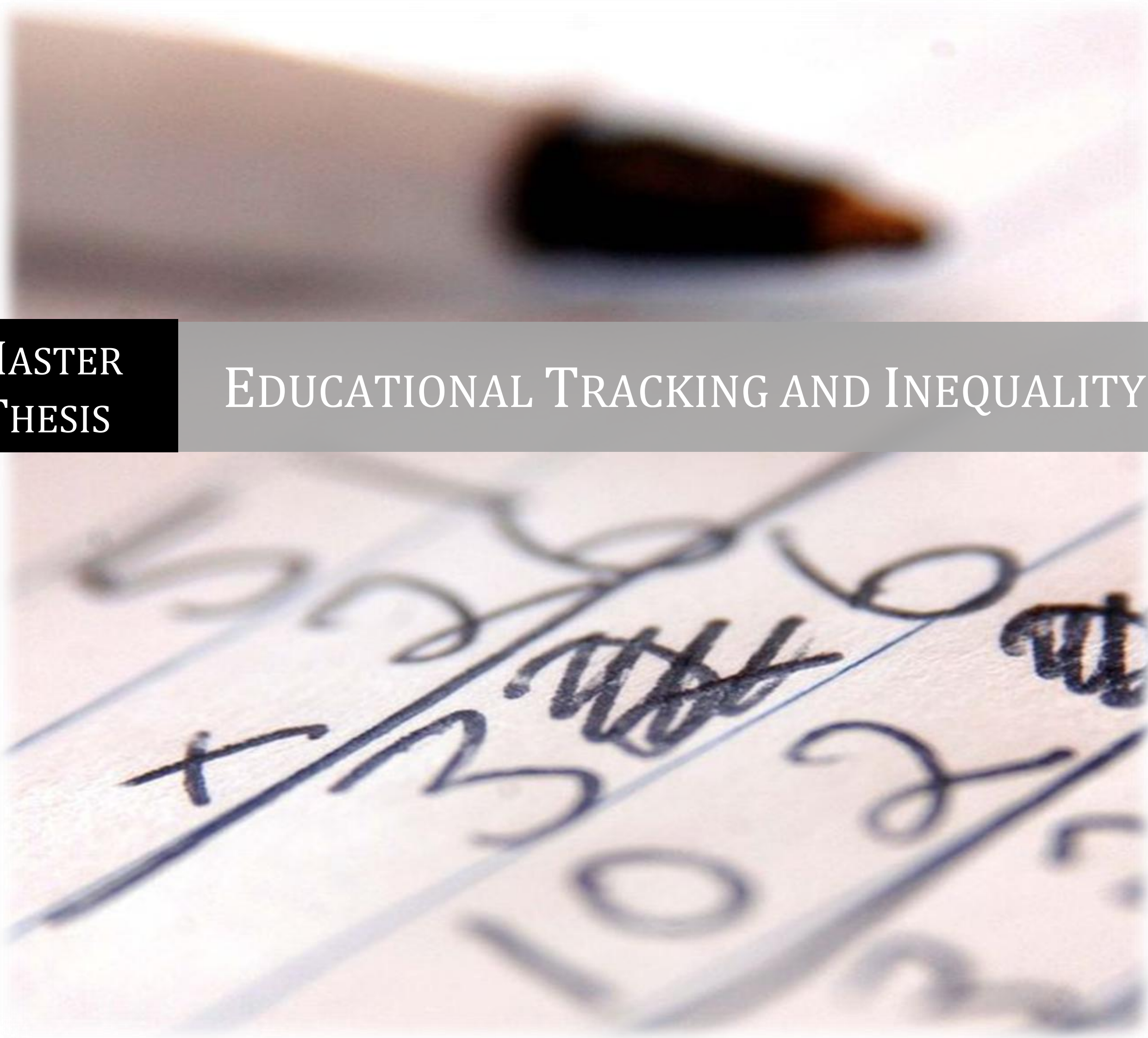
Master Thesis

Educational Tracking and Inequality



Author
Sissy Chen
319847

Supervisor
Prof. dr. Dinand Webbink

Co-reader
Dr. Hans van Kippersluis

Econometrics and Management Science
Erasmus University Rotterdam

January 2013

Abstract

Educational tracking is not internationally implemented, as its merits and demerits are disputed. This research analyzes the differences between tracked and untracked students regarding a novel measure of education quality, namely the cognitive value gained by the students within one year of school attendance. I employ quantile regression to allow differences caused by tracking to vary in different levels of cognitive ability, and validate the differences between these levels using interquantile regression. The findings show remarkable differences between the top and bottom 10% of students, and suggest that tracking significantly enlarges inequality within students' performances, in the sense that top students profit from being in a tracked system, whereas bottom students are better off in an untracked system.

Acknowledgements

I would like to express my gratitude towards my supervisor Prof. dr. Dinand Webbink, for his guidance and support during the process of this research and the writing of my master thesis.

1 Introduction

The impact of educational tracking is a much debated subject which has been investigated in many studies. Educational tracking refers to the practice of separating students into certain levels of curricula based on cognitive ability. There are many disputes about the merits and demerits of tracking. Proponents of tracking argue that homogeneous classrooms permit appropriately paced instruction, which maximizes learning efficiency. This way the fast students will not be held back by the pace of fellow students with lower learning ability. Proponents of untracked school systems argue that homogeneous classrooms have a great disadvantage for slow students, as they are not pushed forward by the so-called peer effects, but left to fall further and further behind the students with high learning ability. This also implies that tracking will not only allow the existing inequality between students, but also enlarge it. However, the empirical evidence for these arguments is still very limited.

It is statistically quite difficult to estimate the impact of educational tracking. In order to obtain estimates which are reliable and statistically valid, variation is an important matter. However, usually there exists no variation in the education system within a country, as the country's policy deciding whether educational tracking will be applied in the school system usually applies for all schools nationwide. There are previous studies which were able to bypass this statistical difficulty. Duflo et al. (2011) performed a randomized evaluation of tracking in schools in Kenya, by conducting an experiment where they apply different policies concerning tracking to a number of schools. They find a higher average score among students in tracked schools after 18 months, and the effect persisted up to one year after the program ended. Also, they report that students at all levels of education in the tracked schools benefited from the tracking system. Elk et al. (2011) exploited variation within the Dutch education system to investigate the effect of the timing of educational tracking on completion of higher education. This variation is based on a division of comprehensive schools which postpone tracking by one or two years, while the other schools implement tracking right from the start of secondary school. They find negative effects for schools which implement educational tracking earlier, meaning that students who were tracked early in their educational course are less likely to complete higher education. These studies provide quite contradicting results. This might come from differences between Kenya and the Netherlands. Also, because these studies are nationwide, the implemented models probably differ in many ways, such as choice of control variables, choice of instruments, choice of dependent variable, the applied scale of dependent variable, and so on.

