). 21% of the investment was ICT-related: 15% went to new ICT-equipment and 6% to infrastructural adjustments. Hence, most of the money was invested in school labs.

Panel C gives information on the timing of the implementation of the projects.[64] In 2007, 17% of the schools had finished their planned projects. This percentage was 70% in 2008. In 2010 all schools that received subsidy had finished their projects.

64 In the survey the schools were asked about their expectation to finish their projects

To sum up, the subsidy led to an additional investment of €500,000 in facilities for schools that received the subsidy. Most of these schools did not finish the implementation of their project until 2008, meaning that most schools were not treated until this year.

5.5 Assignment procedure & empirical strategy

Assignment procedure for receiving subsidy

The assignment procedure for receiving the extra subsidy is crucial for our empirical strategy. The Ministry of Education commissioned the agency *Platform Beta Techniek* to carry out the procedure. All 272 applications were reviewed by a commission that was specifically appointed for the selection. The selection criteria for eligibility were based on three measures: (1) a score related to student achievement and enrollment in nature profiles, (2) a score related to the quality of the school facilities, and (3) a score related to the school's level of ambition for improving enrollment and student achievement in nature profiles.[65] With these scores the subsidy was assigned in two steps. First, based on the first two measures, four types of schools (i.e. quadrants) were determined: i) high achievement, bad facilities (93 schools); ii) high achievement, good facilities (74 schools); iii) low achievement, bad facilities (65 schools); and iv) low student achievement, good facilities (40 schools). All schools in the first quadrant were assigned the extra subsidy. All applications of the schools in the fourth quadrant were rejected.

In the second step, the schools in the second and third quadrant were pooled and the ambition level (3) was used to rank these schools. The higher the ambition level, the higher the school emerged in the ranking. Schools ranked 94-150[66] were assigned the extra subsidy (57 schools). Schools ranked 151-232 were not assigned the subsidy (82 schools). Hence, a cutoff was used for assignment, which was 150. The corresponding ambition level cutoff score was 3.853, and remained unknown to the schools. This threshold was used because the money was limited and the commission determined that a maximum of 150 schools could receive the subsidy. In figure 5.2 we sketch the assignment procedure.

65 The scores were grades that were given by the commission and varied between 1 and 5. A higher score means a higher achievement, quality or ambition level. The scores do not have a clear meaning by themselves.

66 All 93 schools in the first quadrant of which the application was accepted were ranked 1-93. All 40 schools in the fourth quadrant of which the application was rejected were ranked 233-272.

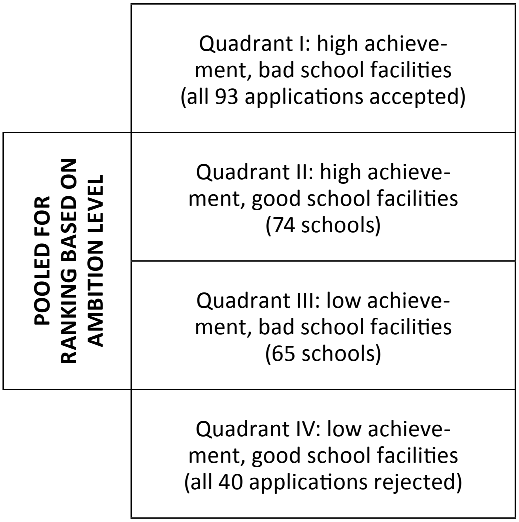


Figure 5.2: Sketch of assignment procedure

In table 5.II we present the number of schools that received the subsidy by quadrant. There were more schools eligible for the subsidy in the second quadrant (high student achievement) than in the third quadrant (low student achievement) because the commission's score with respect to the ambition level is highly correlated with student achievement (correlation=0.74). For our empirical strategy we focus on the schools in quadrants ii) and iii) because we can exploit the ranking of these schools in a regression discontinuity design. The schools in the first and fourth quadrant cannot be used in such a framework because these schools were not ranked according to ambition level (or any other variable).

Table 5.II: Number of schools that received subsidy by quadrant

	Quadrant:				
Received subsidy:	**I**	**II**	**III**	**IV**	**Total**
yes	93	51	6	0	150
no	0	23	59	40	122
Total	93	74	65	40	272

Empirical strategy

We estimate the causal effect of the subsidy by exploiting the assignment procedure. Because the schools in quadrants ii) and iii) are ranked with respect to ambition, this variable can be used as the 'running variable' in a regression discontinuity framework (Lee and Lemieux, 2010). The basic assumptions in this framework are that schools around the cutoff are very similar and that the relationship between the running variable (ambition) and the outcome is smooth around the discontinuity. For each outcome we estimate the following equation:

$$(5.1) \qquad Y_{st} = \beta_0 + \beta_1 Subsidy_s + f\left(Z_s\right) + Subsidy_s * f\left(Z_s\right) + \beta_3{}' \boldsymbol{X}_s + \varphi_t + \varepsilon_s$$

In this equation Y_{st} is the outcome of school *s* in year *t* and $Subsidy_s$ is a dummy variable that equals 1 if school *s* received the subsidy and 0 if not. Z_s is the ambition level with its cutoff level rescaled to 0. That is,

$$Subsidy_s = \begin{cases} 1 \;\; if\ Z_s = Ambition_s - Ambition_{cutoff} = Ambition_s - 3.853 \geq 0 \\ 0 \;\; if\ Z_s = Ambition_s - Ambition_{cutoff} = Ambition_s - 3.853 < 0 \end{cases}$$

Hence, schools with an ambition level higher than the cutoff value (3.853) were assigned the subsidy and schools with a level lower than this value were not. We include a polynomial of Z_s, $f(Z_s)$, in the specification and allow this polynomial to be different either side of the cutoff by including interactions of $Subsidy_s$ with this polynomial. For choosing the appropriate order of the polynomial in the RD-design, we used the Akaike Information Criterion. In all analyses, this statistic indicated that a linear term in ambition was flexible enough for the RD-design.[67] In (5.1) $\boldsymbol{X}_s$ is a vector of baseline covariates from the (pre-intervention) year 2007. We control for the commission's score for each school's student achievement, the quality of the labs, and for the amount of subsidy requested. In some specifications we will also include the outcomes from the baseline year 2007 if available. Furthermore, we also include year dummies, φ_t, for the years 2008-2012 and interactions of these dummies with the 2007 covariates in our models as we pool school data from multiple years. For the estimation of equation (5.1) we apply OLS and adjust standard errors for heteroskedasticity and clustering at the school level.

The parameter of interest is β_1 which represents the effect of receiving the extra subsidy on the outcome. We interpret this coefficient as the effect of better school facilities with a value of approximately €500,000 euro (in 2007 prices, based on table 5.I). Because we do not have information about the actual investments for each school we cannot apply a two-stage-least-squares approach in which we estimate the effect of an additional euro

67 In all analyses the statistic is smallest when using a linear term. Moreover, the estimated coefficient of this term is often not significantly different from zero.

spent on new school facilities. Hence, estimates of β_1 should be considered as intent-to-treat (i.e. reduced form) estimates.

Tests on the validity of the RD-design

The key assumption of the RD-design is that the covariates either side of the cutoff are locally balanced. This means that there should be no differences between treated and non-treated schools around the cutoff. To investigate this, we perform two tests. First, we present a density test in figure 5.3. We show two frequency histograms of the running variable Z (rescaled ambition level) to investigate whether there are discontinuous changes around the cutoff (McCrary, 2008). If schools knew the cutoff level, then they could have manipulated the running variable in order to obtain treatment status (that is, receiving the subsidy). In that case there might be more schools that just received the subsidy than schools that didn't, which could indicate that there are differences between treated and non-treated schools around the cutoff. However, this seems unlikely because the schools did not know the cutoff. This is confirmed by the two histograms: they show that there are no discontinuities/ irregularities (for example spikes) in the number of schools around 0.

