How much do children learn in school?

International evidence from school entry rules

Dinand Webbink* and Sander Gerritsen**

Abstract

This study investigates the effect of time in school on cognitive skills for many countries around the world, for multiple age groups and for multiple subjects. We use data from international cognitive tests and exploit variation induced by school entry rules within a regression discontinuity framework. How much children learn in school depends on their age and on their country. For 9-year-olds a year of school time on average increases performance in cognitive tests with 0.2 to 0.3 standard deviations, and for 13-year-olds with 0.1 to 0.2 standard deviations. The effect of time in school on cognitive skills strongly differs between countries. The estimates also enable a comparison of the performance of education systems based on gain scores instead of level scores. Remarkably, we find no association between the level of test scores and the estimated gains in cognitive skills. As such, a high ranking in international cognitive tests might hide a poor performance in the year before the test.

JEL Codes: I2, J24

Keywords: cognitive skills, time in school, educational policy, international comparison

* Erasmus School of Economics, Rotterdam, Tinbergen Institute, IZA Bonn;

** Erasmus School of Economics, Rotterdam, CPB, Netherlands Bureau for Economic Policy Analysis, The Hague. Corresponding author: webbink@ese.eur.nl;

We thank Roel van Elk, Bas ter Weel, Justin McCrary, Eric Hanushek, Coen Teulings, Thomas Buser, Hessel Oosterbeek, Erik Plug, Sacha Kapoor and seminar participants at the Tinbergen Institute, University of Amsterdam, University of Rotterdam, CPB Netherlands, conferences in Aarhus and Torino for helpful comments.

1. Introduction

Many studies have found a strong association between the economic outcomes of nations and their cognitive skills (e.g. Hanushek & 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.[1] 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 & 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 & 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 & 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 & Woessmann 2011).[2]

[1] 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).

[2] Some recent studies apply a quasi-experimental approach for investigating specific factors such as the effects of class size (Woessmann & West 2006), central exams (Jürges et al. 2005), relative age (Bedard & Dhuey 2006) or private school competition (West & Woessmann 2010) using data from international cognitive tests.

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.[3] Students born in adjacent months are assigned to different grades due to these school entry rules. As a result, students that 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.[4] We apply this framework for estimating the effect of spending one year in school for a selection of countries that participated in international cognitive tests. Moreover, we address identification issues related to sampling bias, relative age effects and violations of the exclusion restriction, which were not addressed in previous studies that applied the same approach for specific countries.

This framework enables us to perform three 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.

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.[5] 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 and

[3] 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).

[4] This approach was introduced by development psychologists for separating schooling and age effects on test scores and was recently applied in the economic literature (see Section 2).

[5] More recent TIMSS studies only sample students in one grade.

have a high compliance with these rules; 10 countries for the 9-year-olds and 13 countries for the 13-year-olds.

Our empirical results can be summarized in three main findings. First, 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, 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. Third, and most remarkable, we find no association between the estimated gains in achievement and the average 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 might yield misleading information about the performance of the education system. Using the 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.

Our paper 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. 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 also used a 'between-grades design' did not address identification issues related to sampling bias, relative age effects and violations of the exclusion restriction. In this paper we perform analyses that aim to mitigate estimation biases due to these concerns. In particular, we apply a recently developed approach for obtaining sharp lower and upper bounds for average treatment effects in the presence of sampling bias (Lee, 2009) to our 'between-grades design'. Fourth, we contribute to the literature that uses international cognitive tests for investigating the determinants of international differences in educational achievements. In this literature it has been argued that 'further exploration of quasi-experimental settings in the international data should be high on the agenda' (Hanushek & Woessman 2011). This is exactly what this paper does; we apply a quasi-experimental approach using international data.

This study is organized as follows. Section 2 reviews the previous literature on the effects of time in school. Section 3 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 4. Section 5 shows the estimates of the effect of one year of time in school for pooled samples of countries. In Section 6 differences between countries are investigated. Section 7 compares country rankings of level scores with country rankings of gains scores. Section 8 concludes.

2. Previous studies

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 & 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.[6] The effect of schooling has been analyzed from different perspectives. A first strand of the literature 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 & 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 (2006) 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.[7] 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.[8] A third strand of the literature focuses on grade retention. For instance, Jacob & 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.[9] Our study is also related to a fourth strand of the

[6] For surveys of the development psychology literature on the estimation of schooling effects, see Ceci (1991) and Stipek (2002).
[7] Early childhood interventions like Head Start or the Perry Preschool Project have been studied intensively. For surveys, see Currie (2001) and Almond & Currie (2010).
[8] For other recent studies, see Berlinsky et al. (2009) and Leuven et al. (2010).
[9] For other recent studies, see Manacorda (2012) or Schwerdt & West (2012).

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.

3. Empirical strategy

In this paper 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). 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 that 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 paper we apply this approach to samples of countries that participated in international cognitive tests. Figure 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.[10] 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.[11] In both countries we also observe that scores in math and science decline

[10] It should be noted that these grades also include 8-year-olds and 10-year-olds.
[11] The first stage estimates (Equation (1)) for Singapore and England respectively are 0.96 and 0.93.

(linearly) with age which confirms previous findings about the importance of age at entry for test performance (Bedard & 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 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.

Estimation

The typical approach in economic studies on the effect of time in school is to estimate instrumental variables models. This approach aims to take account of non-compliance with the school entry rules, for instance due to grade retention, acceleration or redshirting (holding children back for one year at the start of their school career).[12] In the first stage equation the school entry rule is used as an instrumental variable for the grade level. In the second stage equation the variation in time in grade that is induced by the school entry rule is used for estimating the causal effect of time in a specific grade on cognitive skills.

(1) $$G_i = \alpha_0 + \alpha_1 D_i + f(B_i - c) + \alpha_2 X_i + \varepsilon_i$$

$$D_i = 1[B_i < c]$$

$$f(B_i - c) = f_l(B_i - c) + D_i[f_r(B_i - c) - f_l(B_i - c)]$$

(2) $$Y_i = \beta_0 + \beta_1 \hat{G}_i + f(B_i - c + \beta_2 X_i + \eta_i$$

[12] Studies in development psychology have dealt with this problem by excluding non-compliers (e.g. Cahen & Cohen 1989). This creates, however, a non-random sample that might induce biased estimates. In the education literature, Luyten (2006) assessed the effect of time in grade for eight countries separately in a study that starts with a similar approach as our study. However, he does not address the issue of non-compliance or other important identification issues (see below).

where Y_i is the performance of student i, G_i is a dummy variable for being in the higher grade, D_i is a dummy variable for being born on the left side of the cutoff date, B_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). Estimation of β_1 will yield the causal effect of time in grade if the usual IV-assumptions with regard to the first stage, the independence assumption and the exclusion restriction hold. 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 increases by one year due to the school entry rule. The estimate of the effect of 'being born before the cutoff date' on the grade level (α_1) in the first stage equation indicates the proportion of the students of a specific country that moves to the next grade when their date of birth would change from the right side to the left side of the cutoff date. In addition, reduced form estimates will yield the causal effect of being eligible for one extra year of time in school:

(3) $$Y_i = \delta_0 + \delta_1 D_i + f(B_i - c) + \delta_2 X_i + \upsilon_i$$

In this paper we also apply this approach for obtaining reduced form and instrumental variables estimates of time in school. For the estimation we focus on discontinuity samples of student born six months around the cutoff date. Hence, we are using a local approach. Although this estimation framework has been used in several previous studies the application is not straightforward because of several identification issues. To mitigate these issues we also estimate lower and upper bounds of the average treatment effect as introduced in Lee (2009).

The first identification issue is about the plausibility of the exclusion restriction; 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 that are on track. However, if students on the left side of the cutoff have been retained they have been treated with an additional year in school. Hence, it is assumed that grade retention has no effect on cognitive skills. This assumption might not hold because recent studies report statistically significant effects of grade retention (see Section 2). The exclusion restriction would hold if the non-compliance of students would not be the result of retention or acceleration, but the result of redshirting. We address this issue in our analysis by only selecting countries in which retention or acceleration of students is not allowed or very uncommon. For these countries it seems likely that the observed non-compliance is driven by redshirting.

A related concern is that 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 & Dhuey 2006). We address this 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.[13]

The second issue for identification is the conditional independence assumption; students born and observed near the cut-off date should be 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.[14] 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 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

[13] 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.

[14] 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 2013).

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.

A further and related issue is sampling bias, which seems not to be recognized in previous studies that also used a between-grades design. 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 we do not observe students from these age cohorts that are not in these grades, for instance they might have been held back for one year, retained or accelerated. At the left side of the cutoff date we don't observe students that have been accelerated; at the right side of the cutoff date we don't observe student that have been delayed. If the sampling bias results from retention or acceleration of students we might expect a downward bias for the estimated effects because students that have been accelerated will probably have a relatively high ability, and students that have been retained will probably have a relatively low ability. However, redshirting seems to be more likely for students with wealthy backgrounds, which might induce an upward bias of the estimates. We address this concern about sampling bias in two steps. First, we only select countries that apply clear nationwide school-entry rules and that have a high compliance with these rules. In addition, within this group of countries we separately analyze countries that do not allow for retention or acceleration of students. Second, for this selective group of countries we estimate lower and upper bounds of the treatment effect. Lee (2009) has recently introduced an approach for obtaining sharp lower and upper bounds for average treatment effects in the presence of sampling bias within a setting with random assignment of treatment. In this approach sample selection is modeled within Rubin's potential outcome framework and the treatment status (D_i) determines whether the outcome Y_i is observed or not. S, S_0 and S_1 are all binary indicators of sample selection. S denotes whether an individual is observed, and S_0 and S_1 are potential sample selection indicators for the treated and control states. Hence,

$$S = S_1 D + S_0 (1 - D)$$

For example, an individual that is observed in the treatment status but has a missing outcome in the control status has $S_1 = 1$ and $S_0 = 0$. The main assumption in Lee (2009)'s approach is that the treatment assignment can only affect sample selection in one direction:

(4) $S_1 \geq S_0$ with probability 1.

This monotonicity assumption implies that we can distinguish three groups: 1) individuals that are never observed ($S_0 = 0, S_1 = 0$), 2) individuals that are always observed ($S_0 = 1, S_1 = 1$) and 3) individuals that are only observed in the treatment status ($S_0 = 0, S_1 = 1$). If this assumption holds then the observed outcomes in the treatment group will consist of individuals from the second and third group, and the observed outcomes in the control group will only consist of individuals from the second group ('the always observed'). Lee (2009) shows that sharp lower and upper bounds can be obtained for the group of 'always observed' individuals by trimming the treatment group with the proportion of excess individuals, which are the individuals that are observed because of the treatment (the third group).