In this study, I aim to investigate the effect of tracking on education quality at an international level. This way, the comparison is more general as all countries are treated similarly. There have been many previous studies considering the issue from the same perspective. Amongst others, Hanushek and Woessmann (2006) investigated the effect of tracking on test scores using a differences-in-differences (DD) approach and cross-country comparisons. A DD-model is suitable for the effect of any given cause variable. The cause variable is a binary indicator included in the model as regressor. The effect of the cause variable is estimated, given the common trend assumption. In the DD-model employed by Hanushek and Woessmann (2006), the effect of tracking on inequality (standard deviation) and mean performance is estimated through an indicator for the existence of tracking. The chosen dependent variables are aggregated from country specific samples, which means the variables have one value per country. They find that tracking increases educational inequality, in the sense that in tracked systems, the standard deviation among students' performances is larger. Also, their findings suggest that tracking students from the beginning of secondary school reduces the mean performance of the country. However, in this approach the estimation of the tracking effect can only be based on the variation between countries, and the statistical power of the evidence is limited, as it is very hard to find a suitable dataset containing a large number of countries.

Education quality can be defined in many ways. The most straightforward estimator of education quality are students' test scores, which is what the majority of previous studies have implemented. However, this requires the assumption that the students' performances are solely based on the quality of education they receive from school attendance, which is obviously not likely to be the case. The students with high test scores could very well be influenced by, for example, ambitious parents, whereas the students with low test scores may be dealing with certain home situations hindering them to study properly. Also, the students are all different individuals with different characteristics, which can also determine a great deal of their performance in school. A better way to measure education quality is to focus on the cognitive achievement of students, i.e. the educational value added to the student through one year of school. The aim of this study is to investigate the effect of tracking on the cognitive achievement of students in one year of school attendance. Cognitive achievement will be obtained from a regression discontinuity model. This novel approach allows comparison of the effect of time in school, isolated from the effect of time outside of school. This study contributes to the literature by investigating the effect of tracking on the value added of the education system. Comparing results across countries worldwide will give us insight into the differences between tracked and untracked education systems.

The findings of this study suggest differences between tracked and untracked students. Moreover, using quantile regression techniques, I also find that tracking significantly enlarges inequality in students' school performances, as the effect of tracking differs among students at different cognitive ability levels. The top 10% of students benefit from being in a tracked education system, whereas the bottom 10% of students perform better in untracked education systems.

2 Data

In this study, I will use data from the TIMSS test conducted in 1995, with a sample of more than 400.000 students across more than 40 countries all over the world, and tests in the two subjects: mathematics and science. The sample includes populations of 3 different ages: age 9, age 13, and age 18. In order to measure the most direct effect of tracking, this study focuses on tracking implemented in early stages of education, and therefore will only make use of the first two populations of students, where the 9-year-olds are not (yet) tracked and the 13-year-olds are tracked in certain countries. Choosing the TIMSS 1995 test scores is based on the unique principle of the selection of students, which is based on chronological age instead of grade of curriculum. This means that the 9-year-old sample consists of the two grades containing the most 9-year-old students, idem principle for the 13-year-old sample. This selection principle was not employed in other editions of the TIMSS, and therefore makes the TIMSS 1995 database unique. Moreover, this feature of the data is essential for analysis concerning cognitive achievement (this will be explained further in Section 3.1).

In order to obtain comparable populations for the 9-year-olds and 13-year-olds, I exclude countries which have only participated in either one of the two populations, leaving in the sample the following 23 countries: Australia, Austria, Canada, Cyprus, Czech Republic, Greece, Hong Kong, Hungary, Iceland, Iran, Ireland, Japan, Korea, Latvia, the Netherlands, New Zealand, Norway, Portugal, Scotland, Singapore, Slovenia, Thailand, and United Kingdom. With these countries in the sample, I have a total of 165.594 students in the 9-year-old population, and 168.229 students in the 13-year-old population, with respectively 23.053 students and 22.520 students in tracked school systems. The main variable of interest are the test scores[1], which can be directly extracted from the dataset. I have standardized these in order to prevent interpretational difficulties. In the set of regressors, I include a variable controlling for chronological age, serving as assignment variable which is needed in a regression discontinuity model (further explained in following sections). Also, I use a collection of country dummy variables to control for country fixed effects. Lastly, I use a selection of personal variables to control for individual differences: sex, the number of books the student possesses, and indicators for whether the country in which the test is taken is the student's place of birth, whether the language in which the test is taken is the student's mother tongue, and whether the student lives with his/her mother and/or father. Missing values have been imputed with the country's sample mean.

[1] I use the first plausible value of the test scores in the dataset.

3 Empirical strategy

3.1 Cognitive Value

As introduced in Section 1, this study will use the students' cognitive achievement in one year of school attendance as indicator for education quality. In other words, I measure education quality by the educational value added to the student by the education system. There are various, but not all efficient ways to obtain this value added.

A very straightforward approach is to use repeated observations of students' test scores, and analyze the change in time. The adequacy of this approach is, however, criticized by the study by Rothstein (2010), in which this model setup is used to assess teacher quality. The findings state that the assumptions needed for causal interpretations of the model's estimates, including random assignment, are violated. He suggests incorporating information about the actual assignment process. This is done earlier in the study by Cahan and Cohen (1989). They investigated the effect of time spent in school, as opposed to chronological age, on intelligence development. This research was conducted using a dataset of students in grade 4 to grade 6 in Israel. In Israel, children born in the same calendar year start school at the same time. This fact can be exploited to estimate the value added in one year of school attendance, as children born in adjacent months around the time of the calendar year cut-off point differ one year of time spent in school. Using the test scores from these students, they estimate the effect of both chronological age and time spent in school. The findings suggest that one year of school attendance adds approximately twice the value of one year in age.

I will use this design to obtain the value added to the student through one year in school, and compare the results across the tracked and untracked countries in order to retrieve the effect of tracking on the value added. Consider the following simple regression model, which will be referred to as Model 1:

$$y_i = \alpha_0 + \alpha_1 S_i + \boldsymbol{X}_i\boldsymbol{\beta} + f(age) + \boldsymbol{C}_i\boldsymbol{\gamma} + \varepsilon_i, \tag{1}$$

where the test scores of each individual student (y_i) is regressed on S_i, an indicator for whether the student i was born before or after the cut-off date of his/her country, and control variables X, country dummies C, and the chronological age. This regression discontinuity model (RD) is elaborated extensively in the work of Lee and Lemieux (2010). The chronological age function serves as continuous assignment variable for the discontinuous variable S. In this setup, the estimated value of α_1 is the value added. This is the effect of one year of time in school. By performing this model separately for the populations at age 9 and 13 in tracked and untracked countries, I am able to observe any differences in the estimated value added directly.

Table 1 contains the list of countries, along with the cut-off date for school enrollment, and the age at which tracking is first applied in the country. The estimate of α_1 is based on the variable S_i, which is constructed using the cut-off date information provided in this table combined with student i's birthday information.

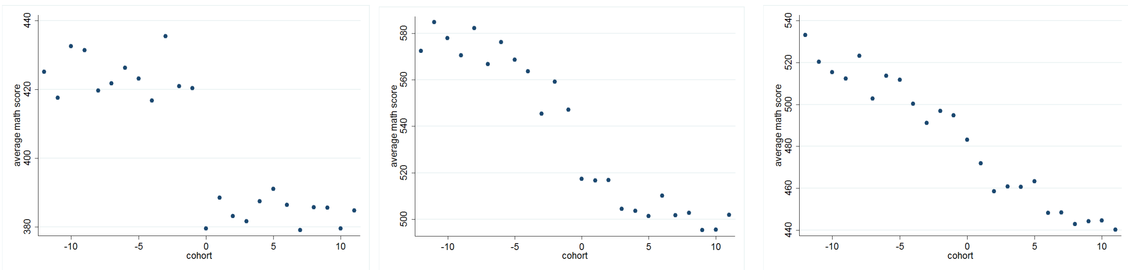
Table 1: Cut-off dates and age of first tracking per country.