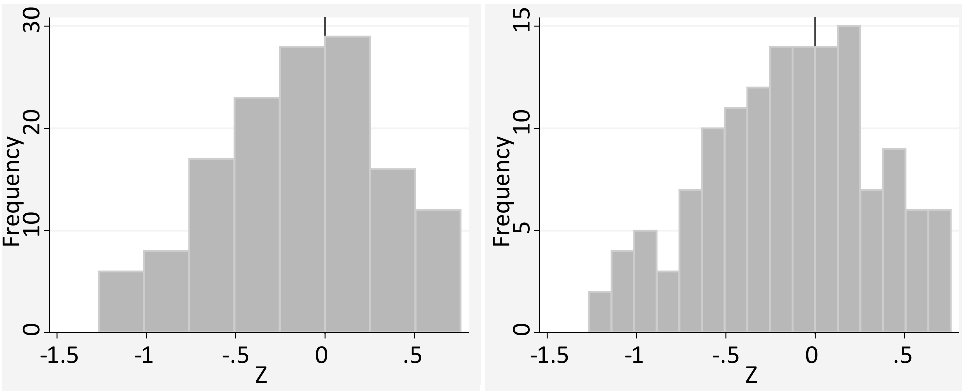


Figure 5.3: Density test: frequency histograms of running variable; 5 bins (left) and 10 bins (right) to the left of the cutoff

Second, we present a balancing test. In table 5.III we investigate whether treated and non-treated schools significantly differ around the cutoff in the amount of subsidy requested, the commission's scores related to the quality of the school's facilities and student achievement, and the fraction of students with nature profiles.[68] Because these scores are derived

68 See next section for descriptive statistics of these variables.

from the pre-intervention year 2007 we expect these differences to become insignificant when the running variable is included. In column (1) we regress the outcome on the dummy for if the school received a subsidy without controls. In column (2) we include the running variable. As expected, there are significant differences between treated and non-treated schools in student achievement and quality of the school facilities (see columns (1)). By including the running variable, these differences become much smaller, and are no longer statistically significant at the 5% level (see columns (2)). Hence, we do not find strong evidence for discontinuities around the cutoff in our baseline covariates, nor in the density of the running variable, suggesting that the key assumption of the RD-design holds. However, it should be noted that the estimates for student achievement and the subsidy requested are marginally significant at the 10% level, meaning that treated schools around the cutoff had higher student achievement and requested a lower subsidy than non-treated schools. In our analyses we include these covariates to control for these small differences.

Table 5.III: Balancing test, do schools differ around the cutoff in 2007-variables (pre-intervention)?

	Dependent variable (from pre-intervention year 2007):							
	quality of lab spaces		**student's achievement**		**subsidy requested**		**fraction of students with nature profiles**	
Independent variable:	**(1)**	**(2)**	**(1)**	**(2)**	**(1)**	**(2)**	**(1)**	**(2)**
Dummy=1 if received subsidy	0.453***	0.203	0.721***	0.224*	5658	-24995*	0.009	0.016
	(0.0808)	(0.138)	(0.0729)	(0.119)	(7770)	(13367)	(0.013)	(0.209)
Running variable included (ambition level)	no	yes	no	yes	no	yes	no	yes
Observations	139	139	139	139	139	139	136	136
R-squared	0.170	0.213	0.418	0.524	0.004	0.060	0.003	0.013

Notes: Each column is an OLS-regression. Columns (1) do not include controls. Columns (2) include the running variable. Outcomes from the year 2007 are used. Robust standard errors (between brackets) are used. * significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

5.6 Data

In our analysis we use school level data for years 2007-2012. We obtained all scores from the assignment procedure from the *Platform Beta Techniek*. This includes information on the ambition level, the quality of the labs, student achievement, and the amount of subsidy requested for each school. The outcomes were obtained from an online source from the Ministry of Education, which includes information on students' profile choice, grades and enrollment in tertiary education.[69] For each year, we could match about 97% of the schools in the second and third quadrant from the assignment procedure with the schools from the online source of the Ministry.[70] Hence, we created a panel dataset with information on 137 schools from 2007-2012.

For our dependent variables we focus on three outcomes: 1) profile choice in secondary education (measured by enrollment in nature profiles in the fourth grade), 2) grades in biology, physics and chemistry (measured at the end of the school year), and 3) enrolment in S&E related courses in tertiary education (measured one year after graduation). Hence, the dependent variables have been chosen such that we can follow cohorts of students over a part of their school career (profile choice -> high school graduation -> enrollment tertiary education). We use the information from multiple years to study the long term impact of the intervention. Descriptive statistics of these variables are listed in table 5.IV for our main estimation sample. They are shown for treated (T) and non-treated (NT) schools for each year. The number of observations differs between years because some schools do not appear in the central database of the ministry in all years. However, as mentioned, the percentage matched is never below 97%. Also, to adjust for the difference in the size of schools, the school's averages and fractions have been weighted by the number of students used in the calculation of these statistics.[71]

First, we look at student choices with respect to their profile in grade 4 (first row of table 5.IV). We may expect that better school facilities affect this choice after finishing grade 3. There is a 1 percentage point difference in the fraction of students with nature profiles between treated and non-treated schools in 2007 (42% versus 41%). Thereafter, this difference increases to 4 percentage points in 2012 (48% versus 44%).

Second, we look at students' (unstandardized) biology, physics, and chemistry grades in the final year of secondary school (rows 2-7). They are given for both HAVO and VWO

69 http://www.duo.nl/organisatie/open_onderwijsdata/databestanden/vo/default.asp

70 Some schools were not administered in the online source from the Ministry of Education in all years.

71 We do this by using the [aweight] option in Stata. If one school has 10 students with average biology grade of 6 and another school 30 students with average biology grade of 8, then the weighted average will be (10*6+30*8)/40=7.5 (which is the correct average). Not using the aweight option will give an incorrect average grade of (6+8)/2=7.

Table 5.IV: Descriptive statistics main estimation sample

Row	Variables	2007 (pre-intervention) T	2007 (pre-intervention) NT	2008 T	2008 NT	2009 T	2009 NT	2010 T	2010 NT	2011 T	2011 NT	2012 T	2012 NT
		Main estimation sample (schools in quadrants II and III)											
	Dependent variables:												
	A. Fraction of students that choose nature profiles in secondary education												
1	% of students with nature-profile in fourth year	42%	41%	45%	42%	46%	42%	46%	42%	45%	43%	48%	44%
	N	*56*	*80*	*56*	*80*	*56*	*81*	*56*	*81*	*55*	*81*	*55*	*80*
	B. Unstandardized grades												
2	Average grade biology HAVO	-	-	-	-	6.4	6.4	6.4	6.5	6.3	6.3	6.2	6.2
3	Average grade physics HAVO	-	-	-	-	6.4	6.4	6.2	6.2	6.2	6.2	6.7	6.7
4	Average grade chemistry HAVO	-	-	-	-	6.1	6.2	6.2	6.2	6.3	6.4	6.5	6.6
	N	-	-	-	-	*54*	*76*	*54*	*76*	*54*	*77*	*53*	*77*
5	Average grade biology VWO	-	-	-	-	6.5	6.5	6.5	6.5	6.5	6.5	6.4	6.5
6	Average grade physics VWO	-	-	-	-	6.3	6.4	6.3	6.4	6.4	6.4	6.4	6.5
7	Average grade chemistry VWO	-	-	-	-	6.3	6.4	6.4	6.5	6.4	6.4	6.6	6.7
	N	-	-	-	-	*56*	*81*	*56*	*80*	*56*	*80*	*55*	*82*
	C. Fraction of students that moves on to S&E related studies in tertiary education												
8	% of students enrolled in S&E related studies one year after graduation	-	-	-	-	21%	21%	21%	21%	22%	21%	25%	23%
	N	-	-	-	-	*56*	*81*	*56*	*81*	*56*	*82*	*55*	*82*
	2007-Covariates												
9	Commission's score related to the school's ambition level (the running variable)	4,2	3,4	-	-	-	-	-	-	-	-	-	-
10	Commission's score related to the school's student achievement	3,6	2,9	-	-	-	-	-	-	-	-	-	-
11	Commission's score related to the school's quality of space labs	3,1	2,6	-	-	-	-	-	-	-	-	-	-
12	Amount of subsidy requested	€ 115.705	€ 110.047										
	N	*57*	*82*	-	-	-	-	-	-	-	-	-	-

students because the nationwide exams are different for these educational levels.[72] The grades have not been standardized as the standard deviation of the grades is approximately equal to 1.[73] This means that differences in grades can be interpreted as a percentage of standard deviation. The figures in rows 2-7 suggest that the increased enrollment in nature profiles did not have a negative impact on grades. There are hardly any differences in grades between treated and non-treated schools, regardless of the year. The maximum difference is only about 0.1 standard deviations.

Finally, we investigate student choices with respect to the field of study in higher education (row 8). After graduation, students make a choice in which field they want to pursue their career. If better school facilities increase the fraction of students that choose nature profiles in secondary education, we may also expect an increase in the fraction of the *entire school population in the final year of secondary school* that choose S&E related courses in tertiary education after graduating from high school.[74] It should be noted that we do not expect to see a change in this variable until 2012 because the third graders from schools that were treated in 2008 are still in secondary school in the years 2008-2011. Recall that after their profile choice (in 2008/2009) they had to spend two or three additional years in school. This means that we do not observe them in tertiary education until 2012. Hence, 2012 should be considered the first post treatment year of the intervention if we use enrollment in tertiary education as the outcome variable. This seems to be confirmed by the figures in table 5.IV: there are hardly any differences in this fraction between treated and non-treated schools up to 2011. In 2012, however, the difference is 2 percentage points (25% versus 23%).

The covariates for our analysis are shown in the bottom panel of table 5.IV (rows 9-12). Treated schools had higher ambition levels, higher student achievement, and better labs than the non-treated schools in the (pre-intervention) year 2007. They also requested more money (about €5,000 more). However, as shown in the previous section, these differences become smaller and turn statistically insignificant for the schools around the cutoff.

