6

Does Early Educational Tracking Increase Inequality? Short and Long Term International Evidence[1]

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

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

6.1 Introduction

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

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

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

Puhani and Weber, 2006; Jurges and Schneider, 2007; Muhlenweg and Puhani, 2010; Dustmann *et al.*, 2014). Given that age differences at the start of a tracked school system are relatively larger when countries track earlier, because a difference of 11 months in age at start is relatively larger for 10-year-olds than for 15-year-olds, the effects described above are therefore likely to be even larger in countries that track at an earlier age. Subsequently we expect that in countries with early tracking relatively young students more often enroll in the low track than in countries with late tracking.

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

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

6.2 Tracking and relative age

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

6.2.1 Relative age effects

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

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

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

6.2.2 Early tracking

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

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

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

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

Waldinger (2006) looks at the effect of parental background on student performance using a similar difference in difference approach. Using individual level data, he finds no effect of an interaction between early tracking and parental background once elementary school inequality is taking into account.

6.2.3 Relative age effects and tracking

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

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

6.3 Empirical strategy

In our empirical strategy we exploit exogenous variation in the effect of relative age, which itself is caused by exogenous variation in MOB and school cut off dates, across two groups of countries: those that track students in secondary school at an early age and those that track students late. This difference in relative age effects across these two groups of countries leads to exogenous variation in the probability to go to the low track for the relatively young students. Our empirical strategy consists of two parts. First, we focus on short term differences in relative age effects between countries with early and countries with late tracking, both in terms of track placements and student performance. Next, we use within country variation in the age of tracking to obtain estimates of early tracking on long term outcomes. We do this by exploiting changes across cohorts in the age of tracking that occurred within countries in the period between 1939 and 2012.

6.3.1 Conceptual framework

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

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

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

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

In the first part of our empirical strategy we show that, due to exogenous variation in MOB and school starting cut off dates, relatively young students are more likely to go to the lower track, as is consistent with the literature. But, more importantly, we show that relatively young students in countries that track early are even more likely to go to the lower track than relatively old students than they would in countries that track later. Following this we also look at performance differences after tracking has taken place. For our short term outcomes we estimate models similar to difference-in-differences models. The first difference is the difference in performance between relatively young and relatively old students. The second difference is between early tracking and late tracking countries. Equation (3) shows the short run models in which Y_{ic} either is a dummy for being in the low track for student i in country c or the PISA math test score for student i in country c.

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

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

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

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

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

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

6.4 Data: PISA and PIAAC

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

6.4.1 PISA

PISA 2009 contains information on 15 year old students and their schools from all 36 OECD countries.[2] Per country a representative sample of schools is selected and all the students in that school of the relevant age group participate in PISA. Students make tests in reading, mathematics, and science, and fill in a survey on their home situation, school work and other related issues. Subsequently for each student a large number of background characteristics are known, for instance age, gender, parental background, and the country specific track he or she is at age 15. Besides student background characteristics, from each student we have their test scores on mathematics, reading, and science.

Since we are interested in the difference in the relative age effect on track placement between early and late tracking countries, we only keep those countries in which students are placed into tracks at the age we observe them (age 15). This means that we delete the countries that track after the PISA tests are administered (the non-tracking countries). This leaves us with 160,733 students in 20 countries.[3] Table 1 gives an overview of these countries. The age of tracking is obtained from the OECD

2. We restrict ourselves to OECD countries since the data from non-OECD countries often showed unclear school cut off dates and because our two methods for determining which national tracks can be considered low and high tracks often contradicted each other for these countries.
3. We deleted 11 students who had an unknown country specific track and students in tracking countries who were not tracked yet due to grade repetition at the time of the PISA tests (293 students).

and Eurydice (see Table 1) and differs across countries ranging from age 10 for Germany and Austria, to 15 for France, Portugal, and four other countries. For our difference in difference strategy we use three options for the definition of an early tracking country: A country is an early tracking country if tracking takes place before the age of 12, 13, or 14. When the cut off is placed at age 12, only 6 countries from the 20 are early tracking countries; when we use age 13 as a cut off the sample is balanced with 8 countries before the cut off and 12 after; with age 14 there are 9 early tracking countries and 11 late tracking countries.

Table 1: Descriptive statistics of the PISA sample

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

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

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

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

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

Source: Own calculations with PISA 2009 data.

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

Webbink (2013).[4] We define relative age in such a way that the youngest in each cohort have a value 1 and the oldest have a value 12.