(5) $$p = \frac{\Pr(S = 1 | D = 1) - \Pr(S = 1 | D = 0)}{\Pr(S = 1 | D = 1)}$$

Upper bounds are obtained after trimming the lower tail of the outcome distribution; lower bounds are obtained by trimming the upper tail of the outcome distribution. A recent study applies Lee (2009)'s general sample selection model within a regression discontinuity framework (Kim, 2013). In this setting the treatment effect is identified at the cut-off, for instance, when the birth date is equal to the cut-off date ($B = c$). Kim (2013) shows that the lower bound δ_L and the upper bound δ_U are sharp and can be obtained as:

(6) $$\delta_L = E(Y_1 | B = c, S_1 = 1, Y_1 \leq y_{1-p}) - E(Y_0 | B = c, S_1 = 1, S_0 = 1)$$

$$= \lim_{b \downarrow c} E(Y | B = b, S_1 = 1, Y \leq y_{1-p}) - \lim_{b \uparrow c} E(Y | B = b, S_1 = 1, S_0 = 1)$$

$$\delta_U = E(Y_1 | B = c, S_1 = 1, Y_1 \geq y_p) - E(Y_0 | B = c, S_1 = 1, S_0 = 1)$$

$$= \lim_{b \downarrow c} E(Y | B = b, S_1 = 1, Y \geq y_p) - \lim_{b \uparrow c} E(Y | B = b, S_1 = 1, S_0 = 1)$$

where p is the proportion of individuals whose outcomes are only observed in the treated state (the third group of individuals), $y_q \equiv G^{-1}(q)$, with G is the c.d.f. of Y, conditional on being at the cut-off; $G(y) = P(Y_1 \leq y | B = c, S_1 = 1)$.

Based on this approach, introduced by Lee (2009) and extended by Kim (2013), we have developed a trimming-procedure for our specific setting in which sample selection is determined by the sampling of two grades only. In our setting the treatment status changes from the control state to the treated state if we move from the right to the left side of the cutoff. On the right side of the cutoff we observe students who are on track or accelerated, on the left side of the cutoff we observe students who are on track or delayed. Hence, we do not observe delayed students in the control group and accelerated students in the treatment group. For obtaining a group of always observed individuals we restrict the sample of students in the control group to those that are on track. This means that our estimation sample consists of 'on-track students' in the treatment and control group, and delayed students in the treatment group. 'On-track' students in the control group are the oldest in their grade. The monotonicity assumption (Equation (4)) is that students that are on track when they are the oldest in class will also be observed when they would be born on the other side of the cutoff date. Hence, these students would not be accelerated when they would be the youngest in class. This seems not a very strong assumption since relatively old students are more likely to be accelerated than relatively young students. In addition, the proportion of accelerated students is very small (approximately 0.6%), especially for the youngest in class. If this assumption holds then all students that are on track on the right side of the cutoff would also be observed on the left side of the cutoff when they would have been born a few months earlier. A second issue in our setting is that we do not directly observe the excess proportion of students, which is used for the trimming of the treatment group (Equation (5)). However, we can approximate this excess proportion by using the observed proportions of students that are not on track and born further away from the cutoff; the oldest students in the oldest age-cohort and the youngest students in the youngest age-cohort. This estimate of the excess proportion is probably too large which implies that our upper and lower bounds estimates will be conservative. By applying this procedure we obtain lower and upper bound estimates for students that are on track on the right side of the cutoff date. For the selection of countries used in the analysis this is approximately 97 % of the population of 9-year olds and 95 % of the population of 13-year olds. The standard errors are obtained by using Lee (2009)'s analytical standard errors. Further details on this trimming-procedure are provided in Appendix 1.

In sum, to mitigate various identification issues we produce three types of estimates: 1) reduced form estimates, 2) IV-estimates and 3) upper and lower bounds. The reduced form models estimate the effect of being eligible for an additional year of education. The IV-models estimate the effect of one year in grade for those students who move to the next grade if their date of birth would change to the left side of the cutoff date. The lower and upper bound estimates should be interpreted as bounds for the effect of one year in school for students that are on track and born at the right side of the cutoff date. These three types of estimates will be equal if there is full-compliance with the school entry rules. To improve comparability of these estimates we only select countries with high compliance with the school entry rules. The reduced form and IV-estimates are comparable to previous estimates from the literature.

4. Data

The data used in this study come from the 1995 TIMSS study.[15] 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.[16] 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 that are one year younger or older than the age groups that were targeted. This sampling

[15] See, http://timss.bc.edu/timss1995i/Database.html for TIMSS data.
[16] These data were also used in a quasi-experimental study on international differences in class size effects (Woessmann & West, 2006).

strategy enables us to apply the regression discontinuity framework that we discussed in Section 3. After the 1995 TIMSS 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. In addition, to mitigate concerns about sampling bias we only select countries with no retention or acceleration of students, or in which retention or acceleration of students is very uncommon. For the nine-year-olds we included 10 out of 26 participating countries. Norway, Singapore, Iceland, Japan, Scotland and England were included because retention or acceleration of students is not allowed or very uncommon. In addition, we included Greece, Cyprus, Canada and Portugal because they also have very low proportions of students that are not on track, indicated by their first stage estimate (first stage>0.7). For the thirteen-year-olds we included 13 out of 41 participating countries. Singapore, Sweden, Norway, Iceland, Scotland, Japan and England were included because of their policies with respect to retention and acceleration. Italy, Spain, Cyprus, Slovak Republic, Greece and Belgium were also included because they have low proportions of students that are not on track. Bedard and Dhuey (2006) included 11 countries for 9-year-olds and 19 countries for thirteen-year-olds. Although we do not restrict our sample to OECD-countries we have selected fewer countries to mitigate concerns about sampling bias due to non-compliance with the school-entry rules.

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. We use the average of these five plausible values for each student which is the common approach in the empirical literature that uses international test scores.

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

& Dhuey (2006) and several online sources, and empirically checked this information in our data (see Appendix 2).

Table A.3 in the appendix shows mean values for the test scores and covariates by month of birth relative to the cutoff date for the pooled sample of countries. This gives a first impression of the effect of time in school and of the similarity of students born around the cutoff date. At the cutoff date we observe for the 9-years-olds a difference in test scores of 32 point in math and 33 points in science (Table A.3a). This difference is much larger than the other differences between months of birth suggesting that time in school is important for student performance. The mean values for the covariates are quite similar across the cutoff date. For the 13-year-olds we also observe the largest difference in test scores at the cutoff date; 15 points in math and 21 points in science (Table A.3b). The mean values of the covariates are similar but the parental education of students observed at the right side of the cutoff date (relative age = 0) is somewhat higher. This suggests that at this age some sorting might have taken place. To further assess the similarity of students born at either side of discontinuity 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 A.4 in the appendix). In general, students across the cutoff are very similar. For the 9-year-olds we only find a marginally significant difference for being born in country of test. For the 13-year-olds we find that students born just before the cutoff date (the youngest in their age cohort) have somewhat lower educated mothers in the discontinuity sample of six months around the cutoff. However, this difference is no longer statistically significant in the smaller discontinuity sample.

5. The effect of time in school on cognitive skills across countries

This section presents the first part of our empirical analysis. We estimate the average effect of one year of time in school or time in grade for 9-year-olds and 13-year-olds, on the performance in cognitive tests for pooled samples of countries. For the 9-year-olds we include 10 countries, for the 13-year-olds we include 13 countries.

To obtain estimates of the average effect of one year of time in school (grade) in the selected countries we have pooled the data for each test (TIMSS 9, TIMSS 13) and estimated the models specified in Section 3. In these models we have also included country dummies and interactions of these dummies with a linear function of age which is allowed to be different at either side of the cut-off for each country. Table 1 shows the estimated effects for 9-year-olds for math and science. Panel A shows the results for the sample of six countries that do not allow retention or acceleration of students. Panel B includes 4 additional countries that also have low proportions of students that are not on track but allow retention or acceleration of students. We show three types of estimates. Columns (1) to (5) show the reduced form estimates, columns (6) and (7) show the instrumental variable estimates of time in grade, and columns (8) and (9) show the lower and upper bounds estimates for students that are on track. All models include a linear specification of age; the model in column (5) includes a quadratic specification of age. The additional controls include gender, born in country of test, lives with mother/father, language of test spoken at home and number of books at home. Columns (6) and (7) also report the first stage estimate from Equation (1). Columns (8) and (9) report the trimming proportion used for obtaining lower and upper bound estimates. For the reduced form estimates we use two discontinuity samples around the cut-off date: ± 3 months and ± 6 months. Standard errors have been adjusted for clustering at the school level.

The estimates in columns (1) to (5) of Panel A in Table 1 show that one year of time in school increases performance of 9-year olds with 29 to 31 points in math and with 23 to 24 points in science, which is approximately 0.2 and 0.3 standard deviations of test scores (one standard deviation of test scores equals 100 points). The estimates are precise, and robust to the discontinuity sample. The inclusion of controls only slightly changes the estimated effects, which confirms that students born around the cut-off date are quite similar in observed characteristics. Using a quadratic instead of a linear specification does not change the results (column (5)). A linear specification seems most sensible for our local discontinuity samples as this is what the Akaike-test suggests, and this is also consistent with what we see in Figure 1 and similar figures for the pooled sample of countries. The IV-estimates reported in columns (6) and (7) are slightly higher but lie in the same ball park. One year in grade 4 increases performance of 9-year olds with 34 points in math and with 25 points in science. These estimates should be interpreted as the effect for the subpopulation of students that would move from grade 3 to grade 4 if their date of birth would change from the right to the left side of the cutoff date. A concern with the previous estimates is bias resulting from the sampling design based on two grades only. To mitigate this concern we have estimated lower and upper bounds of the treatment effect (columns (8) and (9)). For obtaining the lower (upper) bound estimate the upper (lower) tail of the outcomes distribution of the treatment group has been trimmed with 2 percent which reduces the estimation sample with approximately 400 students. We find that the lower bound estimates are close to the estimates in columns (1) to (5) suggesting that one year of time in school increases the test scores in math with 0.3 standard deviations and in science with 0.2 standard deviations. The upper bound estimates are 5 to 6 points higher than the estimates in columns (1) to (5). The lower and upper bound estimates should be interpreted as the treatment effect for the group of students that is on track at the right side of the cutoff date. It should be noted that the estimates in columns (1) to (7) are closer to the lower bound than to the upper bound estimates. This is consistent with a downward bias of these estimates due to missing outcomes of accelerated and retained students, and suggests that missing outcomes due to redshirting of students is less important.