72 For the analysis presented in the next section, we pool the data of the HAVO and VWO students.

73 Because we do not have information on grades at the individual level, we asked CITO for the standard deviation. For each course (biology, chemistry, physics) the standard error is approximately equal to 1.

74 Hence *we do not look* at the fraction of *those with nature-profiles* that choose S&E related studies in tertiary education. This transition probability may have been unaffected.

5.7 Results

Impact of subsidy on profile choice

Table 5.V presents estimates of the effect of the subsidy on the fraction of students with nature profiles in the fourth grade based on equation (5.1). In all columns we adjust standard errors for heteroskedasticity and clustering at the school level. The first three columns give pooled estimates of this effect when we pool data from 2008-2012. Column (1) only includes the running variable and year dummies for 2008-2012. This estimate is the basic RD-estimate of the causal effect of the subsidy on the outcome because, conditional on the level of ambition, schools around the cutoff should be similar in their pre-intervention characteristics. In column (2) we add the other 2007-covariates. For our RD-approach to be robust, including these pre-intervention controls should not alter the estimates much compared to column (1). In column (3), we also include interactions of the year dummies with all our covariates, such that we allow the coefficients for the covariates to be different for each year. This is our most flexible specification.

The estimated effects in the three columns range between 2.5 and 3.5 percentage points and are significantly different from zero. Including the 2007 controls in column (2) decreases the estimated effect from 3.5 to 2.5 but improves precision as the standard error is smaller than in column (1). Allowing the specification to be different for each year in column (3) does not change the estimated effect compared to column (2). Hence, the results are robust to the inclusion of additional controls and the type of specification used. If we take 3 percentage points as the average estimated effect and 41% as the mean of the control group (based on table 5.IV), then the subsidy raised the fraction that choose nature profiles by approximately 3/41=7.3%.

In columns (4)-(6) we investigate whether short term effects are different from long term effects by allowing the estimated effect to be different for each year. We do this by including interactions of the dummy for whether the school received the subsidy with the year dummies. The rest of the specification used in these three columns is the same as in the first three columns.

The estimated effects differ somewhat between years, but an F-test (shown at the bottom of the table) does not reject the hypothesis that the impact is the same across all years. The p-values range between 0.17 and 0.38, all exceeding the 5% significance level. Hence, we cannot conclude that the short-term impact of better school facilities is different from the long-term impact. The results for 2012, which was 4 years after intervention,[75] suggest that the impact is persistent because the estimated effects are in the order of 3 percentage points. The improved school facilities seem to have structurally increased the enrollment rate in nature-profiles.

One possible concern is that the estimates are driven by selection of students from other schools into the treatment schools. For example, students with strong preferences for well appointed school facilities and nature profiles might change schools if they see that the facilities have been improved in their neighboring schools. However, this seems unlikely as the treatment (i.e. the improvement of the school facilities) affects students in the third grade when they have already spent three years in their school. Hence, the effects we find are likely to be driven by the school's own students who, because of the improvement in the school's facilities, are more likely to enroll in nature profiles.

Impact of subsidy on grades

Table 5.VI presents estimates of the impact of the subsidy on students' (unstandardized) biology, physics, and chemistry grades. In all columns we adjust standard errors for hetero-skedasticity and clustering at the school level. Columns (1) to (3) give pooled estimates of this impact for these three courses, respectively. The specification used in these columns is the same as in column (3) in table 5.V, except that it also includes a dummy for educational level (VWO=1; HAVO=0) and its interaction with the years because the data for HAVO and VWO students have been pooled over all years.[76] The estimates can be interpreted as percentage of standard deviations as the standard error of the grades is about 1.

The estimation results show that, although the enrollment rate in nature profiles has increased, which may indicate that students of lower ability may have selected themselves in these profiles, student achievement did not deteriorate. The estimated effects are not significantly different from zero.

In columns (4)-(6) we investigate whether the short-term impact differs from the long term impact, an analogue to column (6) in table 5.V. Again, the estimated effects differ somewhat, but the F-test does not reject the hypothesis that the impact is the same in all years. The p-values range between 0.66 and 0.83, all clearly exceeding the 10% significance level. Hence, there seems to be no negative short- or long-term impact of the subsidy on student achievement.

75 Using 2008 as the year in which the schools were treated.
76 This also explains the larger number of observations in this table compared to table 5.V.

Table 5.V: Impact of subsidy on fraction of students with nature profiles in fourth grade

	Dependent variable: fraction of students with nature profiles in fourth grade					
	Pooled estimates of impact subsidy			**Estimates of impact subsidy by year**		
Independent variables:	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Dummy=1 if received subsidy (pooled estimate)	0.0356**	0.0250**	0.0247*			
	(0.0169)	(0.0123)	(0.0126)			
Dummy=1 if year=2008 * Dummy=1 if received subsidy				0.0320*	0.0213	0.0289**
				(0.0183)	(0.0145)	(0.0146)
Dummy=1 if year=2009 * Dummy=1 if received subsidy				0.0431**	0.0320**	0.0321*
				(0.0180)	(0.0143)	(0.0191)
Dummy=1 if year=2010 * Dummy=1 if received subsidy				0.0388**	0.0284**	0.0280*
				(0.0188)	(0.0141)	(0.0155)
Dummy=1 if year=2011 * Dummy=1 if received subsidy				0.0221	0.0113	0.00209
				(0.0179)	(0.0139)	(0.0197)
Dummy=1 if year=2012 * Dummy=1 if received subsidy				0.0426**	0.0326**	0.0326*
				(0.0182)	(0.0140)	(0.0168)
p-value F-test: impact is same over years 2008-2012				0.17	0.17	0.38
Controls:						
Running variable (ambition 2007)	yes	yes	yes	yes	yes	yes
Additional 2007 controls	no	yes	yes	no	yes	yes
Year dummies * all 2007 controls (full interacted model)	no	no	yes	no	no	yes
Observations	681	681	681	681	681	681
R-squared	0.085	0.426	0.438	0.088	0.429	0.440

Notes: Each column is an OLS-regression. All models include year dummies. Columns (1) and (4) include a linear term in the running variable that differs at either side of the cutoff. Columns (2) and (5) add the other 2007-variables: the commission's score related to student achievement, the score related to the quality of the lab spaces, the amount of subsidy requested and the fraction of students with nature profiles in fourth grade in 2007. Columns (3) and (6) additionally include interactions of the covariates with the year dummies. Standard errors (between brackets) are clustered at the school level. Number of observations (schools) per year is 136, 137, 137, 136 and 135 for years 2008-2012 respectively. * significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

Table 5.VI: Impact of subsidy on fraction of students that moves on to S&E related fields in tertiary education one year after graduation

	Dependent variable: grades					
	Pooled estimates of impact subsidy			Estimates of impact subsidy by year		
	biology	physics	chemistry	biology	physics	chemistry
Independent variables:	(1)	(2)	(3)	(4)	(5)	(6)
Dummy=1 if received subsidy (pooled estimate)	0.0720	-0.0630	-0.111			
	(0.0561)	(0.0913)	(0.0900)			
Dummy=1 if year=2009 * Dummy=1 if received subsidy				0.0466	0.00436	-0.0488
				(0.0732)	(0.119)	(0.109)
Dummy=1 if year=2010 * Dummy=1 if received subsidy				0.119	-0.0977	-0.0964
				(0.0763)	(0.126)	(0.131)
Dummy=1 if year=2011 * Dummy=1 if received subsidy				0.0684	-0.0708	-0.143
				(0.0704)	(0.0962)	(0.108)
Dummy=1 if year=2012 * Dummy=1 if received subsidy				0.0528	-0.0825	-0.145
				(0.0597)	(0.0936)	(0.0867)
p-value F-test: impact is same over years 2009-2012				0.66	0.74	0.83
Controls:						
Running variable (ambition 2007)	yes	yes	yes	yes	yes	yes
Additional 2007 controls	yes	yes	yes	yes	yes	yes
Year dummies * all 2007 controls (full interacted model)	yes	yes	yes	yes	yes	yes
Observations	1067	1067	1067	1067	1067	1067
R-squared	0.160	0.097	0.133	0.160	0.098	0.134

Notes: Each column is an OLS-regression. In all columns the fully interacted model is used as in columns (3) and (6) in table V. The models also include interactions of the dummy for educational level (1=VWO; 0=HAVO) and its interaction with the year dummies. Standard errors (between brackets) are clustered at the school level. For HAVO the number of observations (schools) per year is 130, 130, 131, 130 for years 2008-2011 respectively. For VWO it is 137, 136, 136 and 137. These numbers add up to 1067. *significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

Impact of subsidy on study choice in tertiary education.

Table 5.VII presents estimates of the impact of the subsidy on the fraction of students that were enrolled in S&E related studies one year after graduation. The table is structured in the same way as table 5.V. The first three columns give the pooled estimates. Columns (4)-(6) investigate differences between the short- and long-term impact of the intervention.