COUNTRY	CUT OFF DATE	AGE OF TRACKING
Australia	1-Jan	16
Austria*	1-Sep	10
Canada	1-Jan	16
Cyprus	1-Mar	>14
Czech Republic*	1-Sep	11
Greece	1-Apr	15
Hong Kong	1-Jan	15
Hungary*	1-Jun	11
Iceland	1-Jan	16
Iran	1-Oct	>14
Ireland	1-Sep	15
Japan	1-Apr	14
Korea	1-Mar	14
Latvia	1-Sep	16
Netherlands*	1-Oct	12
New Zealand	1-May	16
Norway	1-Jan	16
Portugal	1-Jan	15
Scotland	1-Jan	>14
Singapore	1-Jan	>14
Slovenia	1-Jan	15
Thailand	1-Oct	>14
United Kingdom	1-Sep	16

The analysis will only include students with birthdays deviating up to one year from the cut-off date. Because I focus on the effect of early tracking, the sample of tracked countries will only contain countries which apply tracking before the age of 13 (the age of the second test population in the TIMSS database), that is Austria, Czech Republic, Hungary, and the Netherlands. These countries are marked with an asterisk in Table 1.

In order to shortly elaborate on the definition of the cognitive value added, I would like to visualize the value added based on maths scores of three randomly chosen countries in my sample. In Figure 1, maths scores are plotted against birth months. Students are grouped into cohorts according to the month they are born in. The values on the x-axis represent the cohorts' deviation in months from the cut-off date for school enrollment. In the ideal case (when the education system, or school attendance, is responsible for 100% of all educational contribution to the students), we would observe a sharp vertical drop at the cut-off (cohort 0). Figure 1a represents the discontinuity within Iranian students. The scores at the cut-off quite sharply drop about 30 points for students in the lower grade. If we look at the discontinuity in the Dutch sample (Figure 1b), the gap is substantially smaller. And if we look at the British sample of students (Figure 1c), we observe a quite linear decrease in scores, without any gap. Based on the test scores, one might have concluded that the education in United Kingdom is better than the education in Iran, as the test scores in United Kingdom are higher than the test scores in Iran. But based on the difference in performance of two adjacent grades of students, which eliminates cognitive value gained through other sources than school attendance, we would suggest that the British education system is less efficient than the Iranian education system.

Figure 1: Discontinuity in the average maths score across birth dates around the cut-off date of school enrollment for the countries Iran, the Netherlands, and United Kingdom.



(a) Iran. (b) The Netherlands. (c) United Kingdom.

Notes: Average maths scores of the subsample containing 9-year-old students are plotted against birth dates. Students are grouped into cohorts according to the month they are born in. The values on the x-axis represent the deviation in months from the cut-off date for school enrollment, i.e. cohort -10 contains students born 10 months before the cut-off date, and cohort 10 contains students born 10 months after the cut-off. Similar graphs can be obtained for each country within the analysis sample upon request.

Using this value added approach, I have a number of advantages over earlier studies on this subject. The most distinguishing advantage is being able to isolate the cognitive value added to the students through time spent in school from the cognitive value added to the students through other paths than school. Also, I can be sure of the statistical power of my analysis, because I will run the models on individual student level.

3.2 The effect of tracking

In order to be able to validate any observed differences identified using Model 1, we need to use interaction effects. Consider Model 2:

$$y_i = \alpha_0 + \alpha_1 S_i + \alpha_2 T_i + \alpha_3 S_i T_i + \boldsymbol{X}_i \boldsymbol{\beta} + f(age) + \boldsymbol{C}_i \boldsymbol{\gamma} + \varepsilon_i, \tag{2}$$

where S_i is the indicator for whether the student i was born before or after the cut-off date, T_i is the indicator for whether tracking is induced in the student's country before the age of 13. Along with these indicators, the same set of control variables as in Equation 1 is used. We are mostly interested in the estimation of α_3, as this is the coefficient of interaction term S_iT_i. This means that α_3 represents the effect tracking has on the value added.

3.3 Quantile regression

The simple RD-models as suggested in Section 3.1 and 3.2 will provide a broad overview of all students, but merely that. Because educational tracking divides the students into levels based on cognitive ability, it seems rather shortcoming to treat students with different learning abilities the same way. A more suitable regression model is the quantile regression model, as introduced by Koenker and Bassett (1978). I will shortly explain the basic estimation technique of quantile regression.

Using quantile regression, a conditional quantile function (CQF) is needed, much like the conditional expectation function (CEF) in "ordinary" linear regression. In quantile regression, the focus is on describing some given quantile (or percentile) of the data, instead of the conditional mean. Consider the CQF of continiously-distributed random variable Y at quantile τ given regressors $\boldsymbol{X}$:

$$Q_\tau(Y|\boldsymbol{X}) = F_Y^{-1}(\tau|\boldsymbol{X}),$$

where $F_Y(y|\boldsymbol{X})$ is the distribution function for Y conditional on $\boldsymbol{X}$. Now, suppose $\boldsymbol{Z\theta} = \alpha_0 + \alpha_1 S_i + \boldsymbol{X}_i\boldsymbol{\beta} + \delta f(age) + \boldsymbol{C}_i\boldsymbol{\gamma}$ (see Equation 1). In this case, the CQF becomes $Q_\tau(Y|\boldsymbol{Z})$. Consider the minimization problem in Equation 3, where $\rho_\tau(u) = (\tau - 1(u \leq 0))u$ is the so-called check function, and $q(\boldsymbol{Z})$ is the τ-th percentile of the control variables $\boldsymbol{Z}$.

$$Q_\tau(Y|\boldsymbol{Z}) = \arg\min_{q(Z)} E[\rho_\tau(Y - q(\boldsymbol{Z}))] \tag{3}$$

In this minimization, the check function $\rho_\tau(u)$ plays the role of weighing positive and negative terms asymmetrically given τ. By substituting a linear model for $q(\boldsymbol{Z})$, we obtain

$$\theta_\tau \equiv \arg\min_{t\epsilon R^d} E[\rho_\tau(Y - \boldsymbol{Zt})]. \tag{4}$$

The quantile regression estimator $\hat{\theta}_\tau$ is the sample analog of Equation 4. Using quantile regression in the simple model (Equation 1) will not only provide insight into differences and changes in value added, but also show the differences and changes within each quantile of scores. This allows for differentiation in the effect of tracking amongst students with different learning ability. In this study, I will use $\tau = \{.1, .25, .5, .75, .9\}$.

3.4 Validation with interquantile regression

In order to validate differences caused by tracking between students of different potential, I wish to compare the difference in value added between the students at $\tau = .1$ and $\tau = .9$. This can be done quite straightforwardly by modifying the quantile regression model, and "upgrade" the model to an interquantile regression model. As the model name already suggests, this model estimates the interquantile range. In other words, it evaluates the difference in quantiles. Recall the CQF $Q_\tau(Y|\boldsymbol{Z})$ described in Section 3.3. We obtained $Q_{.1}(Y|\boldsymbol{Z})$ and $Q_{.9}(Y|\boldsymbol{Z})$ using quantile regression. The difference between these percentiles can be described as the following:

$$Q_{.9}(Y|\boldsymbol{Z}) - Q_{.1}(Y|\boldsymbol{Z}) = \boldsymbol{Z}(\boldsymbol{\theta}_{\tau=.9} - \boldsymbol{\theta}_{\tau=.1}). \tag{5}$$

Interquantile regression estimates the coefficients exactly as shown, namely as the difference in coefficients of two quantile regression models. The standard errors are obtained appropriately by bootstrapping. Applying interquantile regression to Model 2 gives us our third estimation model, Model 3, which will provide the estimations of difference in value added between the top- and bottom 10% of students, caused by tracking.