6.4.2 PIAAC

To look at long run outcomes, we use data on adults taken from the Program for International Assessment of Adults Competencies (PIAAC). PIAAC is an international survey among the adult populations of 24 OECD countries. The survey collects test score data on literacy, numeracy and problem solving, and on the education and work history of the participants and other life outcomes. The sample consists for each countries of a representative sample of the entire working age population age 16 to age 65. We restrict our sample to males age 25 to 50. This leaves us with 38,809 participants in 21 countries.[5] PIAAC contains a wide variety of questions on education and work history of the participants, but unfortunately does not contain the secondary school track the individuals attended. For ease of interpretation of our difference in difference analyses, we construct dummy variables on education and employment. Descriptive statistics on all variables are displayed in Table 3. It displays a subset of the available variables in PIAAC and groups them into education, employment, and earnings outcomes. Education outcomes are available for the largest number of observations (31,585-46,519 males) and contain a dummy for whether or not the respondent obtained a tertiary education degree and the age at which respondent's left formal education. The employment variables contain whether or not the respondent was employed, whether he was self-employed, whether the job he was in required a tertiary education degree and whether he has a supervisory role in his job or not. Limited earnings information was available for 26,918 males and this was transformed into a dummy for earning above the mean, above the mean of the own age group and belonging to lowest ten percent earners by age group.

For our causal estimates on whether relative age differences in life outcomes are larger in early tracking countries, we supplement the PIAAC data with time changing data on countries age of tracking. We use this panel data on tracking to categorize early tracking countries separately for each cohort, as some countries have changed their tracking regime over time. This time changing data is collected by Braga, Checci, Meschi (2013) and was generously shared by the authors for this chapter. The dataset contains time changing data from 1929 to 2000 on a number of education

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

system characteristics and shows quite some changes over time. For instance, Flanders changed the age of first selection from age 14 to age 15.5 in 1971 and to age 12 in 1985 and England changed from age 14 to age 11 in 1944 and again to age 16 in 1966. For students entering the tracked school system after 2000, we use the age of first selection as used in the PISA analyses as described above. We allocate the age of first selection into tracks to each participant in which he/she entered the tracked system. For example, using the example of Flanders also mentioned above, a person born in 1956 in Flanders was tracked at age 14, while a person born the year later in 1957 was tracked at age 15.5. Among the 14 countries, 17 times countries increased the age of tracking, while the age was decreased 5 times. Nine countries actually moved from an early to a late tracking country or vice versa when we define early tracking as tracking before the age of 13. The school starting months in this long run dataset are taken from the same sources as described for the short run analyses.

Table 3: PIAAC descriptive statistics

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

Notes: Males aged 25 to 50 in 14 countries.

6.5 Results

6.5.1 Short run results

Tracking

Previous studies using German, French, and Canadian data have shown that relatively young students are more likely to go to the lower track. We find the same pattern using cross country data, as illustrated by Figure 1 and Table 4. Figure 1 shows the link between relative age and the probability of enrolment in a low ability track for early tracking countries (tracking occurs before the age of 13) and late tracking countries (tracking occurs on or after the age of 13). We observe in both types of countries that the probability of enrolment in the low ability track is higher when students are relatively younger as compared to those students who were born just after the school starting cut off. However, and most important, this difference in the probability to go to the low track for young and old students is much higher in early tracking countries.