As mentioned in the previous section, we do not expect to see differences between treated and non-treated schools until 2011, because the third graders that were affected by the new school facilities in 2008 or 2009 (when all school were treated) had to spend an additional two or three years in secondary school before they chose a course in tertiary education. This means that 2012 is the first post treatment year of the intervention if we use enrollment in tertiary education as outcome variable. This is also reflected in the results in table 5.VII. The pooled estimates in columns (1)-(3) are not significantly different from zero because it pools data from 2009-2012, which only consist of one post treatment year (2012). In addition, the estimates are also not statistically different from zero if we look at the estimates by year in columns (4)-(6) for years 2009-2011. However, in 2012, estimates are in the order of 2.5 percentage points and significantly different from zero. The F-test rejects the hypothesis that the impact is the same across all years at the 10% level (only in columns (4) and (5)). This estimate suggests that better school facilities resulted in a higher enrollment in S&E related courses four years after intervention. If we take 2.5 percentage points as the average estimated effect and 23% as the mean of the control group (based on table 5.IV), then the subsidy raised the fraction that choose nature profiles by approximately 2.5/23=11%.

5.8 Conclusions

In this chapter we evaluated the effects of a subsidy program targeted at improving school facilities for science and engineering (S&E) related courses in secondary schools. The subsidy triggered schools to invest more in these facilities than they would originally invest. Typical, additional investments were about €500,000, lending support to an interpretation of our estimates as better school facilities with a value equal to this amount. For the estimation of the effects we exploited the fact that schools were ranked according to a priority score and that a cutoff level was used to determine which schools could receive the subsidy.

Table 5.VII: Impact of subsidy on fraction of students that moves on to S&E related fields in tertiary education one year after graduation

	Dependent variable: fraction of students that enrolls in S&E-related studies in tertiary education one year after graduation					
	Pooled estimates of impact subsidy			Estimates of impact subsidy by year		
Independent variables:	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Dummy=1 if received subsidy (pooled estimate)	0.0131	0.0123	0.0119			
	(0.00950)	(0.00952)	(0.00955)			
Dummy=1 if year=2009 * Dummy=1 if received subsidy				0.00492	0.00405	0.00726
				(0.0118)	(0.0118)	(0.0154)
Dummy=1 if year=2010 * Dummy=1 if received subsidy				0.0125	0.0117	-0.000125
				(0.0108)	(0.0108)	(0.0111)
Dummy=1 if year=2011 * Dummy=1 if received subsidy				0.00852	0.00762	0.00614
				(0.0115)	(0.0114)	(0.0136)
Dummy=1 if year=2012 * Dummy=1 if received subsidy				0.0271**	0.0262**	0.0346**
				(0.0113)	(0.0114)	(0.0157)
p-value F-test: impact is same over years 2009-2012				0.10	0.10	0.13
Controls:						
Running variable (ambition 2007)	yes	yes	yes	yes	yes	yes
Additional 2007 controls	no	yes	yes	no	yes	yes
Year dummies * all 2007 controls (full interacted model)	no	no	yes	no	no	yes
Observations	549	549	549	549	549	549
R-squared	0.049	0.053	0.071	0.057	0.060	0.078

Notes: Each column is an OLS-regression. All models include year dummies. Columns (1) and (4) include a linear term in the running variable that differs at either side of the cutoff. Columns (2) and (5) add the other 2007-variables: the commission's score related to student achievement, the score related to the quality of the lab spaces and the amount of subsidy requested. Columns (3) and (6) additionally include interactions of the covariates with the year dummies. Standard errors (between brackets) are clustered at the school level. Number of observations (schools) per year is 136, 137, 137, 136 and 135 for years 2009-2012 respectively. *significant at the 10% level, ** significant at the 5% level,*** significant at the 1% level.

We find that better school facilities caused a significant increase in the fraction of students that choose these S&E-related courses. The estimated effect is about 3 percentage points, which is equivalent to a rise of about 7%. In addition, we find that, after graduation, enrollment in S&E related courses in tertiary education increased by 2.5 percentage points, which is equivalent to a rise of approximately 11%. We do not find that the additional investment in school facilities negatively affected student achievement as measured by students' biology, physics, and chemistry grades. This suggests that supply side policies that make S&E related courses more attractive are capable of increasing the number of S&E students without impacting the quality of the supply of these students. However, it remains unclear to what extent these policies increase the number of S&E *workers* (and hence R&D activity) because S&E *students* might choose other jobs than R&D after graduation when these jobs are much better paid. This is a topic for future research.

Summary

This thesis looks at four different input factors that may contribute to the production of human capital. It investigates the importance of teacher quality, time in school, compulsory education, and school facilities for student outcomes, in particular student achievement. For estimating the effects of educational inputs education policies are used that were implemented in such a way that they lent themselves for an evaluation within a quasi-experimental setting. Within such a setting causal estimates of the parameters of interest can be obtained. The chapters in this thesis are summarized below.

Chapter 2 examines the effect of teacher quality on student achievement using a novel identification strategy that exploits data on twins who entered the same school but were allocated to different classrooms in an exogenous way. The assignment of twins to different classrooms can be viewed as a natural experiment that exposes very similar individuals to different schooling conditions. This quasi-experiment allows investigation of the causal effect of classroom quality on student outcomes using observational data. The variation in classroom conditions to which the twins are exposed can be considered as exogenous if the assignment of twins to different classes is as good as random. This assumption seems quite plausible within the institutional context of this study, which is Dutch primary education. In many Dutch schools twins are assigned to different classes due to an informal policy rule that dictates that twins are not allowed to attend the same class. At school entry neither schools nor parents have much information about the ability or behavior of twins and the ability or behavior of their class mates. Moreover, because in early childhood twins are more similar than different, it seems unlikely that small differences between twins will affect the way they are assigned to different classes. By using this identification strategy, it is shown that that teacher quality is important, and that it is partly measured by teacher experience. It is found that (a) the test performance of all students improve with teacher experience; (b) teacher experience also matters for student performance after the initial years in the profession; (c) the teacher experience effect is most prominent in earlier grades; (d) the teacher experience effects are robust to the inclusion of other classroom quality measures, such as peer group composition and class size; and (e) an increase in teacher experience also matter for career stages with less labor market mobility which suggests positive returns to on the job training of teachers.

Chapter 3 uses school entry rules to provide the first estimates of the causal effect of time in school on cognitive skills for many countries around the world, for multiple age

groups and for multiple subjects. These estimates enable a comparison of the performance of education systems based on gain scores instead of level scores (i.e. average test scores). Data from international cognitive tests are used and variation induced by school entry rules is exploited within a regression discontinuity framework. The effect of time in school on cognitive skills differs strongly between countries. Remarkably, there is no association between the level of test scores and the estimated gains in cognitive skills. As such, a country's ranking in international cognitive tests might misguide its educational policy. Across countries we find that a year of school time increases performance in cognitive tests with 0.2 to 0.3 standard deviations for 9-year-olds and with 0.1 to 0.2 standard deviations for 13-year-olds. Estimation of gains in cognitive skills also yields new opportunities for investigating the determinants of international differences in educational achievement.

Chapter 4 evaluates the effects of raising the minimum school leaving age from 14 to 15 in the Netherlands in 1971. The policy goal was to increase the number of high school graduates. The analysis shows that the change led to a decrease in the high school dropout rate of approximately 20%. However, there were no benefits in terms of employment or higher wages. Several explanations for this finding have been explored. Suggestive evidence is presented in support of the skill-based explanation that no more labor-market relevant skills were learned during this extra year of school compared to those skills previously learned out of school.

Chapter 5 evaluates the effects of a subsidy program targeted at improving facilities for biology, physics and chemistry in secondary schools. The goal of this policy was to increase the enrollment rate in science and engineering (S&E) related courses at secondary and subsequent education institutions. The subsidy was assigned to schools based on a priority score reflecting the ambition level of schools to improve student achievement. Schools with scores below a threshold value did not receive subsidy. The assignment procedure is exploited in a regression discontinuity framework to estimate the impact of the subsidy on student outcomes. It is found that the subsidy increased the enrollment rate in S&E-related courses in secondary school by 3 percentage points (equivalent to a rise of approximately 7.5%). In addition, it is found that the enrollment rate in S&E-related courses in *tertiary* education increased by 2.5 percentage points (equivalent to a rise of approximately 11%). The increased enrollment did not lead to a deterioration in student achievement as measured by students' biology, physics, and chemistry grades. This suggests that supply side policies that make S&E-related courses more attractive are capable of increasing the number of S&E students, while keeping the quality of the supply of S&E students constant.

Nederlandse samenvatting

'Menselijk kapitaal is het geheel van competenties, kennis, sociale en persoonlijke vaardigheden die economisch waardevol zijn'[77]

Menselijk kapitaal is van groot belang voor onze welvaart (zie bijv. Hanushek en Woessman, 2008). Economisch bloei zou niet mogelijk zijn, en we zouden bovendien minder gelukkig zijn als we niet in het bezit waren van een basishoeveelheid aan menselijk kapitaal. Gelukkig kunnen we zelf zorgen voor meer menselijk kapitaal door er in te investeren. Een manier om dat te doen is via het onderwijs. We gaan naar school om vaardigheden te leren die nodig zijn voor de productie van complexe goederen en diensten. Daarnaast leren we er vaardigheden die van belang zijn voor ons geluk later in het leven, zoals sociale vaardigheden.