4 Estimation Results

In the following result tables, estimations of all τ's are presented along with the mean estimate, which is the ordinary least squares (OLS) estimate. In Section 4.1, I present the value added estimates obtained using the RD-model. In Section 4.2, estimation results of the interaction model in Equation 2 are presented, providing insight into differences in value added between tracked and untracked students. And finally, in Section 4.3, the results of the interquantile regression model are presented, validating the differences between different levels of students. I will simply refer to the differences between tracked and untracked students as the effect of tracking, but it is good to note that all differences measured between tracked and untracked students should be appropriately interpreted as the difference between tracked and untracked education systems, as it is very likely that tracked and untracked school systems differ in many ways besides the implementation of tracking, which could also cause differences in value added between students in tracked and untracked systems.

4.1 Value added estimations

First, I present the results of Model 1. The estimations of the effect of school attendance in tracked and untracked countries are shown in Tabel 2. We can see immediately that if we compare the population of age 9 with the population of age 13, we observe a strong decrease in value added. 13-year-olds seem to gain less educational value from time in school than 9-year-olds. We observe that for the 9-year-old population, there seems to be no striking differences between tracked and untracked students in the structure of the estimations. In the 13-year-old population, however, we observe clear differences. If we measure the variation between the 10th and 90th percentile per population[2], we find a quite extreme value for the 13-year-old tracked population, while the other populations have very similar values for of variation. For mathematics, the variation is over 1.2 for the 13-year-old tracked population of students, while the variation in the three other populations lie close to 0.3. For science, the population of tracked 13-year-olds has a variation of close to 1, while it varies around 0 in the three other populations. From this structure, we can conclude that based on solely this analysis, tracking seems to increase variation in students' performance quite substantially. If we compare the results of tracked and untracked 13-year-olds, we notice that students with relatively poor school performance are better off in an untracked school system, whereas the relatively good students would prefer early tracking. This conclusion supports one of the arguments proponents of heterogeneous classrooms - students with less cognitive ability are left to fall further and further behind in a tracked school system.

[2] $\Delta VA = \frac{VA_{\tau=.9} - VA_{\tau=.1}}{VA_{OLS}}$.

Table 2: Estimated effect of one year of school attendance for tracked and untracked students.

9 yr	**tracked**	**untracked**	13 yr	**tracked**	**untracked**
10%	0.385***	0.427***	10%	0.0844***	0.148***
	(0.0208)	(0.00909)		(0.0197)	(0.00932)
25%	0.476***	0.465***	25%	0.0997***	0.172***
	(0.0164)	(0.00755)		(0.0162)	(0.00832)
50%	0.553***	0.496***	50%	0.138***	0.188***
	(0.0139)	(0.00723)		(0.0151)	(0.00730)
75%	0.569***	0.534***	75%	0.191***	0.196***
	(0.0149)	(0.00816)		(0.0145)	(0.00774)
90%	0.558***	0.578***	90%	0.255***	0.211***
	(0.0192)	(0.00979)		(0.0194)	(0.00897)
mean	0.510***	0.501***	mean	0.141***	0.185***
	(0.0118)	(0.00558)		(0.0116)	(0.00558)
# of countries	23	23	# of countries	23	23
# of students	19,119	71,356	# of students	17,870	79,870

(a) Results mathematics.

9 yr	**tracked**	**untracked**	13 yr	**tracked**	**untracked**
10%	0.415***	0.466***	10%	0.0936***	0.269***
	(0.0223)	(0.0101)		(0.0209)	(0.00990)
25%	0.444***	0.465***	25%	0.142***	0.257***
	(0.0174)	(0.00769)		(0.0166)	(0.00714)
50%	0.441***	0.462***	50%	0.201***	0.263***
	(0.0127)	(0.00645)		(0.0160)	(0.00631)
75%	0.438***	0.447***	75%	0.260***	0.270***
	(0.0134)	(0.00721)		(0.0141)	(0.00693)
90%	0.434***	0.437***	90%	0.278***	0.285***
	(0.0177)	(0.00816)		(0.0195)	(0.00861)
mean	0.435***	0.455***	mean	0.191***	0.267***
	(0.0114)	(0.00536)		(0.0120)	(0.00544)
# of countries	23	23	# of countries	23	23
# of students	19,119	115,704	# of students	17,870	120,143

(b) Results science.

*Notes: Quantile estimations obtained through quantile regression. Mean estimate obtained through OLS estimation. Standard errors are presented in brackets below the estimated coefficients. *, **, and *** stand for significance at 10% level, 5% level, and 1% level, respectively.*

4.2 Interaction model results

Estimates of the effect of tracking on the value added are presented in Table 3. These are obtained using Model 2. It is likely to assume differences in various factors of teaching between tracked and untracked education systems. We observe almost no significant effects for mathematics test, in contrast to the effects found in the science test. The effects seem to be negative, which can be due to adjusted pacing of the course curriculum by teachers. For the average and above average students at age 13, the effect is very close to zero. Applying the same theory for the mathematics test, it also seems adequate that among the 13-year-olds, only the top 10% seems to benefit from tracking.

Table 3: Estimated effect of tracking on the cognitive value gained by students through one year of school attendance.

9 yr	**MAT**	**SCI**
10%	-0.0499**	-0.0607**
	(0.0226)	(0.0248)
25%	-0.00399	-0.0499**
	(0.0180)	(0.0226)
50%	0.0202	-0.0424**
	(0.0164)	(0.0167)
75%	0.0026	-0.0519***
	(0.0173)	(0.0174)
90%	-0.0198	-0.0474**
	(0.0199)	(0.0195)
mean	-0.00712	-0.0493***
	(0.0136)	(0.0137)
# of countries	23	23
# of students	133,889	133,889

(a) Results 9-year-old population.

13 yr	**MAT**	**SCI**
10%	0.00569	-0.0904***
	(0.0297)	(0.0252)
25%	-0.0131	-0.0553***
	(0.0210)	(0.0201)
50%	0.00322	-0.0183
	(0.0176)	(0.0167)
75%	0.017	0.000427
	(0.0169)	(0.0180)
90%	0.0494**	0.0000
	(0.0199)	(0.0228)
mean	0.014	-0.0352**
	(0.0144)	(0.0143)
# of countries	23	23
# of students	137,162	137,162

(b) Results 13-year-old population.

*Notes: Quantile estimations obtained through quantile regression using interaction term between tracking and grade dummy variable. Mean estimate obtained through OLS estimation. Standard errors are presented in brackets below the estimated coefficients. *, **, and *** stand for significance at 10% level, 5% level, and 1% level, respectively.*

4.3 Interquantile regression results

In Model 3, I apply interquantile regression to the model in Equation 2 in order to obtain the estimations of difference in value added between the top- and bottom 10% of students, caused by tracking. The results are presented in Table 4.