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

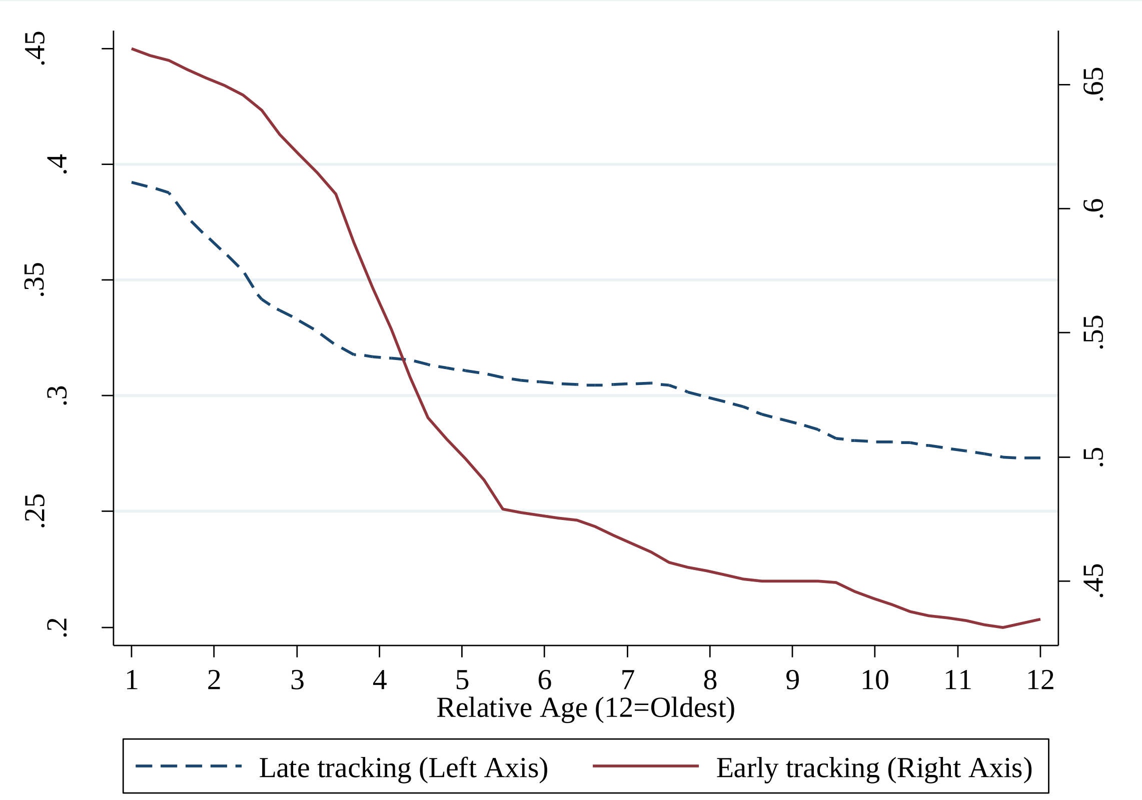


Table 4 shows this result more formally, and estimates equation (3). For three different definitions of early tracking, based on the age threshold, being in the low track is related to relative age and early tracking. As can be seen from the table, the relatively oldest are less likely to be in the low track, as is consistent with the literature. A new insight however is that Table 4 also shows that probability to be in the low track is different for students in countries with early tracking as compared to students in countries with late tracking. When the definition of early tracking is tracking before the age of 13 or 14 (the two right columns of Table 4), relatively older students are even less likely to go to the low track. Or in other words, relatively young students are more likely to go to the low track and even more so in a country that tracks before the age of 13 or 14. Controlling for parental background does not change these results, although students with higher parental background are less likely to go to the low track for any relative age (not shown). In this chapter we exploit the exogenously induced difference between early and late tracking countries for investigating the short term and long term consequences of (early) tracking. From here on we use the threshold at 13 since for this threshold the number of countries before and after the cut off is more balanced (8 before, 12 after) and since we are mainly interested in the largest difference in the probability to go to the lower track and we aim to show the consequences of this difference.[6]

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

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

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

Math scores

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

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

in countries that track early, they do perform better relative to their counterparts in countries that track late.