Dit doen we overigens niet alleen voor onszelf, maar ook voor anderen. We profiteren mogelijk van het onderwijs dat anderen hebben gehad. We kunnen bijvoorbeeld van elkaar leren, en we lopen minder kans om slachtoffer van een misdrijf te worden als anderen hoger opgeleid zijn. Dit wordt vaak aangeduid als positieve spillovers of 'externe' effecten.

Echter, mensen kunnen te weinig investeren in hun menselijk kapitaal. Zij kunnen onvoldoende financiële middelen hebben voor onderwijs, kortzichtig zijn of simpelweg niet de waarde of het belang van onderwijs inzien. Dit schaadt niet alleen henzelf, maar ook de samenleving als geheel. Externe effecten, zoals hierboven beschreven, komen namelijk mogelijk niet tot stand. Daarnaast kan het leiden tot een grotere onderwijs- en inkomensongelijkheid, als sommigen te weinig investeren in hun menselijk kapitaal maar anderen niet. Om deze redenen bemoeit de overheid zich met de keuzes van mensen met betrekking tot scholing. Anders gezegd: de overheid grijpt in op de markt voor onderwijs. Zij doet dit door het maken van wetgeving (bijv. Leerplicht) en door het subsidiëren van het onderwijs. Maar hoe grijpt de overheid nu op een effectieve manier in? Dat is een van de belangrijkste vragen die onderwijseconomen bestuderen.

Bij het onderzoeken van deze vraag kijken economen naar de zogenaamde productiefunctie van menselijk kapitaal. Deze functie beschrijft hoe menselijk kapitaal tot stand komt. Hoeveel menselijk kapitaal iemand bezit wordt gerelateerd aan factoren *buiten* school (zoals zijn genen, familieomgeving en netwerk) en aan factoren *binnen* school (zoals

77 http://www.ensie.nl/definitie/Menselijk_kapitaal

de tijd die is doorgebracht op school, de klassengrootte en de kwaliteit van de docenten). Voor deze onderwijsfactoren (zoals klassengrootte) is het van belang om te weten hoeveel zij bijdragen aan het menselijk kapitaal. Een beter begrip hiervan kan beleidsmakers namelijk helpen bij de besluitvorming over onderwijsmaatregelen. Leidt een klassenverkleining bijvoorbeeld tot betere leerprestaties? En zo ja, hoeveel beter dan? En leidt een verlenging van de leerplicht tot het verwerven van meer vaardigheden? Voor een beleidsmaker die kan kiezen uit een veelvoud aan beleidsopties, is het belangrijk de antwoorden op dit type vragen te weten, wil ze uiteindelijk tot beter beleid komen. Slecht beleid leidt immers tot verspilling van belastinggeld en mogelijk tot onderinvesteringen in menselijk kapitaal. Kennis over dit soort onderwijsvraagstukken is dus zeer belangrijk.

In dit proefschrift worden vier verschillende onderwijsfactoren onderzocht die mogelijk bijdragen aan de productie van menselijk kapitaal. Het proefschrift onderzoekt:

1) hoe belangrijk de kwaliteit van de klas is voor de leerprestaties van leerlingen (en welke rol werkervaring van leerkrachten hierin speelt),
2) hoeveel er in een extra jaar onderwijstijd geleerd wordt op school,
3) wat de effecten zijn van een verlenging van de Leerplicht, en
4) hoe belangrijk schoolfaciliteiten zijn voor leerprestaties en keuzegedrag van studenten.

De belangrijkste uitdaging in dit proefschrift is het identificeren van het causale, d.w.z. oorzakelijke, effect van elk van deze inputfactoren (werkervaring, onderwijstijd, leerplicht, schoolfaciliteiten). Dit is moeilijk omdat de productiefunctie van menselijk kapitaal complex in elkaar zit. Factoren buiten school hangen namelijk vaak samen met factoren binnen school en kunnen allemaal van invloed zijn op de hoeveelheid menselijk kapitaal. Stel dat bijvoorbeeld scholen die extra lestijd geven aan hun leerlingen beter presteren dan andere scholen, maar tegelijkertijd ook vaker leergierige leerlingen hebben. Dan kunnen deze betere prestaties niet zomaar worden toegeschreven aan de extra lestijd, omdat het net zo goed kan worden toegeschreven aan het feit dat de studenten gemotiveerder waren om te leren. Bovendien is het mogelijk dat leerlingen niet alleen verschillen in leergierigheid, maar ook nog eens in veel andere factoren. Deze factoren kunnen, net als leergierigheid, moeilijk te observeren zijn. Dit betekent dat simpele vergelijkingen van de leerprestaties van de leerlingen tussen scholen met en zonder extra lestijd niet het causale effect van extra lestijd zal meten, aangezien de leerlingen allesbehalve vergelijkbaar zijn. Om het causale effect van extra lestijd namelijk te achterhalen, zouden andere factoren, zoals de verschillen tussen de leerlingen, moeten worden uitgesloten. Dit voorbeeld geeft dus aan dat het een probleem kan zijn om een zuivere, geldige vergelijking te maken tussen leerlingen als men geïnteresseerd is in het effect van een inputfactor zoals extra lestijd.

Dit probleem komt overigens vaak voor in evaluaties van inputfactoren van de productiefunctie van menselijk kapitaal, en wordt ook wel het ‘endogeniteitsprobleem’ genoemd. In de economische literatuur is het de afgelopen twee decennia de uitdaging geweest oplossingen voor het endogeniteitsprobleem te vinden (Angrist en Pischke, 2010). Deze oplossingen worden gevonden in het gebruik van gerandomiseerde experimenten of quasi-experimenten waarin studenten ‘exogeen’ (d.w.z. door willekeur) zijn toegewezen aan een controle- of behandelgroep. De studenten in een behandelgroep krijgen dan te maken met de onderwijsmaatregel (bijv. extra lestijd), terwijl studenten in de controlegroep daar niet mee te maken krijgen. Bij een gerandomiseerd experiment wordt de toewijzing dan door de onderzoeker zelf gedaan via een loterij. In dat geval worden studenten door toeval in een behandel- of controlegroep geplaatst. Bij een quasi-experiment gaat de toewijzing vaak via onderwijsbeleid. Het beleid is dan zo ingericht dat er door toeval behandel- en controlegroepen worden gecreëerd. De ‘exogene’ (d.w.z. willekeurige) manier van toewijzen zorgt er dan voor dat beide groepen vergelijkbaar zijn. Dit wil zeggen dat het mogelijk wordt om de behandelgroep op een geldige manier te vergelijken met de controlegroep. De literatuur die gebruik maakt van deze (quasi-)experimenten wordt vaak aangeduid als de experimentele literatuur.

Dit proefschrift levert een bijdrage aan deze literatuur. Er wordt onderwijsbeleid gebruikt om te onderzoeken wat het effect is van: 1) de werkervaring van leerkrachten, 2) een jaar extra onderwijstijd, 3) een verlenging van de leerplicht en 4) de kwaliteit van de schoolfaciliteiten. De effecten worden gemeten op allerlei studentenuitkomsten die een benadering zijn van menselijk kapitaal, zoals de leerprestaties van de studenten. De hoofdstukken en onderwerpen in dit proefschrift kunnen afzonderlijk van elkaar worden gelezen en worden hieronder samengevat.

Hoe belangrijk de kwaliteit van de klas is voor de leerprestaties van leerlingen wordt besproken in hoofdstuk 2. Dit onderzoek maakt gebruik van een nieuwe strategie om causale effecten te kunnen achterhalen; er worden gegevens benut van tweelingparen die naar dezelfde school gaan, maar aan verschillende klassen zijn toegewezen. De toewijzing aan de verschillende klassen is het gevolg van onderwijsbeleid: op veel Nederlandse basisscholen worden tweelingen in verschillende klassen geplaatst door een (vaak informele) richtlijn die bepaalt dat tweelingen niet naar dezelfde klas mogen. De toewijzing van tweelingen aan de verschillende klassen kan worden gezien als een quasi-experiment, waarbij twee zeer vergelijkbare individuen worden blootgesteld aan verschillende klassen. Dit quasi-experiment maakt onderzoek mogelijk naar het causale effect van de klassenkwaliteit op de leerprestaties. De toewijzing aan verschillende klassen kan namelijk als willekeurig worden verondersteld. Bij aanvang van het schooljaar in het eerste leerjaar (groep 1) hebben ouders en leerkrachten namelijk nog weinig informatie over de leercapaciteit van de tweeling en die van hun klasgenoten op basis waarvan ze tweelingen naar verschillende klassen zouden kunnen sturen. Daarnaast lijkt het onwaarschijnlijk dat kleine verschillen tussen tweelingen de manier zal beïnvloeden waarop ze aan verschillende klassen worden toegewezen. Met behulp van deze identificatiestrategie wordt aangetoond dat verschillen in kwaliteit tussen klassen feitelijk neerkomt op kwaliteitsverschillen tussen leerkrachten, en dat deze kwaliteit ertoe doet. Het blijkt dat de kwaliteit van leerkrachten deels kan worden afgemeten aan de ervaring van de leerkrachten: tweelingen die zijn toegewezen aan klassen met de meer ervaren leerkrachten doen het beter in termen van leerprestaties dan hun tweelingbroers/zussen die aan de klassen zijn toegewezen met de minder ervaren leerkrachten.