Table 4: Estimated effect of tracking on the difference between students from the 10th and 90th percentile ($\tau = .1$ and $\tau = .9$).

	MAT	**SCI**
9YR	0.0300	0.0133
	(0.0188)	(0.0335)
13YR	0.0437**	0.0904***
	(0.0207)	(0.0299)
# of countries	23	23
# of students	271,051	271,051

*Notes: Interquantile estimations obtained through interquantile regression using between the 10th and 90th percentile of students' test scores. Standard errors are presented in brackets below the estimated coefficients. *, **, and *** stand for significance at 10% level, 5% level, and 1% level, respectively.*

The findings seem to confirm the previous conclusions. There seems to be no significant difference in value added between the two levels of students at age 9, in contrast to the evident difference detected between students at age 13. Note that despite the difference in significance, the coefficients of the 9-year-old sample and the 13-year-old sample do not seem to differ drastically. However, this is anticipated, as in earlier findings we observed that the value added for 13-year-olds lies relatively lower than for 9-year-olds. These results validate the conclusion that the difference in value added between the top- and bottom 10% is significantly enlarged after tracking is introduced.

5 Sensitivity

To test the robustness of the outcomes presented in Section 4, I perform several sensitivity checks. In Section 5.1, I reduce the sample of countries to Western countries, excluding the countries in Asia and the Middle-East. This way, I hope to purify the comparability of the countries' data. In Section 5.2, I look into the structure found earlier in the results of the simple model (see Section 4.1), concerning the difference in value added between the top- and bottom 10% of students, within each country. And lastly, I investigate possible differences between male and female students, as gender differences is frequently encountered in previous studies on education.

5.1 Comparable country selection

It is fairly possible that, for instance, students from Japan live in a different environment from students from the Netherlands. There is a different view on education, and children live in a different set of ideas and thinking. These are only a few of the unmeasurable factors which could very likely affect school performance. In order to make my sample of countries more homogeneous with regard to unobservable factors, I eliminate all countries in Asia and the Middle-East. This means clearing the database of students from Hong Kong, Iran, Japan, Korea, Singapore, and Thailand, leaving in the subsample students from Australia, Austria, Canada, Cyprus, Czech Republic, Greece, Hungary, Iceland, Ireland, Latvia, the Netherlands, New Zealand, Norway, Portugal, Scotland, Slovenia, and United Kingdom. Note that this means using the subsample may only alter the results for the untracked students, as no tracked countries are eliminated from the sample.

First, I rerun the simple model for the subsample. Results can be found in Table 5. If we compare these outcomes with the outcomes using the full sample (see Table 2), we observe that the value added increases a bit for students performing below average, while it drops for students performing above average. However, this does not change (and only strengthens) the fact that the variation between th 10th and 90th percentile remains the highest for the tracked 13-year-olds.

Table 5: Estimated effect of one year of school attendance for tracked and untracked students within the subsample of comparable countries.

9 yr	**tracked**	**untracked**	13 yr	**tracked**	**untracked**
10%	0.385***	0.462***	10%	0.0844***	0.181***
	(0.0208)	(0.00986)		(0.0197)	(0.0116)
25%	0.476***	0.494***	25%	0.0997***	0.191***
	(0.0164)	(0.00782)		(0.0162)	(0.0145)
50%	0.553***	0.511***	50%	0.138***	0.201***
	(0.0139)	(0.00816)		(0.0151)	(0.00746)
75%	0.569***	0.516***	75%	0.191***	0.210***
	(0.0149)	(0.00841)		(0.0145)	(0.00789)
90%	0.558***	0.511***	90%	0.255***	0.204***
	(0.0192)	(0.00816)		(0.0194)	(0.00872)
mean	0.510***	0.496***	mean	0.141***	0.195***
	(0.0118)	(0.00637)		(0.0116)	(0.00608)
# of countries	17	17	# of countries	17	17
# of students	19,119	71,356	# of students	17,870	79,870

(a) Results mathematics.

9 yr	**tracked**	**untracked**	13 yr	**tracked**	**untracked**
10%	0.415***	0.496***	10%	0.0936***	0.257***
	(0.0223)	(0.0127)		(0.0209)	(0.0122)
25%	0.444***	0.474***	25%	0.142***	0.249***
	(0.0174)	(0.0102)		(0.0166)	(0.00980)
50%	0.441***	0.469***	50%	0.201***	0.264***
	(0.0127)	(0.00903)		(0.0160)	(0.00769)
75%	0.438***	0.440***	75%	0.260***	0.258***
	(0.0134)	(0.00912)		(0.0141)	(0.00869)
90%	0.434***	0.417***	90%	0.278***	0.249***
	(0.0177)	(0.0115)		(0.0195)	(0.00980)
mean	0.435***	0.459***	mean	0.191***	0.254***
	(0.0114)	(0.00678)		(0.0120)	(0.00662)
# of countries	17	17	# of countries	17	17
# of students	19,119	71,356	# of students	17,870	79,870

(b) Results science.

*Notes: Quantile estimations obtained through quantile regression using the subsample of comparable countries. Mean estimate obtained through OLS estimation. Standard errors are presented in brackets below the estimated coefficients. *, **, and *** stand for significance at 10% level, 5% level, and 1% level, respectively.*

Next, I rerun the interaction model estimating the effect of tracking on value added. Results are presented in Table 6. All estimates seem to have decreased for the 13-year-old population, whereas in the 9-year-old population, the value added increases for the students performing above average. However, the conclusion to be drawn here is still that tracking is negatively significant for students with relatively less potential. The beneficial effect for the relatively good students, however, disappears.

Table 6: Estimated effect of tracking on the cognitive value gained by students in the sub-sample of comparable countries through one year of school attendance.

9 yr	**MAT**	**SCI**
10%	-0.0789***	-0.0856***
	(0.0225)	(0.0260)
25%	-0.00688	-0.0606***
	(0.0192)	(0.0192)
50%	0.0256	-0.0515***
	(0.0166)	(0.0167)
75%	0.0304	-0.0391**
	(0.0191)	(0.0184)
90%	0.0297	-0.0439**
	(0.0223)	(0.0212)
mean	0.00157	-0.0550***
	(0.0133)	(0.0140)
# of countries	17	17
# of students	89,549	89,549

(a) Results 9-year-old population.

13 yr	**MAT**	**SCI**
10%	-0.0591**	-0.0940***
	(0.0257)	(0.0268)
25%	-0.0500**	-0.0655***
	(0.0195)	(0.0217)
50%	-0.0401**	-0.0376**
	(0.0161)	(0.0176)
75%	-0.0116	-0.0132
	(0.0180)	(0.0181)
90%	0.0195	-0.0116
	(0.0205)	(0.0180)
mean	-0.0335**	-0.0470***
	(0.0139)	(0.0148)
# of countries	17	17
# of students	96,890	96,890

(b) Results 13-year-old population.

*Notes: Quantile estimations obtained through quantile regression using interaction term between tracking and grade dummy variable. Mean estimate obtained through OLS estimation. Estimation sample contains comparable countries only. Standard errors are presented in brackets below the estimated coefficients. *, **, and *** stand for significance at 10% level, 5% level, and 1% level, respectively.*

I also perform interquantile regression on this subsample, estimating the difference in value added between the 10th and 90th percentile caused by tracking. Results are presented in Table 7. The findings correspond to the findings in the full sample, as the inequality in performance amongst 13-year-olds significantly increases due to tracking. However, for the subject mathematics, this effect is already present amongst 9-year-old students where tracking has not yet been introduced in any country. This suggests the existence of substantial differences between the education systems of tracked and untracked countries, and the effect of the education system on the students' value added.