Panel B shows the estimates after extending the sample with four countries that also have low proportions of students that are not on track but allow retention or acceleration of students. In general, the estimated effects are quite similar to the results in panel A. The lower bound estimates suggest that one year of time in school increase performance in math with 27 points and in science with 19 points. The IV-estimates are also very similar to the estimates obtained for the sample of six countries. A concern with the IV-estimates is that the exclusion restriction might not hold for countries that allow retention or acceleration. The estimates for this sample of ten countries also show that one year of time in school increase performance in math with 0.3 standard deviations and in science with 0.2 standard deviations.

Table 2 shows the effects of time in school for 13-year-olds. Panel A shows the results for a sample of seven countries that do not allow retention or acceleration of students. The reduced form estimates in columns (1) to (5) suggest that one year of time in school increases performance by 12 to 14 points in math and by 22 to 24 points in science. These estimates are robust to the discontinuity sample, the specification of age and the inclusion of additional controls. We also find that the IV-estimates are only slightly larger than the reduced form estimates which results from the high proportion of students that is on track in the countries that have been selected. The lower bound estimates are very close to the reduced form estimates; 11 points for math and 21 points for science. These results are confirmed when we increase the sample of countries with six additional countries. In Panel B of Table 2 we find estimates in the same ball park. The range of the lower and upper bound estimates is larger which reflects the increase of the trimming proportion from 3 to 6 percent. In sum, these estimates suggest that one year of time in school increase performance of 13-year-olds with approximately 0.1 standard deviations in math and 0.2 standard deviations in science.

We investigated the sensitivity of the results for using the exact date of birth as assignment variable and for a potential non-randomness of birth around the cutoff. Table A.5 in the appendix shows the results. The first two columns show the results for using the exact date of birth instead of the month of birth as assignment variable. The third column shows the results after excluding students born in the period of three days around the cutoff date. The estimates are robust to these changes.

In sum, the cross-country estimates of the effect of time in school and the effect of time in grade yield two important findings. First, a year of school time or a year of time in grade 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. The cross-country estimates of the effect of time in grade are slightly larger but lie in the same ballpark. The estimated 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 or time in grade matters more for performance in math at the age of 9 than at the age of 13. At the age of 9 one year in school increases performance with 0.3 standard deviations, at the age of 13 the effect reduces to 0.1 standard deviations. For science we do not observe this pattern but when we exclude Singapore, which has a very large gain in science (see next section), we find a similar result. Hence, the effect of time in school or time in grade 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.

6. International differences in gains in cognitive skills

The second part of 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 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 being eligible for one additional 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 Portugal and England. Column (2) shows IV-estimates of the effect of one year of time in grade. Columns (3) and (4) show the lower and upper bound estimates for the students that are on track at the right side of the cutoff date. These estimates are obtained after trimming the estimation sample with the trimming proportion (column (12)). Column (5) shows the mean level score of the highest of the adjacent grades for each country. Singapore and Japan have the highest scores, whereas Iceland and Portugal have the lowest level scores in math in the upper grades. We call these average scores the country level scores and we will compare them with our estimates of the effect of time in school. Columns (6) to (10) show the results for the science test. Column (11) 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 and six month after the cutoff date.

The estimates of the effect of time in school or time in grade on the performance in math and science show large differences in the gains in cognitive skills between countries. The estimated effects differ between 0.1 and 0.5 standard deviations of test scores. In Norway, Singapore and Iceland one year of time in school increases the average performance of students in both tests with 0.3 to 0.5 standard deviations. On the other hand, students in Portugal or England approximately gain only 0.1 to 0.2 standard deviations of test scores. These differences between countries are quantitatively meaningful and might have serious implications for the economic outcomes of nations over time. For instance, Hanushek & Woessman (2008) conclude that one standard deviation of test scores is associated with 1 to 2 %-points additional annual GDP-growth. We observe that the three types of estimates in Table 3 show a consistent picture.

The estimated effects of time in school and time in grade are quite similar, which follows from the small proportions of delayed or accelerated students. The lower bound estimates are also very similar to the reduced form estimates; the range of the upper bound estimates is somewhat larger. We also observe that countries that have high gains in achievement in math also achieve high gains in science; the correlation between the estimates in columns (1) and (6) is 0.9. It seems not likely that the large differences in cognitive gains between countries can be explained by differences in the sampling of students for our estimation. We have selected countries with small proportions of delayed or accelerated students. In addition, for the six countries that do not allow retention or acceleration of students we observe countries with high gains and countries with low gains in achievement.

Differences in achievements of 13-year-olds between countries

The sample of countries used for estimating the effect for 13-years-olds consists of 13 countries. Table 4 shows the estimation results. In general, the estimates of the gains in cognitive skills are substantially smaller for 13-years-olds than for 9-year-olds. Again we observe large differences between countries. Singapore achieves the highest gains in cognitive skills in both subjects; the results for science are far ahead of all other countries. Sweden and Norway also generate relatively high gains of approximately 0.2 standard deviations for both tests. Remarkably, at the bottom of the ranking in Table 4 we observe several countries, for instance England, for which time in school has a very small and statistically insignificant effect on performance in math and science. This suggests that in these countries one additional year of time in school does not matter for the performance on the TIMSS math or science tests. The three types of estimates show a consistent pattern for countries with high proportion of students that are on track as indicated by a high first stage estimate. For countries with larger proportions of students that are not on track (Italy, Spain, Cyprus, Greece and Belgium) the range of the lower and upper bound estimates are larger. For these countries the trimming proportions are larger which affects the estimates of the lower and upper bounds. As for the 9-year-olds 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 (6) is 0.81. We have also investigated whether countries that have relatively high achievement gains for 9-year-olds also have relatively high achievement gains for 13-year-olds, and whether countries with relatively low gains for 9-year-olds also have relatively low gains for 13-year-olds. We find a correlation of 0.6 for the reduced form estimates for math and a correlation of 0.5 for the reduced form estimates for science. This suggests 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. Again, it seems not likely that differences in the proportions of observed students can explain the large differences in gains in cognitive skills between countries. We also observe large differences in cognitive gains between countries that do not allow retention or acceleration of students (countries in grey). For instance, Singapore, Iceland, Japan and England have similar high first stage estimates but strongly differ in the gains in achievement in math and science. 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 (see Table A.6 in the appendix).

In sum, we find large differences between countries in the gains in cognitive skills from one year in school or one year in grade. These differences are found for both tests and both age groups. The three types of estimates show a consistent pattern. In addition, countries that perform well in math also perform well in science and countries that perform well for 9-year-olds also do well for 13-year-olds. Differences in the proportions of observed students cannot explain the differences in cognitive gains.

7. Do gains 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. This raises the question whether the rankings of the estimates of gains in cognitive skills in Tables 3 and 4 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. Low level scores might also be misleading about the performance of the education systems if the conditions outside school are unfavorable for learning.[17] 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 4).

For investigating whether the two measures show a consistent ranking of countries we compare the estimates of the gain scores with the mean upper grade scores.[18] 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 2 and 3 we have plotted the mean upper grade scores for the different age groups and subjects on the vertical axis against the reduced form estimates of the gains in achievement on the horizontal axis. Figure 2 shows the results for the 9-years-olds in math and science. Figure 3 shows the results for the 13-years-olds. We have included axes at the median level of gains 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 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 close to the median level scores. Hence, countries close to the median 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. For science we observe a more even distribution of countries across the four quadrants of performance. An almost similar pattern is found for the 13-year-olds (Figure 3). We observe no association between mean level scores and gains in cognitive skills for math. 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. For science we observe a positive association but this association is completely driven by the exceptional performance of Singapore. The association disappears after the

---

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

[18] 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.

exclusion of Singapore and the correlation becomes negative for both subjects. Figures based on the IV-results or lower (upper) bound estimates are very similar and show no association between level scores and gain scores.

The results from Figures 2 and 3 have been summarized in Table 5. This table shows correlations between the estimates of the gains in achievement and the country level scores by test and age group. The left panel shows the correlations for the reduced form estimates, the middle panel shows the correlations for the IV-estimates and the right panel shows the correlations for the lower and upper bounds estimates. The main finding of Table 5 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 reduced form estimates, the IV-estimates, and for the lower and upper bound estimates. The largest associations are found for 13-year-olds. However, these correlations are completely driven by the large gains of Singapore. The correlations become negative and statistically insignificant when the outlier Singapore is excluded.

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.[19] For countries in these two quadrants a mere focus on the ranking along the vertical axis might yield misleading information for educational policy.

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. The estimates of the effect of time in grade, which should be interpreted as local average treatment effects, are similar but slightly higher. 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.

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

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

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 only 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.

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, than level scores. The estimation of gain scores is likely to be useful for improving decisions on educational policies.

References

Angrist, Joshua. D., and Alan B. Krueger. (1991). "Does compulsory school attendance affect schooling and earnings?" *Quarterly Journal of Economics*, 106(4): 979–1014.

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

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

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

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, Samuel, Sebastian Galiani, and Patrick J. McEwan. (2011) "Preschool and maternal labor supply: Evidence from a regression discontinuity design," *Economic Development and Cultural Change*, 59(2): 313-344.

Black, Sandra, Paul J. Devereux, and Kjell G. Salvanes. (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. and D.A. Jaeger, (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. and D. Hungerman, (2013), Season of birth and later outcomes: old questions, new answers, *Review of Economics and Statistics*, 95 (3), 711-724.

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

Cahan, S. & D. Davis, (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.

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.

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

Dickert-Conlin, Stacy and Amitabh Chandra. (1999) “Taxes and the Timing of Births,” *Journal of Political Economy*, 107(1): 161–177.

Dickert-Conlin, Stacy and Todd Elder. (2010) “Suburban Legend: School Cutoff Dates and the Timing of Births,” *Economics of Education Review*, 29(5): 826–841.

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

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

Fredriksson, Peter, and Björn Öckert. (forthcoming), “Life-cycle Effects of Age at School Start,” *Economic Journal*

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.

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., 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.

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

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

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

Leuven, Edwin, Mikael, Lindahl, Hessel, Oosterbeek, and Dinand, Webbink. (2010) “Expanding schooling opportunities for 4-year-olds,” *Economics of Education Review*, 29: 319–328.

Lee, D., Lemieux T., (2010) “Regression Discontinuity Designs in Economics,” *Journal of Economic Literature* 48(2), 281-355.

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

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

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.

McCrary, Justin, and Heather Royer. (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, Patrick J., Joseph S. Shapiro. (2008) “The benefits of delayed primary school enrollment: Discontinuity estimates using exact birth dates,” *Journal of Human Resources*, 43(1): 1–29.

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

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

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

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.