In hoofdstuk 3 wordt onderzocht wat het causale effect is van een jaar onderwijstijd op de leerprestaties van leerlingen in het basis- en voortgezet onderwijs.[78] Met behulp van schoolinstroom regels[79] wordt dit voor verschillende landen gedaan: voor elk land wordt bepaald wat het effect is. Dit effect kan worden gelabeld als een 'toegevoegde waarde' effect: hoeveel voegt een onderwijssysteem van een land toe aan de leerprestaties van de leerlingen in een jaar onderwijs. Het kan worden beschouwd als een kwaliteitsmaatstaf van het onderwijssysteem: hoe meer een systeem toevoegt, hoe beter het systeem. De resultaten laten zien dat het effect van een jaar onderwijstijd op de leerprestaties sterk verschilt per land. Deze 'toegevoegde waarde' effecten worden vervolgens gerankt: van het land met de hoogste toegevoegde waarde naar het land met de laagste toegevoegde waarde. Hierdoor ontstaat een landen ranking. Deze ranking wordt vervolgens vergeleken met een andere, traditionele ranking, namelijk een die gebaseerd is op de gemiddelde leerprestaties van een land. De gemiddelde leerprestaties van een land kan ook worden beschouwd als een indicator van de kwaliteit van het onderwijssysteem: hoe hoger de

78 Er worden twee groepen leerlingen onderscheiden: 9-jarigen (groep 5/6 in het basisonderwijs) en 13-jarigen (klas 1/2 in het voortgezet onderwijs)

79 Ook dit is een vorm van onderwijsbeleid. Er wordt hier niet verder ingegaan op de techniek. Schoolinstroom regels zijn regels die betrekking hebben op het beginnen met school. In Nederland is er lange tijd de 1-oktober regeling geweest: kinderen geboren voor 1 oktober mochten in het huidige schooljaar beginnen, terwijl kinderen geboren na 1 oktober nog een jaar moesten wachten. De regels zorgen ervoor dat kinderen die bijna even oud zijn (geboren rond 1 oktober) verschillen in de tijd doorgebracht op school.

gemiddelde score, hoe beter het systeem. Nu is de vraag of de twee rankingen, namelijk de toegevoegde waarde ranking en de ranking gebaseerd op het gemiddelde, hetzelfde beeld opleveren. Is het zo dat een land dat het gemiddeld goed doet ook veel toevoegt in een jaar onderwijs? En is het zo dat een land dat het gemiddeld slecht doet weinig toevoegt? Uit de analyse blijkt dat dit niet het geval is. Er zijn landen die het gemiddeld goed doen, maar weinig toevoegen, en andersom (landen die het gemiddeld slecht doen maar veel toevoegen). Anders gezegd: er wordt geen verband gevonden tussen de traditionele ranking, welke gebaseerd is op het gemiddelde, en de nieuwe ranking, welke gebaseerd is op de toegevoegde waarde. Een verklaring hiervoor is dat in tegenstelling tot het toegevoegde waarde effect, een hoog gemiddelde ook door andere factoren dan het onderwijssysteem kan worden veroorzaakt. Bijvoorbeeld als de condities buiten school gunstig zijn voor het opdoen van vaardigheden. Dit betekent dat de traditionele onderwijsrankings die vaak de kranten halen (zoals TIMMS en PISA), en welke gebaseerd zijn op het gemiddelde, mogelijk misleidend kunnen zijn als kwaliteitsmaatstaf voor het onderwijssysteem. Immers, de onderwijssystemen van de landen die in een dergelijke ranking bovenaan staan hoeven niet per se veel toe te voegen aan de leerprestaties van hun leerlingen.

Wat is het effect van een verhoging van de leerplicht in Nederland? Dit effect wordt besproken in hoofdstuk 4. In 1971 werd de leerplichtige leeftijd verhoogd van 14 naar 15 jaar. Deze wetswijziging had als doel om meer leerlingen de mogelijkheid te geven een middelbaar schooldiploma te halen. De analyse in hoofdstuk 4 toont aan dat deze wijziging inderdaad geleid heeft tot een daling van het aantal voortijdig schoolverlaters van ongeveer 20% (d.w.z. er zijn meer diploma's uitgereikt). Het heeft er echter niet toe geleid dat de leerlingen meer zijn gaan verdienen of betere kansen op de arbeidsmarkt hebben gekregen. In dit hoofdstuk worden verschillende verklaringen voor deze bevinding onderzocht. Er wordt suggestief bewijs gegeven voor een 'skill-based' verklaring: de leerlingen die een extra jaar op school hebben gezeten, hebben waarschijnlijk geen extra vaardigheden geleerd die relevant waren voor de arbeidsmarkt. De analyse laat dus zien dat het uitbreiden van de leerplicht tot een hoger opgeleide bevolking leidt (meer diploma's) maar dat het daarmee niet automatisch de productiviteit (d.w.z. het inkomen) van de bevolking verhoogt.

In het laatste hoofdstuk wordt geëvalueerd wat de effecten zijn van een subsidieprogramma gericht op het verbeteren van de schoolfaciliteiten voor biologie, natuurkunde en scheikunde in het voortgezet onderwijs (zoals laboratoria). Dit beleid was bedoeld om de instroom in bètaopleidingen in het voortgezet en hoger onderwijs te verhogen. De subsidie werd toegewezen aan een bepaalde selectie van scholen. Alleen ambitieuze scholen kwamen in aanmerking voor deze subsidie. De ambitie van scholen om de leerprestaties van leerlingen te verbeteren werd bepaald door een score. Scholen met hoge ambitie scores kwamen in aanmerking voor de subsidie, terwijl scholen met iets minder hoge scores daar niet voor in aanmerking kwamen. Er zijn dus scholen die net wel of net niet aan een 'drempelwaarde' (bepaalde ambitie score) voldeden. Deze toewijzingsprocedure kan worden gebruikt in een zogeheten 'regressie discontinuïteit analyse' om het effect van betere schoolfaciliteiten op de instroom in bètaopleidingen en de leerprestaties te bepalen. In deze analyse worden leerlingen van scholen met scores vlak onder en boven de drempelwaarde met elkaar vergeleken. De analyse laat zien dat de subsidie het instroompercentage in bètaopleidingen in het *voortgezet* onderwijs heeft verhoogd met 3 procentpunten (equivalent aan een stijging van ongeveer 7,5%). Daarnaast laat de analyse zien dat de subsidie het instroompercentage in bètaopleidingen in het *hoger* onderwijs met ongeveer 2,5 procentpunt heeft verhoogd (gelijk aan een stijging van circa 11%). De hogere instroom heeft daarentegen niet geleid tot een daling van de leerprestaties, zoals afgemeten aan de cijfers voor het centraal schriftelijk eindexamen voor biologie, natuurkunde en scheikunde. Dit suggereert dat het type subsidie zoals boven besproken, het aantal bètastudenten kan vergroten, zonder daarbij de kwaliteit van het aanbod aan bètastudenten te laten verschralen.

Curriculum Vitae

Sander Gerritsen was born in 1983 in Amsterdam. He studied econometrics at the University of Amsterdam from 2001 to 2008. He obtained his bachelor degree in 2006 and his master degree in 2008. His master thesis was about the effect of cannabis use on educational attainment. He has been working as a researcher for the CPB Netherlands Bureau for Economic Policy Analysis since 2009. In 2011 he started writing his PhD thesis.

Bibliography

Aaronson, D., Barrow, L., Sander, W. (2007). 'Teachers and student achievement in Chicago public high schools,' *Journal of Labor Economics*, 25, 95-135.

Almond, D., Currie, J. (2010). 'Human capital development before age 5,' in: Eric A. Hanushek, Stephen Machin and Ludger Woessmann, *Handbook of Labor Economics*, Volume 4b.

Angrist, J.D., Krueger, A.B. (1991). 'Does Compulsory Schooling Attendance Affect Schooling and Earnings,' *Quarterly Journal of Economics* 106, 979-1014.

Angrist, J.D., Lavy, V. (2002). 'New Evidence on Classroom Computers and Pupil Learning,' *Economic Journal*, 112:482, 735-765.

Angrist, J.D., Pischke, J-S. (2010). 'The Credibility Revolution in Empirical Economics: How Better Research Designs is Taking the Con out of Econometrics,' *Journal of Economic Perspectives*, Volume 24, Number 2, 3-30.