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

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

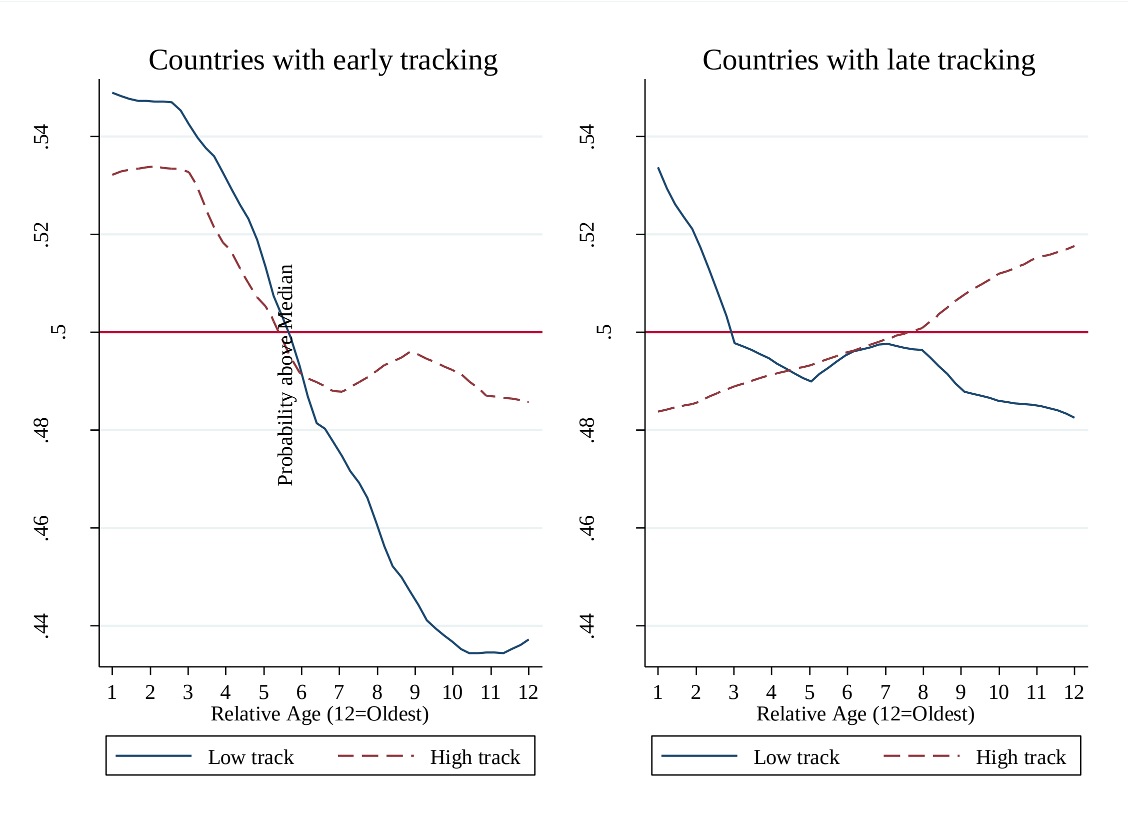
Notes: The table presents coefficients (standard errors in parenthesis) from fixed effect models, as depicted in equation (3), using individual weights provided by PISA. The superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Early tracking countries are those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

6.5.2 Mechanism

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

Figure 2 illustrates this principle. Figure 2 shows the probability to have a PISA mathematics test score above the median score within each track over relative age and for students in countries that track early and those that track late. In countries that track late, the relatively young students hardly outperform the relatively old students in the low track, while the relatively old students hardly outperform the young in the high track. If we turn to students in countries that track early we see that in the high track, the relatively young do slightly better than the relatively old, similarly to the pattern we saw in countries that track late. However, we also see that the relatively young perform more often above the median than the relatively old students in the low track.

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



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

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

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

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

6.5.3 Long run effects

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

Second, the relatively young students turn out to be more able than the relatively old students in the lower track in countries that track early. When in countries that track early students are placed into tracks, the observed performance of the relatively young students is a biased measure of their ability due to the relative age bias. Their ability level is actually higher than observed at the time, but this bias will disappear over time. When this bias has disappears and the true ability of the relatively young students is mirrored in increased performance, the relatively young students turn out to be the best in their class and even in their track. This might motivate the students and cause even better performance resulting in a positive effect of being in the lower track. The effect running through being the best of the track could even cancel out the lower initial attainment and explain the findings of Muehlenweg and Puhani (2010) and Dustmann *et al.* (2014) who find no long term relative age effect in Germany due to track mobility.

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

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

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

R^2	0.05	0.04	0.89	0.90	0.18
Country*cohort clusters	377	377	371	371	371

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

Table 7 displays the causal effects of relative age and early tracking on a number of long run outcomes for males between the age of 25 and 50. We make use of within country variation on both relative age and age of tracking. The within country variation on age of tracking creates changes in the composition of the group of countries who track early or late for some cohorts. These analyses are therefore not confounded by time-invariant systematic differences between countries that track early or those that track late. Table 7 shows that we do not find a significant effect of relative age and early tracking on educational attainment (column 1) or employment (column 2), but we do find effects on earnings at later ages (columns 3-5). Thus relative age and early tracking do not seem to matter for obtaining a tertiary degree and for being employed, but do matter for earnings at later ages. In general, relatively old students have a small earnings advantage compared to the relatively young: The oldest are one percent more likely to have earning above the average. But column 3 also shows that being a relatively old student in school and being tracked early means that at later ages it is more likely to have earnings below the average, as compared to peers in a late tracking country: with early tracking those who were relatively old when in school are two percent less likely to have earning above the average than relatively old students in late tracking countries.

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

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

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

6.6 Conclusions

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

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

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

The advantage for the relatively young in countries with early tracking means a disadvantage for the relatively old students in the lower track. These students are grouped together with young students who turn out to be more able than them and the young outperform them at all stages in life: early cognitive outcomes and labor market outcomes. Summarizing, we can say that the relative age effect which caused more relatively young students to be placed into the lower track has long lasting effects. Although the relative age effect has not disappeared in late tracking countries, it is much smaller than the effect in early tracking countries. Early tracking therefore induces an inequality in outcomes based on month of birth.