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

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

Tables

Table 1: Estimates of the effect of one year of school time by subject for pooled sample of countries, 9-year-olds.

	Reduced Form (RF)					Using ± 6 months sample			
	± 3 months		± 6 months			IV		LB	UB
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: 6 countries with no retention/acceleration									
math	29.4***	29.1***	30.9***	31.0***	30.6***	33.6***	33.7***	29.7***	36.9***
	(4.0)	(3.7)	(3.1)	(2.9)	(4.2)	(3.3)	(3.1)	(2.8)	(2.9)
N	10680	10680	21340	21340	21340	21340	21340	20942	20942
Coefficient first stage/Trimming proportion						*0.92*	*0.92*	*0.02*	*0.02*
science	22.6***	22.5***	23.1***	23.4***	23.8***	25.1***	25.5***	21.1***	28.7***
	(3.9)	(3.6)	(2.8)	(2.6)	(4.1)	(3.0)	(2.8)	(2.6)	(2.6)
N	10680	10680	21340	21340	21340	21340	21340	20942	20942
Coefficient first stage/Trimming proportion						*0.92*	*0.92*	*0.02*	*0.02*
Panel B: 10 countries (first stage>0.7)									
math	26.6***	26.2***	29.2***	28.4***	25.9***	34.1***	33.2***	26.5***	38.0***
	(2.9)	(2.6)	(2.1)	(1.9)	(3.0)	(2.4)	(2.2)	(1.9)	(2.0)
N	18286	18286	37297	37297	37297	37297	37297	36311	36311
Coefficient first stage/Trimming proportion						*0.86*	*0.85*	*0.03*	*0.03*
science	18.9***	18.6***	22.2***	21.5***	18.6***	25.9***	25.2***	19.2***	30.8***
	(3.0)	(2.6)	(2.0)	(1.8)	(2.9)	(2.3)	(2.1)	(1.9)	(1.8)
N	18286	18286	37297	37297	37297	*0.86*	*0.85*	36311	36311
Coefficient first stage/Trimming proportion								*0.03*	*0.03*
Birth month control (forcing variable)	linear	linear	linear	linear	square	linear	linear	linear	linear
Additional controls	no	yes	no	yes	yes	no	yes	yes	yes

Notes: Each cell shows results from a separate regression. The outcome (math/science test score) is regressed on the dummy for being born left of the cutoff date, country dummies and interactions of these dummies with the polynomial in birth month/day. For each country the functional form is allowed to be different on either side of the cutoff. The models in the columns (2), (4), (5), (7), (8) and (9) control for gender, born in country of test, lives with mother/father, language of test spoken at home and number of books at home and dummies for when one of the variables is missing.

Table 2: Estimates of the effect of one year of school time by subject for pooled sample of countries, 13-year-olds.

	Reduced Form (RF)					Using ± 6 months sample			
	± 3 months		± 6 months			IV		LB	UB
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: 7 countries with no retention/acceleration									
math	11.9***	12.7***	12.9***	11.9***	14.0***	14.2***	13.1***	10.3***	17.2***
	(3.3)	(2.6)	(2.4)	(2.1)	(3.4)	(2.6)	(2.3)	(2.3)	(2.2)
N	10478	10478	21371	21371	21371	21371	21371	20913	20913
Coefficient first stage/Trimming proportion						*0.91*	*0.91*	*0.03*	*0.03*
science	21.9***	22.7***	23.0***	21.9***	23.6***	25.3***	24.0***	20.2***	28.3***
	(3.5)	(3.0)	(2.5)	(2.2)	(3.3)	(2.8)	(2.4)	(2.3)	(2.2)
N	10478	10478	21371	21371	21371	21371	21371	20913	20913
Coefficient first stage/Trimming proportion						*0.91*	*0.91*	*0.03*	*0.03*
Panel B: 13 countries (first stage>0.7)									
math	12.0***	12.2***	11.0***	10.6***	13.7***	13.1***	12.6***	5.4***	20.6***
	(2.4)	(2.2)	(1.7)	(1.5)	(2.3)	(2.0)	(1.8)	(1.7)	(1.6)
N	19228	19228	39191	39191	39191	39191	39191	37512	37512
Coefficient first stage/Trimming proportion						*0.84*	*0.84*	*0.06*	*0.06*
science	19.2***	19.4***	17.7***	17.3***	19.4***	21.0***	20.5***	11.9***	28.4***
	(2.7)	(2.4)	(1.8)	(1.6)	(2.6)	(2.1)	(1.8)	(1.6)	(1.6)
N	19228	19228	39191	39191	39191	39191	39191	37512	37512
Coefficient first stage/Trimming proportion						*0.84*	*0.84*	*0.06*	*0.06*
Birth month controls (forcing variable)	linear	linear	linear	linear	square	linear	linear	linear	linear
Additional controls	no	yes	no	yes	yes	no	yes	yes	yes

Notes: Each cell shows the results of a separate regression. The outcome (math/science test score) is regressed on the dummy for being born left of the cutoff date, country dummies and interactions of these dummies with the polynomial in birth month/day. For each country the functional form is allowed to be different on either side of the cutoff. The models in the columns (2), (4), (5), (7), (8) and (9) control for gender, born in country of test, lives with mother/father, language of test spoken at home and number of books at home, parental education, and dummies for when one of the variables is missing.

Table 3: Gain in achievement (standard error) and mean upper grade by country for 9-year-olds, math & science

		Math					Science							
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Ranking	Country	RF	IV	LB	UB	mean	RF	IV	LB	UB	mean	first stage	trimming %	N
1	Norway	43.3 (6.4)	48.4 (7.1)	40.1 (6.3)	52.6 (6.4)	502	38.8 (8.0)	43.4 (8.9)	39.1 (8.4)	46.8 (7.8)	530	0.89 (0.02)	2.0	2133
2	Singapore	40.0 (7.2)	41.7 (7.5)	35.3 (7.1)	46.0 (7.1)	625	27.8 (6.4)	29.1 (6.6)	23.1 (6.4)	33.0 (6.4)	547	0.96 (0.01)	2.8	6986
3	Greece	36.3 (8.1)	40.9 (8.9)	31.0 (7.6)	48.2 (7.6)	492	21.7 (7.2)	24.5 (8.0)	19.6 (7.2)	30.6 (6.6)	497	0.89 (0.02)	2.7	2981
4	Iceland	34.6 (5.5)	35.9 (5.7)	36.6 (6.0)	39.8 (5.6)	474	30.9 (7.5)	32.1 (7.8)	32.9 (7.8)	35.9 (7.6)	505	0.96 (0.01)	2.3	1702
5	Cyprus	30.5 (5.7)	38.8 (7.2)	29.9 (6.0)	48.0 (6.1)	502	20.4 (5.7)	25.9 (7.1)	19.6 (5.6)	39.4 (5.2)	475	0.79 (0.02)	5.2	3230
6	Japan	28.6 (4.5)	30.3 (4.7)	28.1 (4.4)	31.4 (4.4)	597	19.1 (4.2)	20.2 (4.4)	19.0 (4.2)	22.9 (4.2)	574	0.94 (0.01)	0.8	4343
7	Canada	24.4 (4.1)	34.6 (5.7)	23.0 (4.1)	40.3 (4.2)	532	21.7 (4.6)	30.8 (6.4)	17.3 (4.7)	38.6 (4.5)	549	0.70 (0.02)	4.8	7436
8	Scotland	20.1 (6.1)	26.1 (7.7)	18.8 (5.9)	33.0 (6.3)	520	12.9 (6.9)	16.7 (8.9)	10.7 (7.1)	23.6 (7.5)	536	0.77 (0.03)	3.1	3089
9	Portugal	17.7 (7.0)	22.4 (8.7)	11.2 (6.9)	30.6 (6.8)	475	17.9 (8.0)	22.7 (9.9)	10.1 (7.3)	34.5 (7.7)	480	0.79 (0.03)	6.5	2310
10	England	13.5 (5.9)	14.4 (6.2)	12.2 (6.0)	20.2 (5.8)	513	12.9 (6.9)	13.8 (7.3)	6.4 (6.8)	21.0 (6.8)	551	0.93 (0.02)	2.2	3087

Notes: The sample for estimating gains in achievement scores consists of students born in the period 6 months before and after the cut-off date. The means are the officially reported means for the upper grade, see http://timssandpirls.bc.edu/timss1995i/TIMSSPDF/P1HiLite.pdf. Retention/acceleration is not allowed or very uncommon in countries in grey.

Table 4: Gain in achievement (standard error) and mean upper grade by country for TIMSS 13-year-olds, math & science

		Math					Science							
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Ranking	Country	RF	IV	LB	UB	mean	RF	IV	LB	UB	mean	first stage	trimming %	N
1	Singapore	27.8 (5.2)	29.0 (5.4)	24.4 (5.4)	31.7 (5.3)	643	55.9 (6.8)	58.4 (7.0)	52.1 (7.0)	60.6 (6.8)	607	0.96 (0.01)	2.3	3567
2	Sweden	19.1 (6.0)	20.9 (6.5)	18.0 (6.1)	25.6 (6.0)	519	25.2 (7.1)	27.6 (7.7)	25.6 (7.3)	33.8 (6.8)	535	0.91 (0.01)	1.4	3403
3	Italy	18.5 (7.9)	25.9 (10.9)	-7.4 (8.4)	46.3 (7.8)	493	15.0 (7.2)	21.0 (9.9)	-4.8 (7.3)	49.0 (7.5)	498	0.71 (0.03)	17.9	2186
4	Norway	18.2 (6.0)	21.7 (7.1)	22.2 (6.2)	29.2 (5.9)	503	22.1 (6.3)	26.4 (7.4)	19.1 (6.5)	33.8 (6.0)	527	0.84 (0.02)	4.1	2751
5	Spain	13.6 (5.8)	18.5 (7.7)	-8.2 (5.9)	38.2 (5.7)	487	10.9 (5.3)	14.9 (7.1)	-8.1 (5.5)	30.3 (5.4)	517	0.73 (0.03)	17.6	3157
6	Iceland	13.5 (6.3)	14.3 (6.7)	14.4 (6.5)	16.5 (6.8)	487	17.1 (7.2)	18.1 (7.6)	17.3 (7.5)	24.7 (7.1)	494	0.94 (0.02)	2.1	1834
7	Scotland	9.9 (5.9)	13.1 (7.8)	4.5 (5.9)	25.5 (6.0)	498	12.7 (6.5)	16.8 (8.6)	9.1 (6.9)	28.5 (6.7)	517	0.75 (0.03)	5.5	2824
8	Cyprus	9.5 (7.1)	11.8 (8.8)	-3.6 (7.2)	26.0 (6.4)	474	14.7 (7.2)	18.3 (8.9)	6.6 (6.8)	33.3 (7.7)	463	0.80 (0.02)	7.1	2837
9	Slovak Republic	8.6 (5.5)	10.9 (6.9)	3.6 (5.4)	18.0 (5.3)	547	11.9 (5.6)	15.1 (7.1)	5.7 (5.7)	19.9 (5.6)	544	0.79 (0.02)	5.3	3475
10	Japan	4.8 (4.4)	5.0 (4.5)	3.5 (5.0)	5.7 (4.5)	605	14.1 (4.1)	14.5 (4.2)	13.3 (4.6)	15.7 (4.2)	571	0.98 (0.01)	0.4	5158
11	Greece	3.0 (5.6)	4.1 (7.5)	-9.2 (5.5)	21.1 (6.0)	484	5.9 (5.5)	7.9 (7.3)	-2.4 (5.4)	23.7 (5.8)	497	0.75 (0.02)	7.7	3543
12	Belgium (Flemish)	2.8 (4.8)	3.6 (6.2)	-3.9 (5.0)	18.9 (4.6)	565	12.2 (4.8)	15.9 (6.2)	0.8 (5.1)	30.0 (4.4)	550	0.77 (0.03)	11.8	2622
13	England	-9.0 (9.1)	-9.5 (9.7)	-8.7 (9.0)	-6.8 (9.0)	506	1.8 (9.7)	1.9 (10.3)	-2.3 (9.1)	7.6 (9.6)	552	0.94 (0.02)	1.7	1834