Table 7: Estimated effect of tracking on the difference between students from the 10th and 90th percentile ($\tau = .1$ and $\tau = .9$) from the subsample of comparable countries.

	MAT	**SCI**
9YR	0.109***	0.0417
	(0.0301)	(0.0298)
13YR	0.0785***	0.0824**
	(0.0252)	(0.0338)
# of countries	17	17
# of students	186,439	186,439

*Notes: Interquantile estimations obtained through interquantile regression using between the 10th and 90th percentile of students' test scores. Estimation sample contains comparable countries only. Standard errors are presented in brackets below the estimated coefficients. *, **, and *** stand for significance at 10% level, 5% level, and 1% level, respectively.*

Notice that a peculiar structure is derived from the analysis without the countries in Asia and the Middle-East, namely the fact that the models seem to suggest a lower value added than the value added found using the full sample. This might suggest that, compared to students in the countries left in the subsample, students in the excluded countries acquire an overall higher cognitive development in the years between age 9 and age 13. However, this is not the main focus of this study, and therefore I will withhold any conclusions concerning this matter.

5.2 Analysis at country level

Recall the peaking variance between the 10th and 90th percentile of students found for the 13-year-old tracked population (see Section 4.1). In order to look even further into this structure, I perform the simple quantile model on country specific level for the subsample of comparable countries (see Section 5.1). This allows me to check if this peaking structure observed in the tracked 13-year-old population is also present in the individual countries. Note that even though I only included comparable countries in this sensitivity analysis, there are still unmeasured differences amongst both the tracked and untracked countries, which means differences between the estimations of tracked and untracked countries cannot entirely be interpreted as the effect of tracking. In Table 8, the variances between the 10th and 90th percentile of students are presented, i.e. the difference in value added between the top- and bottom 10%, divided by the mean value added. The 4 countries above the dashed line apply early tracking, and the countries below the dashed line do not introduce tracking before the age of 14. The pattern in the variances seem quite irregular, but overall, excluding outliers like Ireland and Portugal, the variances at age 13 seem to lie higher in the tracked countries, while the variances at age 9 seem to lie higher in the untracked countries. This approximation can be tested straightforwardly through calculating the mean variances per population. The average variance of the tracked 9-year-old population is 0.205, while the average variance of the untracked 9-year-old population is equal to 0.467. Simultaneously, in the 13-year-old population, tracked countries have an average variance of 1.063, while for the untracked countries, this is equal to 0.526.

Table 8: Variances in value added between 10th and 90th percentile of students.

COUNTRY	9YR/MAT	9YR/SCI	13YR/MAT	13YR/SCI
Austria	*0.387*	*0.221*	*0.234*	*0.960*
Czech Republic	*0.343*	*-0.093*	*0.572*	*0.632*
Hungary	*0.164*	*-0.345*	*1.174*	*0.802*
Netherlands	*0.496*	*0.470*	*3.282*	*0.851*
Australia	0.241	0.097	0.587	0.513
Canada	0.216	-0.045	0.347	0.065
Cyprus	0.275	0.054	0.370	0.414
Greece	7.010	3.238	-0.017	-0.170
Iceland	0.159	-0.014	0.160	-0.075
Ireland	-0.256	0.067	6.085	1.457
Latvia	1.107	0.513	0.903	0.545
New Zealand	0.054	-0.056	0.031	0.151
Norway	0.077	-0.246	-0.027	-0.329
Portugal	0.038	-0.432	1.530	0.883
Scotland	0.195	-0.142	-0.045	-0.039
Slovenia	0.125	-0.104	0.564	0.172
United Kingdom	0.121	-0.154	-0.230	-0.166

Notes: The variances between the 10th and 90th percentile of students are calculated as the difference in estimated value added between the top- and bottom 10%, divided by the estimated mean value added (obtained from OLS).

To validate any differences between the 10th and 90th percentile of students, I perform interquantile regression for each country separately. The results are presented in Table 9. The structure becomes vague in the country specific results. We observe significant differences in the tracked countries at the age of 13, which are also found in some untracked countries. However, the sample size varies across the countries, which means that the results per country might not be suitable for direct comparison. Despite the lack of evidence provided by interquantile regression, the country specific analysis generally shows support for the structure with higher sample variance for tracked 13-year-olds. It also shows us that students in tracked countries have a lower value added than students in countries without early tracking, before tracking is introduced. Again, we need to keep in mind that no strong conclusions can be based on these comparisons, as there are most likely many unobserved differences between the countries which influences the performance inequality of students.

Table 9: At country level estimated effect of tracking on the difference between the 10th and 90th percentile of students.

COUNTRY	9YR/MAT	9YR/SCI	13YR/MAT	13YR/SCI	# of students
Austria	*0.213****	*0.0900*	*0.0373*	*0.166***	*8661*
	(0.0607)	*(0.0563)*	*(0.0631)*	*(0.0647)*	
Czech Republic	*0.185****	*-0.0466*	*0.131****	*0.156****	*11985*
	(0.0395)	*(0.0411)*	*(0.0434)*	*(0.0304)*	
Hungary	*0.0554*	*-0.111**	*0.138***	*0.0980*	*8721*
	(0.0535)	*(0.0606)*	*(0.0560)*	*(0.0768)*	
Netherlands	*0.299****	*0.194****	*0.354****	*0.252****	*7622*
	(0.0516)	*(0.0649)*	*(0.0652)*	*(0.0618)*	
Australia	0.102**	0.0354	0.0880***	0.121***	20947
	(0.0398)	(0.0518)	(0.0325)	(0.0341)	
Canada	0.120***	-0.0216	0.0862***	0.0145	29609
	(0.0218)	(0.0333)	(0.0202)	(0.0348)	
Cyprus	0.162***	0.0273	0.0643	0.123**	11399
	(0.0498)	(0.0471)	(0.0428)	(0.0539)	
Greece	0.736	0.289	0.0104	0.0938	4413
	(0.603)	(0.584)	(0.0860)	(0.0733)	
Iceland	0.0931	-0.00883	0.0381	-0.0224	7108
	(0.0636)	(0.0709)	(0.0545)	(0.0470)	
Ireland	-0.0831	0.0187	0.192***	0.220***	9297
	(0.0589)	(0.0806)	(0.0486)	(0.0697)	
Latvia	0.268***	0.0817	0.176***	0.216***	7347
	(0.0867)	(0.0808)	(0.0476)	(0.0665)	
New Zealand	0.0245	-0.0255	0.00850	0.0557	10984
	(0.0699)	(0.0642)	(0.0625)	(0.0410)	
Norway	0.0599	-0.192***	-0.00969	-0.137***	10018
	(0.0403)	(0.0450)	(0.0355)	(0.0355)	
Portugal	0.0165	-0.212***	0.231***	0.302***	9451
	(0.0547)	(0.0594)	(0.0532)	(0.0461)	
Slovenia	0.0680	-0.0532	0.114**	0.0229	9897
	(0.0556)	(0.0423)	(0.0578)	(0.0405)	
United Kingdom	0.0630	-0.0791	-0.0759	-0.0662	9584
	(0.0395)	(0.0579)	(0.0723)	(0.0861)	
Scotland	0.105**	-0.0678	-0.0153	-0.0168	11172
	(0.0447)	(0.0688)	(0.0463)	(0.0768)	

*Notes: Tracked countries are Austria, Czech Republic, Hungary, and the Netherlands. Standard errors are presented in brackets below the estimated coefficients. *, **, and *** stand for significance at 10% level, 5% level, and 1% level, respectively.*

5.3 Differences by gender

Students' performance in the subjects mathematics and science have been investigated frequently in existing literature. Quite oftenly, differences were found between boys and girls (e.g. Hyde et al. (1990), Lee and Burkam (1996)). In order to check whether the findings on the effect of tracking differ between the genders, I rerun the three models separately for male and female students.