Ashenfelter, O., Krueger, A. (1994). 'Estimating the returns to schooling using a new sample of twins,' *American Economic Review*, 84: 1157–1173.

Bacolod, M.P. (2007). 'Do Alternative Opportunities Matter? The Role of Female Labor Markets in the Decline of Teacher Quality,' *Review of Economics and Statistics*, 89: 737–751.

Baltes, P.B., Reinert, G. (1969). 'Cohort Effects in cognitive development as revealed by cross-sectional sequences,' *Development Psychology*, 1(2): 169-77.

Bedard, K., Dhuey, E. (2006). 'The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects,' *Quarterly Journal of Economics*, 121(4): 1437–1472.

Behrman, J.R., Rosenzweig, M.R. (2002). 'Does Increasing Women's Schooling Raise the Schooling of the Next Generation?' *American Economic Review*, 92 (1): 323-334.

Berlinski, S., Galiani, S., Gertler, P. (2009). 'The effect of pre-primary education on primary school performance,' *Journal of Public Economics*, 93, 219–234.

Berlinski, S., Galiani, S., McEwan, P.J. (2011). 'Preschool and maternal labor supply: Evidence from a regression discontinuity design,' *Economic Development and Cultural Change*, 59(2): 313-344.

Bhuller, M., Mogstad, M., Salvanes, K.G. (2011). 'Life-Cycle Bias and the Returns to Schooling in Current and Lifetime Earnings,' NHH Dept. of Economics Discussion Paper No. 4/2011.

Bishop, J.H. (1997). 'The effect of national standards and curriculum-based examinations on achievement,' *American Economic Review*, 87 (2), 260-264.

Black, S., Devereux, P.J., Salvanes, K.G. (2011). 'Too Young to Leave the Nest? The Effects of School Starting Age,' *Review of Economics and Statistics*, 93(2): 455–467.

Boardman, A.E., Murnane R. (1979). 'Using panel data to improve estimates of the determinants of educational achievement,' *Sociology of Education*, 52, 113-121.

Bound, J., Jaeger, D.A. (2000). 'Do compulsory school attendance laws alone explain the association between quarter of birth and earnings?' *Research in Labor Economics*, 19, 83-108.

Breakspear, S. (2012). 'The Policy Impact of PISA: An Exploration of the Normative Effects of International Benchmarking in School System Performance,' *OECD Education Working Papers*, No. 71, OECD Publishing. http://dx.doi.org/10.1787/5k9fdfqffr28-en.

Buckles, K., Hungerman, D. (2012). 'Season of birth and later outcomes: old questions, new answers,' *Review of Economics and Statistics*, forthcoming.

Cahan, S., Cohen, N. (1989). 'Age versus schooling effects on intelligence development,' *Child Development*, 60, 1239–1249.

Cahan, S., Davis, D. (1987). ‘A between-grade-levels approach to the investigation of the absolute effects of schooling on achievement,’ *American Educational Research Journal*, 24(1):1-12.

Card, D. Payne, A. A. (2002). ‘School Finance Reform, the Distribution of School Spending, and the Distribution of Student Test Scores,’ *Journal of Public Economics*, 83, 49-82.

Cascio, E. U., Lewis, E. G. (2006). ‘Schooling and the armed forces qualifying test: Evidence from school entry laws,’ *Journal of Human Resources*, 41(2), 294–318.

Cellini, S.R., Ferreira, F., Rothstein, J. (2010). ‘The Value of School Facility Investments: Evidence from a Dynamic Regression Discontinuity Design,’ *Quarterly Journal of Economics*, 125, 215-261.

Chay, K.Y., McEwan P.J., Urquiola, M. (2002). ‘The Central Role of Noise in Evaluating Interventions that Use Test Scores to Rank Schools,’ *Journal of Public Economics*, 83, 49-82.

Chetty, R., Friedman, J.N., Hilger, N., Saez, E., Whitmore Schanzenbach, D., Yagan, D. (2011). ‘How does you kindergarten classroom affect your earnings? Evidence from project STAR,’ *Quarterly Journal of Economics*, 126 (4): 1593-1660.

Clotfelter, C.T., Ladd, H. F., Vigdor, J.L. (2006). ‘Teacher-Student Matching and the Assessment of Teacher Effectiveness,’ *Journal of Human Resources*, 41(4): 778-820.

Corcoran, S. P., Evans, W.N., Schwab, R.M. (2004). ‘Changing Labor-market Opportunities for Women and the Quality of Teachers,’ 1957–2000, *American Economic Review*, 94 (2004), 230–235.

Currie, J. (2001). ‘Early Childhood Education Programs,’ *Journal of Economic Perspectives*, 15(2): 213-38.

Dee, T. S. (2004). ‘Teachers, Race and Student Achievement in a Randomized Experiment,’ *The Review of Economics and Statistics*, 86 (1): 195-210.

Devereux, P.J., Hart, A.H. (2010). ‘Forced to be Rich? Returns to Compulsory Schooling in Britain,’ *The Economic Journal* 120 (December), 1345-1364.

Dickert-Conlin, S., Chandra, A. (1999). ‘Taxes and the Timing of Births,’ *Journal of Political Economy*, 107(1): 161–177.

Dickert-Conlin, S., Elder, T. (2010). ‘Suburban Legend: School Cutoff Dates and the Timing of Births,’ *Economics of Education Review*, 29(5): 826–841.

Dobkin, C., Ferreira, F. (2010). ‘Do school entry laws affect educational attainment and labor market outcomes?’ *Economics of Education Review*, 29(1): 40-54.

Driessen, G., Van Langen, A., Vierke, H. (2004). Basisrapportage PRIMA-cohortonderzoek, Vijfde meting 2002-2003 (Report on PRIMA-longitudinal research project, Survey 2002-2003). Nijmegen.

Earthman, G. I. (2002). ‘School Facility Conditions and Student Academic Achievement,’ UCLA’s Institute for Democracy, Education, and Access (IDEA) Paper No. wws-rr008-1002.

Feng, Li. (2009). ‘Opportunity wages, classroom characteristics, and teacher mobility,’ *Southern Economic Journal*, 75, 1165-1190.

Figlio, D., Loeb, S. (2011). ‘School accountability, In Eric A. Hanushek, Stephen Machin and Ludger Woessmann (Eds.), *Handbook of the Economics of Education, Vol. 3, Amsterdam: North Holland*, pp. 383-421.

Fredriksson, P., Öckert, V. (2013). ‘Life-cycle Effects of Age at School Start,” *Economic Journal*, forthcoming.

Freeman, R.B. (2005). ‘Does Globalization of the Scientific/Engineering Workforce Threaten U.S. Economic Leadership,’ NBER working paper no. 11457.

Geluk, A., Hol, J. (2001). Samen of apart? Meerlingkinderen naar school, peuterspeelzaal of kinderopvang [Together or Apart? Multiples to school, playground or daycare; Brochure]. Bergen, the Netherlands: NVOM (Dutch Society for Parents of Multiples).

Goolsbee, A. (1998). 'Does Government R&D Policy Mainly Benefit Scientists and Engineers?,' *American Economic Review*, Vol. 88, No. 2, Papers and Proceedings of the Hundred and Tenth Annual Meeting of the American Economic Association.

Goolsbee, A., Guryan, J. (2006). 'The Impact of Internet Subsidies in Public Schools,' *Review of Economics and Statistics*, 88, 336-347.

Gormley, W. T., Gayer, T. (2005). 'Promoting school readiness in Oklahoma: An evaluation of Tulsa's pre-K program,' *Journal of Human Resources*, 60, 533–558.

Grenet, J. (2013). 'Is Extending Compulsory Schooling Alone Enough to Raise Earnings? Evidence from French and British Compulsory Schooling Laws,' *Scandinavian Journal of Economics* 115(1), 176-210.

Guarino, C., Reckase, M.D., Wooldridge, J.M. (2012). 'Can Value-Added Measures of Teacher Performance Be Trusted?' IZA Discussion Papers 6602, Institute for the Study of Labor (IZA).

Guryan, J. (2002). 'Does Money Matter? Regression-Discontinuity Estimates from Education Finance Reform in Massachusetts,' NBER working paper no. 8269.

Hansen, K., Heckman, J., Mullen, K. (2004). 'The effect of schooling and ability on achievement test scores,' *Journal of Econometrics*, 121(1–2), 39–98.

Hanushek, E.A. (1971). 'Teacher characteristics and gains in student achievement: Estimation using micro data,' *American Economic Review*, 60 (2), 280-288.

Hanushek, E.A. (1992). 'The Trade-Off Between Child Quantity and Quality,' *Journal of Political Economy*, 100, 84–117.

Hanushek, E.A. (1997). 'Assessing the effects of school resources on student performance: an update.' *Educational Evaluation and Policy Analysis* 19(2): 141-164.

Hanushek, E.A. (2011). 'The Economic Value of Higher Teacher Quality,' *Economics of Education Review*, 30(2): 466-479.