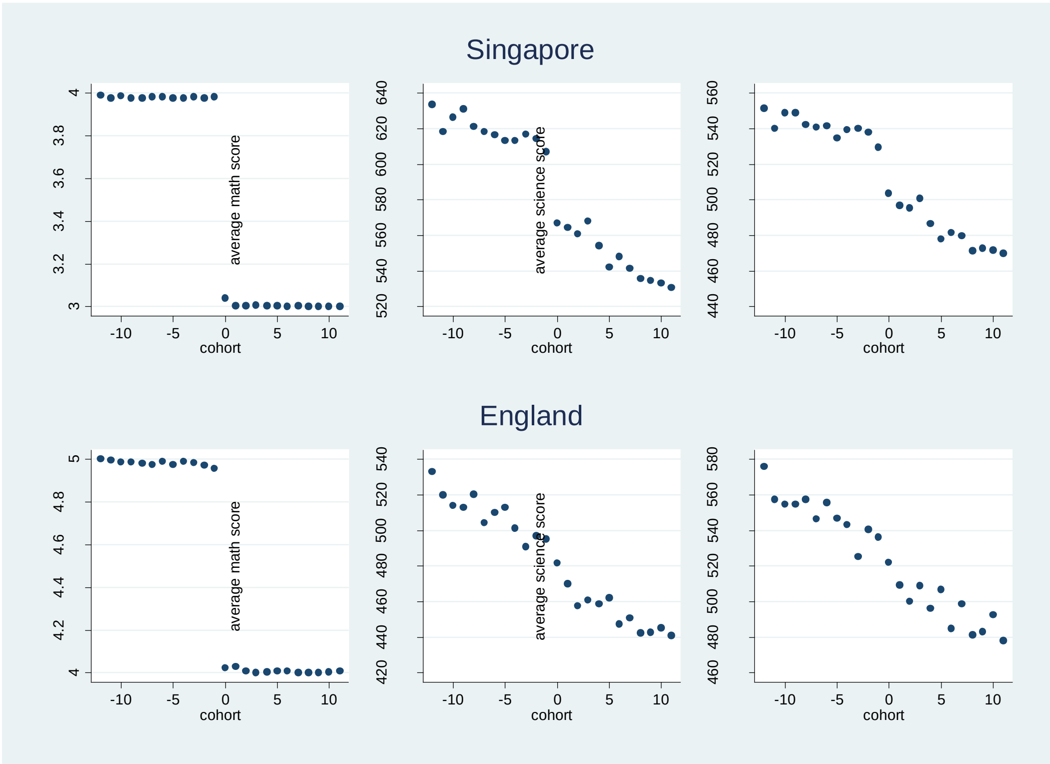
Notes: The sample for estimating gains in achievement scores consists of students born in the period 6 months before and after the cut-off date. The means are the officially reported means for the upper grade, see http://timssandpirls.bc.edu/timss1995i/TIMSSPDF/P1HiLite.pdf. Retention/acceleration is not allowed or very uncommon in countries in grey.

Table 5: Correlations between estimates of gain score and mean upper grade of countries by age group and subject

		± 3 months		± 6 months								
		RF		RF		IV		LB		UB		
test	**subject**	**correlation**	**p-value**	**correlation**	**p-value**	**correlation**	**p-value**	**correlation**	**p-value**	**correlation**	**p-value**	**N**
TIMSS 9	math	0.23	0.53	0.21	0.56	0.13	0.73	0.20	0.59	-0.01	0.99	10
TIMSS 9	science	0.22	0.53	-0.07	0.84	-0.14	0.70	-0.06	0.86	-0.44	0.20	10
TIMSS 13	math	0.42	0.15	0.22	0.47	0.12	0.70	0.42	0.16	-0.17	0.58	13
TIMSS 13	science	0.48	0.10	0.52	0.07	0.47	0.10	0.55	0.05	0.12	0.69	13

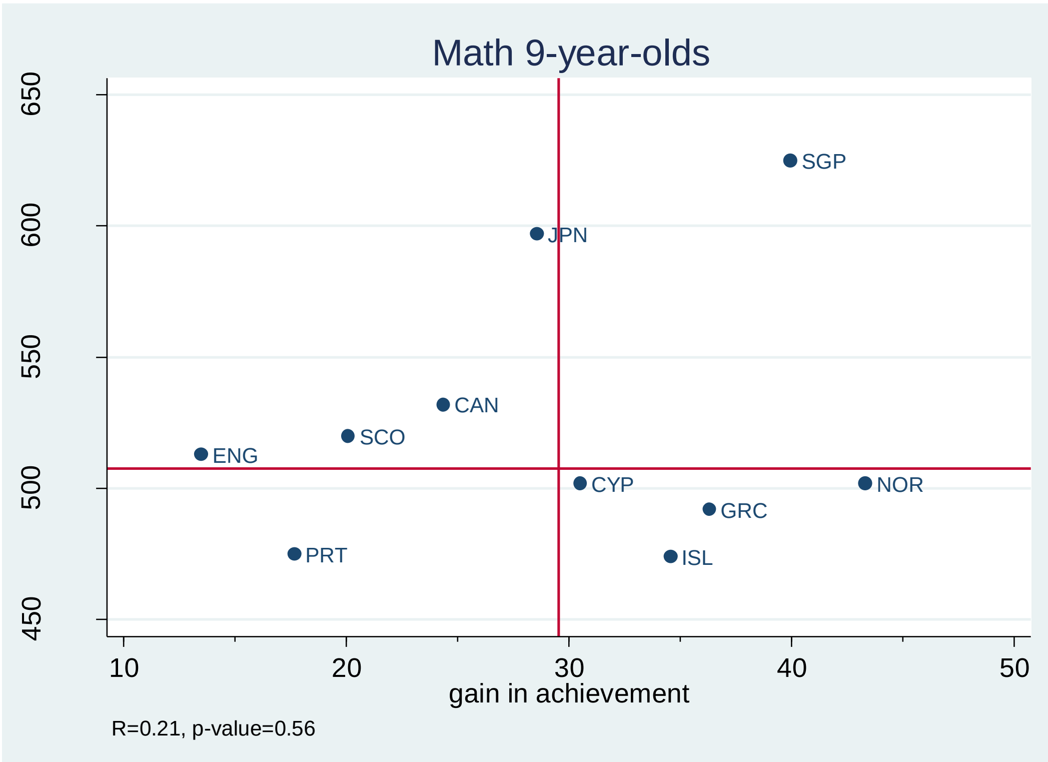
Figures

Figure 1: Grade level and math/science scores around the cut-off date for 9-year-olds from Singapore and England



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 A.3 for the cutoff dates per country.

Figure 2: The association between country level scores and the reduced form estimates of the effect of one year of school time on cognitive skills for 9-year-olds.