I start by estimating the value added separately male and female students. Results are presented in Table 10. Comparing the value added between boys and girls regarding mathematics, we observe overall higher values for girls before tracking takes place, suggesting that school attendance adds more to the performance of female students. After tracking takes place, girls still have an overall higher value added than boys. However, if we look at the 90% estimations, male students in tracked systems seem to have a higher value added than their female classmates, whereas male students in untracked systems do not. Moving on to the subject science, we observe in the 9-year-old sample similar values added amongst male and female students in tracked and untracked countries, which persists in the 13-year-old-sample, suggesting no gender differences in value added for the subject science.

Next, I estimate the effect of tracking on value added separately for male and female students. Results can be found in Table 11. For the interpretation of these results, it is useful to recall the results obtained using the full sample (see Table 3). Recall the conclusions we have drawn from the full sample analysis, namely that a tracked education system negatively affects the value added at the age of 9 for the subject science, and continues to do so at the age of 13 only for students with less potential. Above average students at age 13 are advantaged by tracking, or unaffected, for both mathematics and science. Now, observe the gender specific results. There are hardly any significant effects for mathematics, which is equivalent to the full sample outcomes. However, for the subject science, we observe negative significant effects mainly in the 9-year-old male student sample. Recall that this was also observed in the full sample outcomes. This negative effect is also found in the 9-year-old female student sample, but it is statistically less convincing. This suggests that tracked education systems seems to only negatively affects boys at age 9 for the subject science, and does not necessarily have any negative effect on girls. However, if we focus on the 13-year-old sample, we observe strong negative effects of tracking in the 10th and 25th percentile of female students for the subject science. This was also one of the findings from the full sample analysis. Male students also experience negative effects, but far less evident. Conclusively, this analysis has shown us that the conclusions we have drawn earlier from the full sample should be split up between the genders. In other words, the strong negative effects of tracking for science at age 9 are mostly endured by male students, while the negative effects for bottom students at age 13 are mostly endured by female students. All in all, negative effects of tracking are still only present among bottom students. Above average students are unaffected, or even favored.

Table 10: Estimated effect of one year of school attendance on the cognitive value gained by male and female students.

9 yr	**tracked**	**untracked**	13 yr	**tracked**	**untracked**
10%	0.351***	0.401***	10%	0.0459	0.111***
	(0.0294)	(0.0136)		(0.0292)	(0.0145)
25%	0.449***	0.445***	25%	0.0607**	0.145***
	(0.0283)	(0.0106)		(0.0275)	(0.0122)
50%	0.534***	0.477***	50%	0.127***	0.165***
	(0.0182)	(0.0101)		(0.0218)	(0.0102)
75%	0.542***	0.522***	75%	0.184***	0.193***
	(0.0216)	(0.0115)		(0.0187)	(0.0107)
90%	0.525***	0.556***	90%	0.270***	0.203***
	(0.0278)	(0.0133)		(0.0271)	(0.0123)
mean	0.490***	0.483***	mean	0.126***	0.167***
	(0.0172)	(0.00803)		(0.0175)	(0.00813)
# of countries	23	23	# of countries	23	23
# of students	9,351	57,610	# of students	8,616	59,797

(a) Results mathematics male students.

9 yr	**tracked**	**untracked**	13 yr	**tracked**	**untracked**
10%	0.429***	0.446***	10%	0.113***	0.181***
	(0.0262)	(0.0129)		(0.0284)	(0.0123)
25%	0.487***	0.486***	25%	0.122***	0.196***
	(0.0203)	(0.0108)		(0.0199)	(0.0107)
50%	0.565***	0.514***	50%	0.142***	0.212***
	(0.0215)	(0.00995)		(0.0201)	(0.00944)
75%	0.587***	0.553***	75%	0.199***	0.200***
	(0.0222)	(0.0115)		(0.0207)	(0.0102)
90%	0.575***	0.602***	90%	0.232***	0.223***
	(0.0266)	(0.0129)		(0.0252)	(0.0130)
mean	0.526***	0.520***	mean	0.154***	0.204***
	(0.0162)	(0.00785)		(0.0155)	(0.00767)
# of countries	23	23	# of countries	23	23
# of students	9,708	56,693	# of students	9,241	60,085

(b) Results mathematics female students.

9 yr	**tracked**	**untracked**	13 yr	**tracked**	**untracked**
10%	0.416***	0.456***	10%	0.105***	0.241***
	(0.0340)	(0.0153)		(0.0309)	(0.0149)
25%	0.445***	0.454***	25%	0.151***	0.250***
	(0.0251)	(0.0115)		(0.0235)	(0.0110)
50%	0.442***	0.460***	50%	0.225***	0.272***
	(0.0174)	(0.0101)		(0.0250)	(0.00975)
75%	0.438***	0.447***	75%	0.311***	0.284***
	(0.0216)	(0.0103)		(0.0235)	(0.0103)
90%	0.420***	0.453***	90%	0.342***	0.297***
	(0.0248)	(0.0117)		(0.0295)	(0.0115)
mean	0.435***	0.457***	mean	0.218***	0.267***
	(0.0166)	(0.00783)		(0.0180)	(0.00797)
# of countries	23	23	# of countries	23	23
# of students	9,351	57,610	# of students	8,616	59,797

(c) Results science male students.

9 yr	**tracked**	**untracked**	13 yr	**tracked**	**untracked**
10%	0.426***	0.472***	10%	0.0793***	0.293***
	(0.0313)	(0.0130)		(0.0264)	(0.0144)
25%	0.431***	0.473***	25%	0.131***	0.263***
	(0.0230)	(0.0104)		(0.0261)	(0.00941)
50%	0.433***	0.460***	50%	0.189***	0.253***
	(0.0182)	(0.00865)		(0.0207)	(0.00848)
75%	0.437***	0.449***	75%	0.222***	0.257***
	(0.0176)	(0.00958)		(0.0186)	(0.00956)
90%	0.437***	0.421***	90%	0.231***	0.273***
	(0.0276)	(0.0115)		(0.0264)	(0.0123)
mean	0.432***	0.452***	mean	0.166***	0.267***
	(0.0157)	(0.00738)		(0.0160)	(0.00743)
# of countries	23	23	# of countries	23	23
# of students	9,708	56,693	# of students	9,241	60,085

(d) Results science female students.

*Notes: Quantile estimations obtained through quantile regression using subsamples of male and female students. Mean estimate obtained through OLS estimation. Standard errors are presented in brackets below the estimated coefficients. *, **, and *** stand for significance at 10% level, 5% level, and 1% level, respectively.*

Table 11: Estimated effect of tracking on the cognitive value gained by male and female students through one year of school attendance.