Hanushek, E.A., Kain J.F., O'Brien, D.M., Rivkin, S.G. (2005). 'The market for teacher quality,' NBER Working Paper 11154.

Hanushek, E.A., Rivkin, S.G. (2006). 'Teacher quality.' In: Hanushek, E., Welch, F. (Eds.), *Handbook of Economics of Education, vol 2*. Elsevier.

Hanushek, E.A., Woessmann, L., (2008). 'The role of cognitive skills in economic development,' *Journal of Economic Literature*, 46 (3), 607–668.

Hanushek, E.A., Woessmann, L., (2011). 'The economics of international differences in educational achievement,' In Eric A. Hanushek, Stephen Machin and Ludger Woessmann (Eds.), *Handbook of the Economics of Education, Vol. 3, Amsterdam: North Holland*, 2011, pp. 89-200.

Harris, D.N., Sass, T.R., (2011). 'Teacher training, teacher quality and student achievement,' *Journal of Public Economics*, 95, 798-812.

Hoxby, C. M., Leigh, A. (2004). 'Pulled Away or Pushed Out? Explaining the Decline of Teacher Aptitude in the United States,' *American Economic Review*, 94: 236–240.

IEA, 2011, TIMSS and PIRLS—Informing Educational Policy for Improved Teaching and Learning, document.

Jacob, B., Lefgren, L, (2009). 'The Effect of Grade Retention on High School Completion,' *American Economic Journal: Applied Economics*, 1(3): 33-58.

Jürges, H, Schneider, K., Büchel, F. (2005). 'The effect of central exit examinations on student achievement: Quasi-experimental evidence from TIMSS Germany,' *Journal of the European Economic Association*, 3 (5), 1134-1155.

Kane, T., Rockoff, J.E., Staiger D.O. (2006). 'What does certification tell us about teacher effectiveness? Evidence from New York City,' Working Paper 12155.

Krueger, A.B. (1999). 'Experimental estimates of education production functions,' *The Quarterly Journal of Economics*, 114 (2): 497-532.

Lee, D., Lemieux, T. (2010). 'Regression Discontinuity Designs in Economics,' *Journal of Economic Literature* 48, 281-355.

Leuven, E., Lindahl, M., Oosterbeek H., Webbink, H.D. (2007). 'The Effect of Extra Funding for Disadvantaged Pupils on Achievement,' *Review of Economics and Statistics* 89, 721-736.

Leuven, E., Lindahl, M., Oosterbeek, H., Webbink, D. (2010). 'Expanding schooling opportunities for 4-year-olds,' *Economics of Education Review*, 29:319–328.

Li, H., Rosenzweig, M., Zhang, J. (2010). 'Altruism, Favoritism, and Guilt in the Allocation of Family Resources: Sophie's Choice in Mao's Mass Send-Down Movement,' *Journal of Political Economy*, 118(1): 1-38.

Luyten, H. (2006). 'An empirical assessment of the absolute effect of schooling: regression-discontinuity applied to TIMSS-95'. *Oxford Review of Education*, Vol. 32, No. 3, 397-429.

Machin, S., McNally, S., Silva, O. (2007). 'New Technology in Schools: Is There a Payoff,' *The Economic Journal*, 117(2007), 1145-1167.

Manacorda, M, (2012). 'The Cost of Grade Retention,' *Review of Economics and Statistics*, 94 (2): 596–606.

McCrary, J. (2008). 'Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test,' *Journal of Econometrics*, 142, 698-714.

McCrary, J., Royer, H. (2011). 'The Effect of Female Education on Fertility and Infant Health: Evidence from School Entry Policies Using Exact Date of Birth,' *American Economic Review*, 101(1): 158–195.

McEwan, P. J., Shapiro, J.S. (2008). 'The benefits of delayed primary school enrollment: discontinuity estimates using exact birth dates,' *Journal of Human Resources*, 43(1): 1–29.

Meghir, C., Palme, M. (2005). 'Educational Reform, Ability and Family Background,' *American Economic Review*, Vol. 95, No. 1, pp. 414-424.

Meghir, C., Rivkin, S. (2010). 'Econometric methods for research in education,' In Eric A. Hanushek, Stephen Machin and Ludger Woessmann (Eds.), *Handbook of the Economics of Education, Vol. 3, Amsterdam: North Holland*, 2011, pp. 89-200.

Memorie van Toelichting, Leerplichtwet 1968, Zitting 1967-9039, Nr. 3.

Mendell, M. J., Heath, G.A. (2004). 'Do Indoor Environments in Schools Influence Student Performance? A Critical Review of the Literature,' *Indoor Air*, 15, 27-52.

Montizaan, R.M., Cörvers, F., De Grip, A., Dohmen, T.J. (2013). 'Negative Reciprocity and Retrenched Pension Rights.' IZA Discussion paper No. 6955.

Mueller, S. (2013). 'Teacher experience and the class size effect – Experimental evidence,' *Journal of Public Economics*, 98, 44-52.

Neal, D.A., Johnson, W.R. (1996). 'The role of premarket factors in black-white differences,' *Journal of Political Economy*, 104, 869-895.

Nye, B., Konstantopoulos, S., Hedges, L.V. (2004). 'How large are teacher effects?' *Educational Evaluation and Policy Analysis*, 26 (3): 237-257.

OECD, Danielle Venn (2009), 'Legislation, collective bargaining and enforcement: Updating the OECD employment protection indicators,' www.oecd.org/els/workingpapers.

Oosterbeek, H., Webbink, D.H. (2007). 'Wage effects of an extra year of basic vocational education,' *Economics of Education Review* 26, 408-419.

Oreopoulos, P. (2006). 'Estimating Average and Local Average Treatment Effects of Education when Compulsory Schooling Laws Really Matter,' *American Economic Review* 96, 152-175.

Papke, L.E. (2005). 'The effects of Spending on Test Pass Rates: Evidence from Michigan,' *Journal of Public Economics*, 89, 821-839.

Pischke, J.S., von Wachter, T. (2008). 'Zero Returns to Compulsory Schooling in Germany: Evidence and Interpretation,' *Review of Economics and Statistics* 90, 592-598.

Rivkin, S.G., Hanushek, E.A., Kain, J.F. (2005). 'Teachers, schools and academic achievement,' *Econometrica*, Vol. 73, No. 2: 417-458.

Rockoff, J.E. (2004). 'The impact of individual teachers on student achievement: evidence from panel data,' *The American Economic Review*, Vol. 94, No. 3, Papers and proceedings of the one hundred sixteenth annual meeting of the American Economic Association, pp. 247-252.

Roeleveld, J., Vierke, H. (2003). 'Uitval en Instroom bij de derde meting van het PRIMA-cohortonderzoek,' Amsterdam/Nijmegen: SCO-Kohnstamm Instituut / ITS.

Rothstein, J. (2010). 'Teacher Quality in Educational Production: Tracking, Decay and Student Achievement,' *Quarterly Journal of Economics*, 125(1): 129-174.

Schwerdt, G., West, M.R. (2012). 'The Effects of Test-based Retention on Student Outcomes over Time: Regression Discontinuity Evidence from Florida,' Working Paper.

Smith, A. (1776). An Inquiry into the Nature and Causes of the Wealth of Nations.

Staiger, D.O., Rockoff, J.E. (2010). 'Searching for effective teachers with imperfect information,' *Journal of Economic Perspectives*, 24 (3): 97-117.

TIMSS & PIRLS, (2011) TIMSS and PIRLS—Informing Educational Policy for Improved Teaching and Learning, International Study Center, http://timssandpirls.bc.edu/home/pdf/TP_Impact_Statement.pdf.

Todd, P.E., Wolpin, K.I. (2003). 'On the specification and estimation of the production function for cognitive achievement,' *Economic Journal*, 113 (485), F3-33.

Von Davier, M., Gonzalez, E., Mislevy, R.J. (2009). 'What are plausible values and why are they useful?' IERI Monograph Series. Issues and Methodologies in Large-Scale Assessments, Vol. 2, 9-36.

Webbink, D. (2005). 'Causal Effects in Education,' *Journal of Economic Surveys*, Vol. 19, No. 4.

Webbink, D., Hay, D., Visscher P.M. (2007). 'Does sharing the same class in school improve cognitive abilities of twins? *Twin Research and Human Genetics*, Vol. 10, No. 4: 573-580.

West, M.R., Woessmann, L. (2010). 'Every Catholic child in a Catholic school: Historical resistance to state schooling, contemporary school competition, and student achievement,' *Economic Journal*, 120 (546), F229-255.

Wiswall, M. (2013). 'The dynamics of teacher quality,' *Journal of Public Economics*, 100, 61-78.

Woessman, L, (2003). 'Schooling resources, educational institutions and student performance: The international evidence,' *Oxford Bulletin of Economics and Statistics*, 65 (2), 117-170.

Woessmann, L., West, M. (2006). 'Class-size effects in school systems around the world: Evidence from between-grade variation in TIMSS,' *European Economic Review*, 50 (3): 695-736.