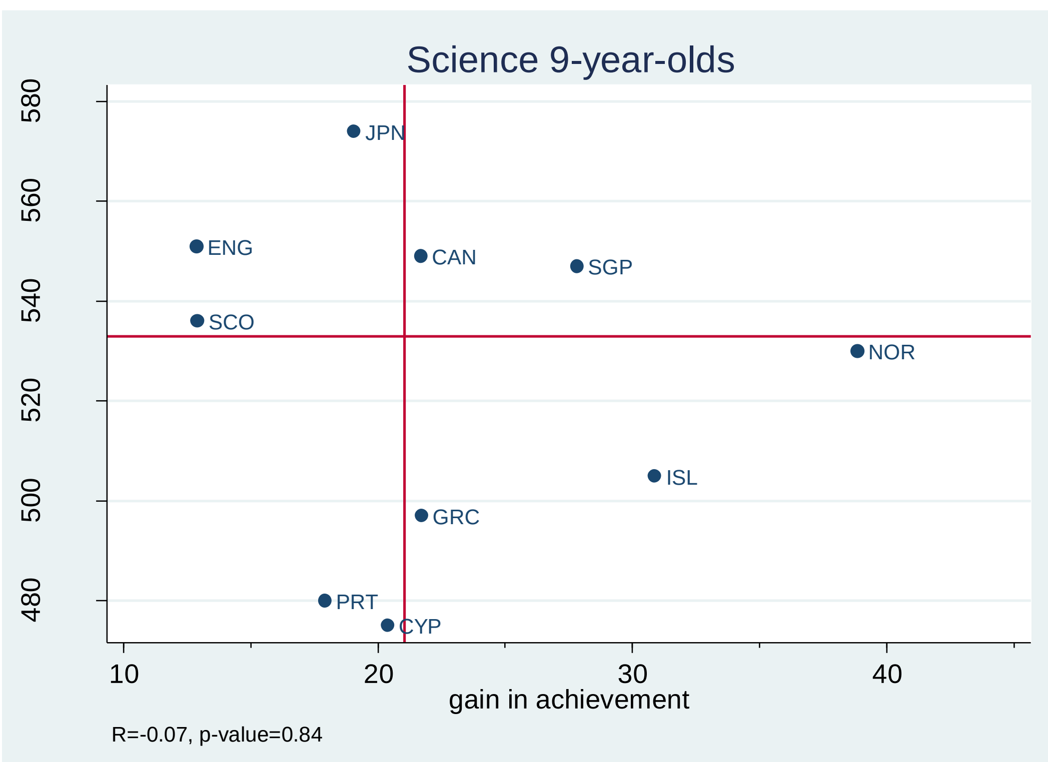
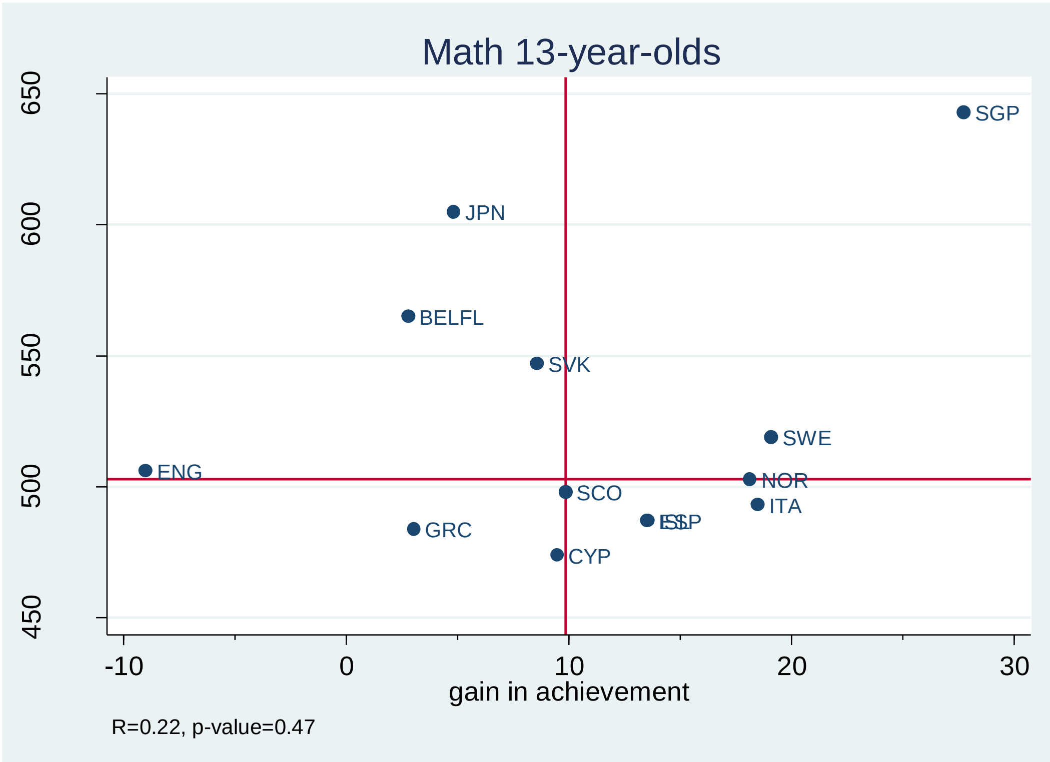
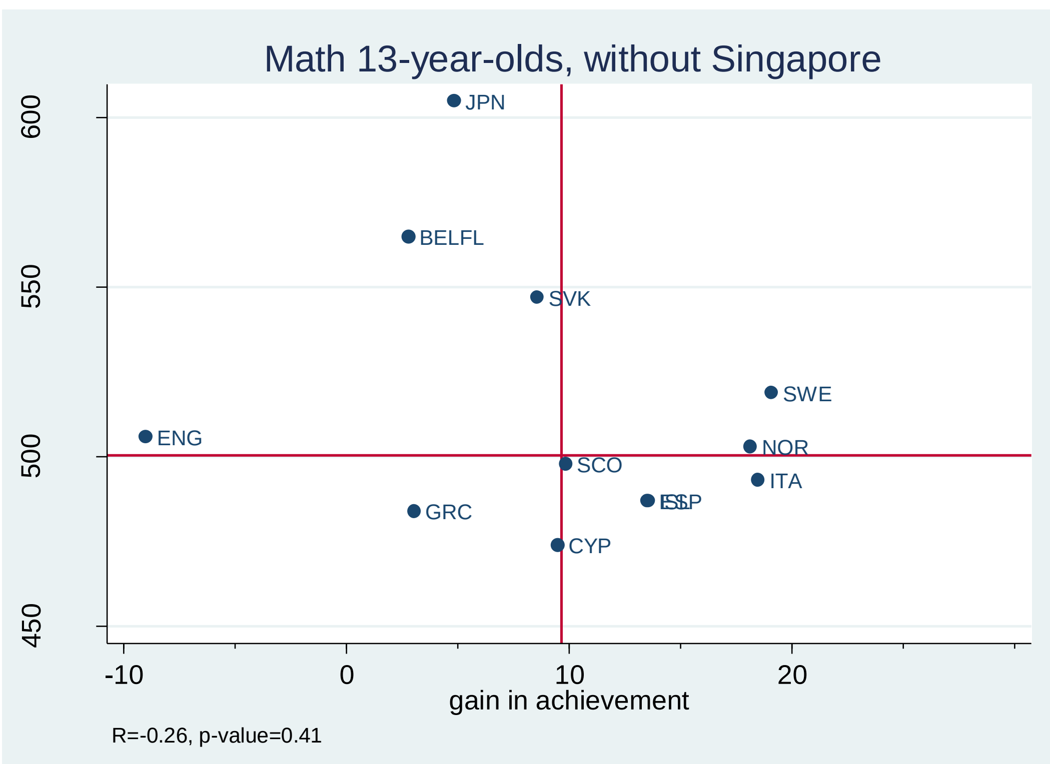
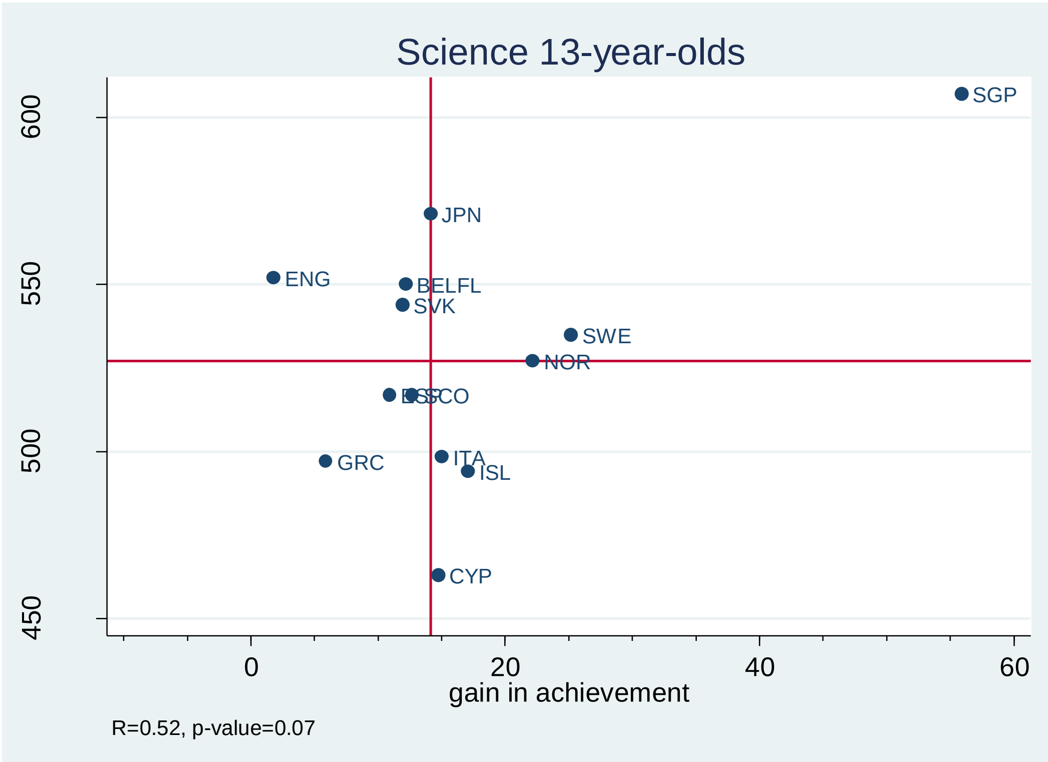


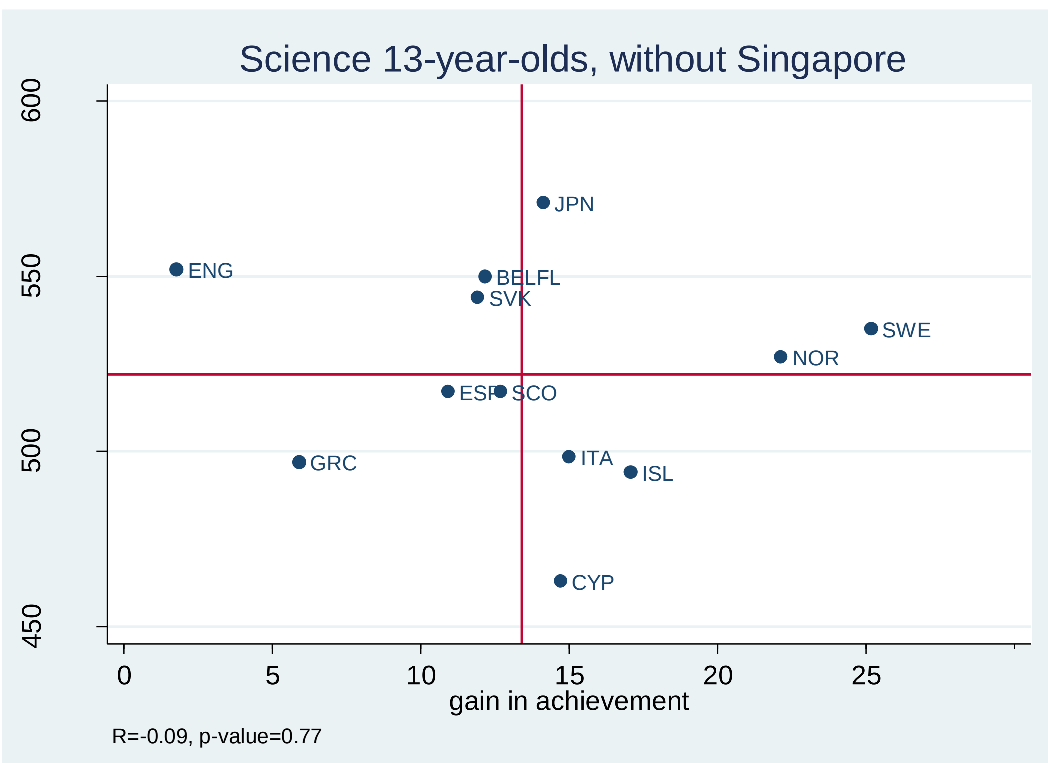
Figure 3: The association between country level scores and the reduced form estimates of the effect of one year of school time on cognitive skills for 13-year olds.