9 yr	**MAT**	**SCI**
10%	-0.0598*	-0.0555
	(0.0341)	(0.0402)
25%	-4.75e-05	-0.0356
	(0.0285)	(0.0292)
50%	0.0143	-0.0504**
	(0.0255)	(0.0253)
75%	0.0123	-0.0602**
	(0.0254)	(0.0247)
90%	-0.0194	-0.0737**
	(0.0313)	(0.0306)
mean	-0.0107	-0.0537***
	(0.0198)	(0.0202)
# of countries	23	23
# of students	66,528	66,528

(a) Results 9-year-old population of male students.

13 yr	**MAT**	**SCI**
10%	0.00592	-0.0758*
	(0.0424)	(0.0421)
25%	-0.0236	-0.0511*
	(0.0317)	(0.0293)
50%	-0.0106	-0.00257
	(0.0238)	(0.0255)
75%	0.0178	0.0264
	(0.0261)	(0.0264)
90%	0.0667**	0.0262
	(0.0287)	(0.0316)
mean	0.0138	-0.0124
	(0.0214)	(0.0216)
# of countries	23	23
# of students	68,000	68,000

(b) Results 13-year-old population of male students.

9 yr	**MAT**	**SCI**
10%	-0.0331	-0.0742**
	(0.0305)	(0.0356)
25%	-0.00240	-0.0401
	(0.0245)	(0.0293)
50%	0.0184	-0.0399*
	(0.0223)	(0.0231)
75%	-0.0159	-0.0413*
	(0.0251)	(0.0230)
90%	-0.0237	-0.0362
	(0.0289)	(0.0292)
mean	-0.00784	-0.0475**
	(0.0187)	(0.0185)
# of countries	23	23
# of students	66,000	66,000

(c) Results 9-year-old population of female students.

13 yr	**MAT**	**SCI**
10%	-0.00428	-0.102***
	(0.0379)	(0.0381)
25%	-0.0180	-0.0704**
	(0.0278)	(0.0287)
50%	0.00887	-0.0379*
	(0.0247)	(0.0215)
75%	0.0272	-0.0291
	(0.0246)	(0.0242)
90%	0.0209	-0.0487
	(0.0282)	(0.0315)
mean	0.00914	-0.0621***
	(0.0193)	(0.0191)
# of countries	23	23
# of students	68,889	68,889

(d) Results 13-year-old population of female students.

*Notes: Quantile estimations obtained through quantile regression using interaction term between tracking and grade dummy variable. Mean estimate obtained through OLS estimation. Estimation sample contains comparable countries only. Standard errors are presented in brackets below the estimated coefficients. *, **, and *** stand for significance at 10% level, 5% level, and 1% level, respectively.*

Finally, I test whether tracking enlarges inequality in students' performances by rerunning Model 3 separately for male and female students. Results are presented in Table 12. We only observe a significant effect in the 13-year-old male student sample for the subject science. Tracking appears to be significantly enlarging inequality in performance for this population of students. Recall the results in Table 4, obtained using the full sample, where we observed significant effects of tracking for both subjects. However, the effect in mathematics was only significant at 5% level, and combining it with the results by gender, the conclusion that tracking enlarges inequality for this subject seems to be less compelling. The conclusion is still valid for the subject science in the male student population, but also drops in significance (from 1% level to 5% level). Conclusively, gender difference analyses have shown that conclusions drawn earlier are different between the genders. Not all positive or negative effects is experienced by both sexes.

Table 12: Estimated effect of tracking on the difference between students from the 10th and 90th percentile.

	MAT	**SCI**
9YR	0.0404	-0.0182
	(0.0496)	(0.0446)
13YR	0.0607	0.102**
	(0.0512)	(0.0438)
# of countries	23	23
# of students	134,528	134,528

(a) Results of male students.

	MAT	**SCI**
9YR	0.0094	0.038
	(0.0503)	(0.0309)
13YR	0.0252	0.0537
	(0.0419)	(0.0406)
# of countries	23	23
# of students	134,889	134,889

(b) Results of female students.

*Notes: Interquantile estimations obtained through interquantile regression using between the 10th and 90th percentile of students' test scores, for the male and female subsample. Standard errors are presented in brackets below the estimated coefficients. *, **, and *** stand for significance at 10% level, 5% level, and 1% level, respectively.*

6 Conclusions

This study provides insights into effects of early tracking on the cognitive value students gain through one year of school attendance. Simple quantile regression results have shown evident differences between the value added coefficients of students with relatively less potential and the value added coefficients of top students. This difference is confirmed to be statistically significant through the use of an interaction model of tracking and value added, as the effect of tracking on the bottom 10% of students appears to be negative, while the top 10% of students seem to be unharmed, or even favored by tracking. This result is validated by means of interquantile regression, where the difference between the top- and bottom 10% is estimated. The difference is only (positively) significant in the 13-year-old population, i.e. tracking seems to amplify the difference between the top- and bottom 10% of students. Therefore, we can conclude that educational tracking enlarges the inequality in students' cognitive ability, in the sense that top students benefit from homogeneous classrooms, whereas the bottom students are better off in heterogeneous classrooms.

In order to test the robustness of the conclusions, three sensitivity checks are performed. First, I improved the comparability of countries by reducing the sample to Western countries only. Results appear to be in line with findings from the full sample analysis. Next, I performed country specific analysis, to check whether the findings are also valid in within each individual country. Despite varying support from the results per country, I was able to state that tracked countries have a higher value added than untracked countries after tracking is introduced. In addition, before the introduction of tracking, results have shown that untracked countries have a higher value added than tracked countries. And lastly, I checked for gender differences in the effect of tracking, and I found that the results are gender specific. Not all positive or negative effects of tracking seem to be equally evident for both sexes.

In sum, the findings in this study supports arguments of both proponents and opponents of tracking. Homogeneous classrooms maximize the efficiency of learning. However, this efficiency is only gained by above average students. This makes sense, as top students presumably put more effort into school work compared to below average students, who are disadvantaged by a tracked system as they are not stimulated by the peer-effects. This way, tracking drifts students with different levels of cognitive ability further and further apart from each other, causing inequality in the school performance to increase.

References

S. Cahan and N. Cohen. Age versus schooling effects on intelligence development. *Child Development*, 60(5):1239–1249, October 1989.

E. Duflo, P. Dupas, and M. Kremer. Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. *American Economic Review*, 101(5):1739–74, August 2011.

R. Van Elk, M. Van Der Steeg, and D. Webbink. Does the timing of tracking affect higher education completion? *Economics of Education Review*, 2011.

E.A. Hanushek and L. Woessmann. Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *The Economic Journal*, 116: C63–C76, March 2006.

J.S. Hyde, E. Fennema, and S.J. Lamon. Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, 107(2):139–155, March 1990.

R. Koenker and G. Bassett. Regression quantiles. *Econometrica*, 46(1):33–50, January 1978.

D.S. Lee and T. Lemieux. Regression discontinuity designs in economics. *Journal of Economic Literature, American Economic Association*, 48(2):281–355, June 2010.

V.E. Lee and D.T. Burkam. Gender differences in middle grade science achievement: Subject domain, ability level, and course emphasis. *Science Education*, 80(6):613–650, November 1996.

J. Rothstein. Teacher quality in educational production: Tracking, decay, and student achievement. *The Quarterly Journal of Economics*, 125(1):715–214, February 2010.