References

Angrist, J. and A. Krueger (1991) Does Compulsory School Attendance Affect Schooling and Earnings? *The Quarterly Journal of Economics,* 106, pp. 979-1014.

Allen, J. and R. Barsnley (1993) The Interaction of Ability, Maturity, and Training in Systems with Age-Dependent Recursive Selection. *The Journal of Human Resources*, 28(3), pp. 649-659.

Ariga, K., and G. Brunello (2007) Does Secondary School Tracking Affect Performance? Evidence from IALS. *IZA Discussion Paper* 2643.

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

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

Braga, M., D. Checci, E. Meschi (2013) Educational policies in a long-run perspective. *Economic policy*, 28(73), pp. 45-100.

Crawford, C., L. Dearden and C. Meghir (2013) When You Are Born Matters: Evidence for England. *IFS Report* R80. London: Institute for Fiscal Studies.

Cunha, F., J. Heckman, L. Lochner, and D. Masterov (2005) Interpreting the Evidence on Life Cycle Skill Formation. *NBER Working Paper* No. 11331.

Duflo E., P. Dupas, and M. Kremer (2011) Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *American Economic Review*, 101, pp. 1739-1774.

Dustmann, C., P. Puhani, and U. Schönberg (2014) The Long-Term Effects of Early Track Choice, *IZA Discussion Papers* 7897.

Fredriksson, P. and B. Oeckert (2013) Life-Cycle Effects of Age at School Start. *The Economic Journal*, 124, pp. 997-1004.

Fuchs, T. and L. Woessmann (2007) What accounts for international differences in student performance? A re-examination using PISA data. *Empirical Economics*, 32, pp. 433- 464.

Gerristen, S. and D. Webbink (2013) How much do children learn in school? International evidence of the effects of time in school. CPB Discussion Paper 255. Den Haag: Netherlands Bureau for Economic Policy Analaysis (CPB).

Grenet, J. (2010) La date de naissance influence-t-elle les trajectoires scolaires et professionnelles? Une évaluation sur données françaises *Revue économique,* 60(3), pp. 589-598.

Hanushek, E. and L. Woessmann (2006) Does Educational Tracking Affect Performance and Inequality? Difference-in-Difference Evidence Across Countries. *Economic Journal,* 116, pp. C63-C76.

Jakubowski, M. (2010) Institutional tracking and achievement growth: Exploring difference-in-differences approach to PIRLS, TIMSS and PISA data. In: *Quality and Inequality of Education* (2010), pp 41-81.

Jurges, H. and K. Schneider (2007) What Can Go Wrong Will Go Wrong: Birthday Effects and Early Tracking in the German School System. *CESifo working papers*, nr. 2055. CESifo.

Muhlenweg, A. and P. Puhani (2010) The Evolution of the School-Entry Age Effect in a School Tracking System. *The Journal of Human Resources*, 45(2), pp. 407-438.

OECD (2010) *School Factors Related to Quality and Equity. Results from PISA 2000.* OECD: Paris.

Kerr, S., T. Pekkarinen and R. Uusitalo (2013) School Tracking and Development of Cognitive Skills. *Journal of Labor Economics*, 31(3), pp. 577-602.

OECD (2003) *Literacy Skills for the World of Tomorrow - Further results from PISA 2000.* OECD: Paris.

OECD (2007) *Science competenties for tomorrows world.* OECD: Paris.

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

Pekkarinen, T., R. Uusitalo and S. Kerr (2009) School tracking and intergenerational income mobility: Evidence from the Finnish comprehensive school reform. *Journal of Public Economics*, 93, pp. 965–973.

Puhani, P. and A. Weber (2006) Does the early bird catch the worm? Instrumental variable estimates of early educational effects of age of school entry in Germany. *Empirical Economics*, 32, pp.359–386.

Stipek, D. (2002) At what age should children enter kindergarten? A question for policy makers and parents. *Social Policy Report*, 16(1).

Van Elk, R., M. van der Steeg and D. Webbink (2011) Does the Timing of Tracking Affect Higher Education Completion? *Economics of Education Review,* 30(5), pp.1009- 1021.

Walldinger, F. (2006) Does Tracking Affect the Importance of Family Background on Students' Test Scores? *Working paper.* www.cepr.org/meets/wkcn/3/3527/papers/Waldinger.pdf, 05-12-2012.