Science 13-year-olds
SGP
JPN
ENG
BELFL
SVK
SWE
NOR
SCO
GRC
ITA
ISL
CYP
600
550
500
450
0
20
40
60
gain in achievement
R=0.52, p-value=0.07



Science 13-year-olds, without Singapore
JPN
ENG
BELFL
SVK
SWE
NOR
ESP
SCO
GRC
ITA
ISL
CYP
600
550
500
450
0
5
10
15
20
25
gain in achievement
R=-0.09, p-value=0.77

Appendix 1 The trimming procedure

The population of students consists of three types: students that are accelerated (a), students that are on time (o) and students that have been delayed (d) due to retention or redshirting. In practice, acceleration of students is not very common, but retention of students is quite common in many countries. Table A.1 shows the proportions (p) within the treatment group (T) and the control group (C). The treatment group are students born before the cut-off date, the control group are students born after the cut-off date. Students in the treatment group are eligible for one additional year in school. The first two columns of Table A.1 show the population of students, columns (3) and (4) shows the students that are observed in our sample.

In our regression discontinuity design we focus on students born six months before or after the cut-off date (columns (5) and (6)). This implies that students in the treatment group are the youngest students within their age cohort, and students in the control group are the oldest students in their age cohort. In our sample we do not observe accelerated students in the treatment group (p_a^T) and delayed students in the control group (p_d^C) because they are not in the two grades that have been sampled. We expect that the proportion of accelerated students in the treatment group (p_a^T) will be smaller than the proportion of accelerated students in the control group (p_a^C) because relatively young students are less likely to be accelerated. Similarly, we expect that the proportion of delayed students in the treatment group (p_d^T) will be larger than the proportion of delayed students in the control group (p_d^C) because relatively young students are more likely to be delayed. This is also what we observe within age cohorts in our data. The aim of the trimming procedure is to construct a group of students that will be observed in both treatment statuses (always observed students). S, S_0 and S_1 are all binary indicators of sample selection. S denotes whether an individual is observed, and S_0 and S_1 are potential sample selection indicators for the treated and control states. Individuals are always observed if $S_0 = 1, S_1 = 1$. The trimming-procedure consists of two steps.

Table A.1 Sample selection, student proportion in population and in sample

Population		Observed in Sample		Estimation sample		Approximating trimming %	
Treated	Control	Treated	Control	-6<=B<=-1	0<=B<=5	-12<=B<=-7	6<=B<=11
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
p_a^T	p_a^C	-	p_a^C	-	-		p_a^{C*}
p_o^T	p_o^C	p_o^T	p_o^C	p_o^T	p_o^C		
p_d^T	p_d^C	p_d^T	-	p_d^T	-	p_d^{T*}	

Notes: B denotes month of birth relative to cutoff date (B=0); p_d^{T*} used for p_d^C and p_a^{C*} used for p_a^T .

Step 1: Creating a group of always observed students ($S_0 = 1, S_1 = 1$)

In the first step the sample of students in the control group is restricted to those that are on track (p_o^C). Hence, students in the control group that are accelerated (p_a^C) are excluded (columns (5) and (6) in Table A.1). The main assumption for applying the trimming procedure is that this restricted sample of students (p_o^C) should also be observed in the treatment group if their treatment status would change (if they had been born before the cut-off date). In our RD-design this assumption holds if students that are on track when they are the oldest in their age-cohort would also be on track or delayed for one year when they would be the youngest in their age-cohort. Hence, these students would not be accelerated when they would be the youngest in their age-cohort. This seems not a very strong assumption since relatively old students are more likely to be accelerated than relatively young students.

Step 2: Determining the trimming proportion

To obtain estimates for the treatment effect for the group of always observed students we need to trim the treatment group with the excess number of students. Hence, we should construct a group of always observed students within the treatment group which is equivalent to the group of always observed students in the control group which has proportion p_o^C . In the second step of the procedure we determine the trimming proportion for the treatment group. In Table A.1 we can observe that this trimming proportion equals:

$$p = (p_a^C - p_a^T) + p_d^C$$

This trimming proportion cannot be obtained directly because we do not observe p_a^T and p_d^C. We approximate these proportions by exploiting data of the youngest and oldest students within their age cohort further away from the cut-off date. Columns (7) and (8) of Table A.1 show that these students are born between birth months -12 and -7 or between birth months 6 and 11. We use the observed proportion of delayed students born between birth months -12 to -7 (p_d^{T*}) to approximate the proportion of delayed students born between birth month 0 and 6 (p_d^C). Similarly, we use the observe proportion of accelerated students born between birth month 7 to 12 (p_a^{C*}) to approximate the proportion of accelerated students born between birth month -6 to -1 (p_a^T). We use these proportions for estimating the trimming proportion p:

$$\hat{p} = (p_a^C - p_a^{C*}) + p_d^{T*}$$

This estimate of p will probably be upward biased because we expect that the proportion of delayed or accelerated students will increase when students have spent more time in the school system. In our data we also observe that the proportions of delayed and accelerated students increase with their grade. For instance, for the 9-year-olds in the pooled sample the proportion of delayed students in the upper grade is 6 % against 4.8 % in the lower grade, the proportion of accelerated students is 1.0 percent in the upper grade and 0.8 percent in the lower grade. For the 13-year-olds we also find higher proportions of delayed or accelerated students in the upper grade than in the lower grade. Hence, we expect that:

$$p_a^T \geq p_a^{C*} \text{ and } p_d^{T*} \geq p_d^C .$$

Using an upward biased trimming proportion implies that the upper and lower bounds of the estimated effect for the always observed students should be interpreted as conservative estimates because we are trimming too many students from the upper or lower tails of the outcome distribution. The trimming proportions are shown in Tables 1 to 4.

We obtained the lower and upper bounds estimates of the effect of time in school by estimating Equation (3) for the trimmed samples. The standard errors are obtained by using the analytical standard errors derived in Lee (2009). We have also adjusted these standard errors for clustering at the school level.

Appendix 2 School entry rules

Table A.2. Cutoff dates per country & source and data availability

country	cutoff date	source	timms 9	timms 13
Belgium-Flemish	January 1	Bedard & Dhuey	no	yes
Canada*	January 1	Bedard & Dhuey	yes	no
England	September 1	Bedard & Dhuey	yes	yes
Greece	April 1	Bedard & Dhuey	yes	yes
Iceland	January 1	Bedard & Dhuey	yes	yes
Italy	January 1	Bedard & Dhuey	no	yes
Japan	April 1	Bedard & Dhuey	yes	yes
Norway	January 1	Bedard & Dhuey	yes	yes
Portugal	January 1	Bedard & Dhuey	yes	yes
Slovak Republic	September 1	Bedard & Dhuey	no	yes
Sweden	January 1	Bedard & Dhuey	no	yes
Singapore[1]	January 1	Internet/TIMMS Data	yes	yes
Scotland[2]	March 1	Internet/TIMMS Data	yes	yes
Cyprus	March 1	Request/ 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. Cutoff dates refer to the situation in 1995. The columns timms 9 and timms 13 show whether the country was used in the analysis. *Because Canada has a first stage>0.7 in timms 9 but <0.7 in timms 13, this country only is used in the former.

1 http://www.moe.gov.sg/education/
2 http://en.wikipedia.org/wiki/Education_in_Scotland

Table A.3a: Means of test scores and covariates by age relative to the cutoff for the pooled sample of 10 countries, TIMMS 9-year-olds

age relative to the cutoff (0)	**math**	**science**	**female**	**born in country**	**speaks language of test at home**	**living with mother**	**living with father**	**number of books at home**	**in highest grade**	**N**
-12	547.5	539.7	0.51	0.92	0.71	0.96	0.82	101.3	0.98	2945
-11	545.1	537.9	0.52	0.92	0.71	0.95	0.81	100.9	0.98	2966
-10	543.1	535.3	0.51	0.92	0.69	0.95	0.81	99.7	0.98	3277
-9	542.6	535.0	0.51	0.92	0.73	0.95	0.81	100.9	0.98	3316
-8	538.6	530.6	0.51	0.93	0.71	0.94	0.81	99.2	0.97	3324
-7	535.8	528.4	0.49	0.93	0.70	0.95	0.81	99.5	0.98	3234
-6	534.4	527.8	0.50	0.92	0.70	0.95	0.81	99.4	0.97	3225
-5	534.0	523.2	0.49	0.91	0.67	0.95	0.82	95.7	0.97	3243
-4	532.6	523.3	0.51	0.91	0.66	0.95	0.81	100.8	0.97	3087
-3	535.6	523.4	0.50	0.92	0.64	0.95	0.82	99.5	0.94	3282
-2	528.8	518.6	0.49	0.92	0.66	0.94	0.81	98.2	0.93	3069
-1	524.8	514.4	0.49	0.92	0.67	0.96	0.82	99.2	0.90	2921
0	492.5	491.1	0.51	0.91	0.69	0.96	0.82	99.4	0.06	3031
1	484.1	483.8	0.50	0.91	0.69	0.95	0.80	102.4	0.02	2951
2	483.4	481.3	0.52	0.92	0.70	0.96	0.81	99.5	0.01	3032
3	481.6	480.8	0.49	0.92	0.70	0.96	0.82	101.4	0.01	3155
4	475.3	473.5	0.52	0.92	0.69	0.95	0.82	97.5	0.01	3227
5	474.4	474.5	0.51	0.91	0.69	0.94	0.81	100.4	0.01	3074
6	474.8	471.3	0.50	0.92	0.68	0.95	0.81	102.8	0.01	3162
7	468.7	466.5	0.50	0.92	0.66	0.94	0.82	98.5	0.01	3026
8	468.9	466.8	0.48	0.91	0.65	0.94	0.81	95.6	0.01	2984
9	469.8	463.4	0.51	0.91	0.63	0.95	0.82	100.0	0.01	2843
10	467.2	462.8	0.50	0.91	0.62	0.94	0.82	100.2	0.01	2592
11	465.5	461.3	0.49	0.93	0.65	0.94	0.82	96.9	0.01	2754

Note: The relative age of the oldest students is -12; relative age 0 means born in the first month at the right side of the cutoff

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

age relative to the cutoff (0)	math	science	female	born in country	speaks language of test at home	living with mother	living with father	number of books at home	high educational level mother*	high educational level father*	in highest grade	N
-12	524.9	524.9	0.51	0.94	0.82	0.97	0.83	104.6	0.24	0.33	0.95	3208
-11	526.3	524.0	0.49	0.94	0.83	0.96	0.84	105.5	0.24	0.33	0.96	3294
-10	524.0	524.1	0.50	0.94	0.83	0.96	0.83	108.7	0.25	0.34	0.96	3514
-9	525.5	523.7	0.53	0.94	0.84	0.96	0.85	106.6	0.28	0.36	0.96	3582
-8	523.2	522.7	0.52	0.95	0.83	0.96	0.83	107.6	0.24	0.33	0.95	3716
-7	518.4	516.4	0.52	0.94	0.82	0.96	0.84	105.1	0.25	0.34	0.95	3492
-6	520.6	516.4	0.50	0.94	0.82	0.97	0.85	104.4	0.24	0.34	0.94	3398
-5	519.3	517.4	0.49	0.94	0.82	0.96	0.85	106.1	0.27	0.35	0.94	3411
-4	517.0	515.5	0.51	0.95	0.81	0.96	0.84	107.3	0.26	0.32	0.93	3431
-3	517.3	514.7	0.49	0.94	0.81	0.97	0.85	106.1	0.24	0.32	0.92	3416
-2	516.7	513.7	0.50	0.95	0.82	0.96	0.84	104.2	0.26	0.34	0.91	3176
-1	515.8	513.1	0.50	0.95	0.81	0.96	0.84	104.7	0.26	0.36	0.89	3252
0	501.1	492.5	0.50	0.95	0.82	0.97	0.84	105.8	0.30	0.38	0.06	3227
1	501.6	492.6	0.51	0.95	0.83	0.97	0.85	106.6	0.29	0.36	0.02	2963
2	498.0	490.7	0.49	0.95	0.82	0.97	0.84	107.2	0.28	0.37	0.02	3194
3	498.1	488.8	0.49	0.95	0.83	0.97	0.85	107.7	0.29	0.35	0.01	3242
4	493.6	485.5	0.51	0.95	0.80	0.97	0.85	106.1	0.26	0.35	0.01	3300
5	495.7	485.4	0.50	0.96	0.82	0.97	0.85	109.6	0.29	0.38	0.01	3181
6	492.1	482.6	0.49	0.95	0.82	0.97	0.86	106.1	0.29	0.38	0.01	3237
7	491.3	482.4	0.49	0.95	0.82	0.97	0.85	104.3	0.29	0.37	0.01	3104
8	490.6	481.4	0.51	0.95	0.79	0.97	0.85	106.6	0.28	0.37	0.01	3130
9	490.4	479.2	0.49	0.95	0.79	0.97	0.85	105.0	0.26	0.34	0.00	2920
10	492.7	481.8	0.50	0.96	0.79	0.97	0.86	104.2	0.28	0.35	0.00	2816
11	489.9	478.7	0.47	0.96	0.79	0.97	0.86	106.1	0.29	0.36	0.00	2774

*High educational level is defined as having vocational education or more.

Table A.4a: Balancing tests TIMMS 9

	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)
Dummy=1 if female	-0.0168	-0.00961	0.00252	0.00120
	(0.0162)	(0.0113)	(0.00203)	(0.00161)
N	18196	37102	18286	37297
Dummy=1 if born in country of test	0.0153*	0.0101	-0.00397	-0.000101
	(0.00891)	(0.00642)	(0.00441)	(0.00319)
N	15762	32145	18286	37297
Dummy=1 if language at home is language of test	0.0132	0.00258	-0.00978	-0.00560
	(0.0134)	(0.00908)	(0.00864)	(0.00585)
N	13353	27205	18286	37297
Dummy=1 if if living with father	-0.00355	-0.00289	0.00249	-0.000936
	(0.0136)	(0.00904)	(0.00430)	(0.00315)
N	15728	32010	18286	37297
Dummy=1 if living with mother	-0.00260	-0.00728	-0.000637	-0.00304
	(0.00700)	(0.00507)	(0.00391)	(0.00299)
N	15773	32092	18286	37297
Number of books at home	-2.647	-1.852	-0.00654	-0.0118**
	(2.610)	(1.745)	(0.00729)	(0.00479)
N	15116	30811	18286	37297

Notes: Each cell shows results of a separate regression. The outcome (in the rows) is regressed on the dummy for being born left of the cutoff, the country dummies and the interactions of these dummies with a linear function in birth month that differs at either side of the cutoff. In columns (1) and (2) results of regressions are reported for the sample for which the variable in the rows is non-missing. In columns (3) and 4) the outcome variable is a dummy for missing.

Table A.4b: Balancing tests TIMMS 13

	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)
Dummy=1 if female	-0.00402	-0.00137	-0.000251	-0.000142
	(0.0161)	(0.0106)	(0.000249)	(0.000168)
N	19227	39189	19228	39191
Dummy=1 if born in country of test	0.000514	-0.000128	0.00524	0.00352*
	(0.00762)	(0.00478)	(0.00327)	(0.00211)
N	16489	33611	19228	39191
Dummy=1 if language at home is language of test	-0.000507	0.00228	-0.00149	-0.00636
	(0.0117)	(0.00819)	(0.00771)	(0.00504)
N	15376	31334	19228	39191
Dummy=1 if if living with father	-0.0121	-0.00958	-0.000477	-0.000800
	(0.0117)	(0.00805)	(0.00362)	(0.00245)
N	16408	33462	19228	39191
Dummy=1 if living with mother	-0.00562	-0.00507	0.00104	-0.00115
	(0.00646)	(0.00409)	(0.00320)	(0.00207)
N	16468	33575	19228	39191
Number of books at home	-1.494	0.136	0.000	0.00190
	(2.425)	(1.658)	(0.00403)	(0.00254)
N	16387	33406	19228	39191
Dummy=1 if educational level mother is high	-0.0195	-0.0256**	-0.00294	-0.0264***
	(0.0181)	(0.0122)	(0.0122)	(0.00796)
N	11276	23092	19228	39191
Dummy=1 if educational level father is high	0.0114	-0.0113	-0.00593	-0.0180**
	(0.0203)	(0.0133)	(0.0124)	(0.00848)
N	10983	22513	19228	39191

Notes: Each cell shows results of a separate regression. The outcome (in the rows) is regressed on the dummy for being born left of the cutoff, the country dummies and the interactions of these dummies with a linear function in birth month that differs either side of the cutoff. In columns (1) and (2) results of regressions are reported for the sample for which the variable in the rows is non-missing. In columns (3) and 4) the outcome variable is a dummy for missing.

Table A.5: Sensitivity analysis for birth day control and timing of birth

	All	All	± 3 days excluded
	(1)	(2)	(3)
Panel A: TIMMS 9 (10 countries)			
math	30.7***	30.4***	31.0***
	(2.6)	(2.6)	(2.6)
N	21924	21924	21576
science	21.5***	21.4***	21.6***
	(2.3)	(2.3)	(2.3)
N	21924	21924	21576
Panel B: TIMMS 13 (13 countries)			
math	9.8***	9.6***	9.5***
	(1.5)	(1.5)	(1.6)
N	33400	33400	32891
science	18.6***	18.2***	17.8***
	(1.5)	(1.5)	(1.6)
N	33400	33400	32891
Birth month controls	linear	no	no
Birth day controls	no	linear	linear
Additional controls	yes	yes	yes

Notes: The sample consists of countries for which exact day of birth is available; not available for Norway, England and Canada in timms 9, and Norway in timms 13. Same specifications as in column (4) of table 1.

Table A.6a: Reduced form estimates of 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 forcing 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 (6.4)	2133	-	-	38.8 (8.0)	2133	-	-
2	Singapore	40.0 (7.2)	6986	38.3 (7.2)	6855	27.8 (6.4)	6986	26.3 (6.5)	6855
3	Greece	36.3 (8.1)	2981	33.4 (11.7)	1055	21.7 (7.2)	2981	23.0 (9.3)	1055
4	Iceland	34.6 (5.5)	1702	34.3 (5.8)	1474	30.9 (7.5)	1702	26.6 (7.6)	1474
5	Cyprus	30.5 (5.7)	3230	32.9 (5.7)	3151	20.4 (5.7)	3230	22.6 (5.5)	3151
6	Japan	28.6 (4.5)	4343	31.7 (4.7)	4268	19.1 (4.2)	4343	21.3 (4.3)	4268
7	Canada	24.4 (4.1)	7436	-	-	21.7 (4.6)	7436	-	-
8	Scotland	20.1 (6.1)	3089	18.5 (6.9)	2588	12.9 (6.9)	3089	7.1 (8.0)	2588
9	Portugal	17.7 (7.0)	2310	16.6 (7.1)	2185	17.9 (8.0)	2310	18.0 (8.0)	2185
10	England	13.5 (5.9)	3087	-	-	12.9 (6.9)	3087	-	-

Table A.6b: Reduced form estimates of 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 forcing 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 (5.2)	3567	25.6 (5.2)	3502	55.9 (6.8)	3567	53.9 (6.7)	3502
2	Sweden	19.1 (6.0)	3403	16.3 (6.1)	3235	25.2 (7.1)	3403	22.3 (7.2)	3235
3	Italy	18.5 (7.9)	2186	15.8 (8.3)	2142	15.0 (7.2)	2186	12.0 (7.6)	2142
4	Norway	18.2 (6.0)	2751	-	-	22.1 (6.3)	2751		-
5	Spain	13.6 (5.8)	3157	13.7 (5.9)	3084	10.9 (5.3)	1834	10.4 (5.3)	3084
6	Iceland	13.5 (6.3)	1834	13.0 (6.6)	1641	17.1 (7.2)	2824	14.1 (7.3)	1641
7	Scotland	9.9 (5.9)	2824	8.8 (6.2)	2451	12.7 (6.5)	2837	8.9 (6.5)	2451
8	Cyprus	9.5 (7.1)	2837	4.2 (6.9)	2764	14.7 (7.2)	3475	9.7 (6.8)	2764
9	Slovak Republic	8.6 (5.5)	3475	8.5 (5.5)	3385	11.9 (5.6)	5158	9.6 (5.4)	3385
10	Japan	4.8 (4.4)	5158	3.0 (4.5)	5075	14.1 (4.1)	3543	13.4 (4.1)	5075
11	Greece	3.0 (5.6)	3543	13.0 (10.1)	1484	5.9 (5.5)	2622	11.1 (10.0)	1484
12	Belgium (Flemish)	2.8 (4.8)	2622	3.8 (5.0)	2479	12.2 (4.8)	2159	12.8 (4.9)	2479
13	England	-9.0 (9.1)	1834	-9.2 (10.2)	1635	1.8 (9.7)	1834	0.3 (11.1)	1635