Does Ability Tracking Exacerbate the Role of Family Background for Students' Test Scores?

Fabian Waldinger (LSE & CEP)*

February 16, 2007

Abstract

This study investigates whether ability tracking exacerbates the role of parental background for students' educational test scores. Using microdata from different educational studies, PISA, PIRLS and TIMSS, this paper exploits the cross-country variation in tracking policies to identify the effect of tracking. Controlling for unobserved country level variables using difference-in-differences, I find that family background is more important in early tracking countries but that the importance of family background does not increase after actual tracking has taken place. This suggests that tracking does not augment the role of family background for students' test scores. Factors other than tracking are more likely to be responsible for the fact that family background is more important in early tracking countries. This result runs contrary to the findings of the current literature. In support of my findings, I show that the results of the current literature are not robust to slight changes in specification.

1 Introduction

The results from the Programme for International Student Assessment (PISA) have triggered a serious debate about the functioning of education systems. An important finding of PISA is the widely varying influence of a student's family background on her educational achievement across different countries. There are a number of possible explanations for this phenomenon, with features of the schooling system being widely seen as key factors. A frequently discussed feature is ability tracking of students. Ability tracking, or simply tracking, means placing students into different school tracks according to some measure of

*I thank Holger Breinlich, Rocco Macchiavello, Steve Machin, Ralph Ossa, Matthias Parey, Giacomo Rodano, Justin Smith and especially Steve Pischke for very helpful discussions and comments. I also thank seminar participants at the LSE and the CEPR conference on the Economics of Education in Padua for helpful suggestions.

their ability. For the purposes of this paper I refer to tracking as a policy which places students of different abilities in different *schools*.[1] In practice, the use of tracking varies greatly across countries; some educate all children in the same type of school until high-school graduation while others separate children as soon as they reach the age of 10. Differences in tracking policies and the widely varying impact of family background on educational outcomes across different countries has lead politicians but also people involved in educational research to argue that tracking has a causal impact on the importance of family background for educational achievement.

One of the main theoretical arguments is that children develop gradually and ability levels are difficult to observe at early ages. In a system that tracks students early, children from higher socioeconomic backgrounds may be more likely to be placed in academic tracks for reasons independent of their true ability. The reasoning behind this argument is that parents from higher socioeconomic backgrounds are more aware of the importance of sending their children to the more academic school tracks.

Also in untracked schooling systems do parents from higher socioeconomic backgrounds have the possibility to ensure a better education for their children; albeit through different channels: parents may move closer to a better school or place their children in a private school. Whether tracking actually exacerbates the importance of parental background for educational achievement is ultimately an empirical question.

The findings of PISA and other cross-country educational studies show a positive correlation between tracking and the importance of the family background for students' test scores. Nonetheless, I argue in this paper that it may be overhasty to interpret this correlation as a causal effect. In fact, I use a difference-in-differences methodology to show that tracking does not exacerbate the importance of parental background for students' test scores once controls for pre-tracking differences are added to the econometric specification.

I use cross-country differences in tracking policies to identify the impact of tracking on the importance of family background for students' test scores using data from three large international educational studies: PISA (Programme for International Student Assessment), PIRLS (Progress in International Reading Literacy Study), and TIMSS (Third International Mathematics and Science Study).

The following two graphs illustrate the effect I estimate in this paper. For Figure 1, I first regressed post-tracking reading test scores on the three parental background characteristics shown on the vertical axes of the graphs, separately for each country. I then plot each country's coefficient against its tracking grade for those background characteristics.[2]

[1] Placing students of different abilities into different classes within a school type is less strong form of tracking. In this paper, however, I do not analyze this within-school tracking.

[2] The estimating equation is: $TestScore = \beta_1 + \beta_2 ParentBackground + \beta_3 StudentCharacteristics + \epsilon$. The β_2 coefficient for each country is then plotted on the vertical axis of each graph.

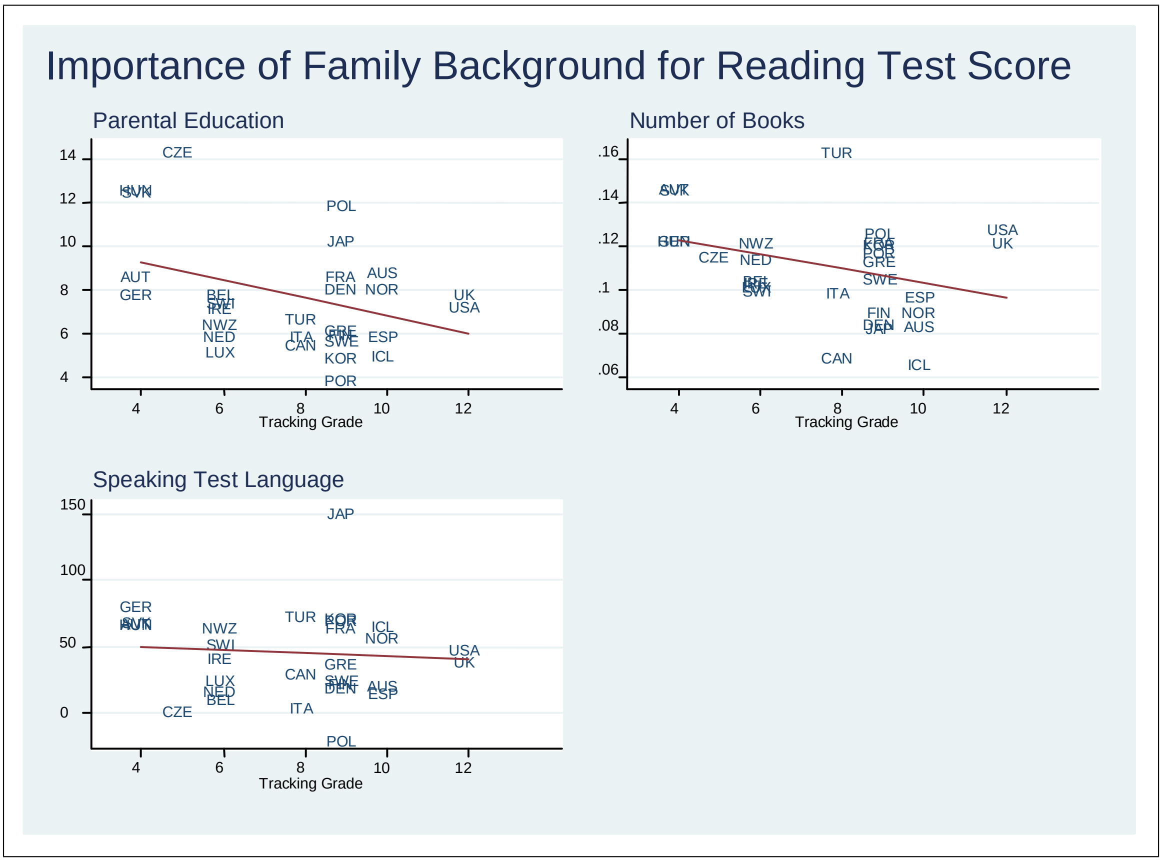

Figure 1: Family Background and Tracking Grade PISA 2003 (Grades 9/10)

Figure 1 shows clearly that background has a higher impact on reading test scores in early tracking countries for all parental background measures. In Figure 2 I use the same methodology as for Figure 1 but replacing PISA reading scores with TIMSS mathematics scores. This figure shows that the same relationship exists for mathematics.

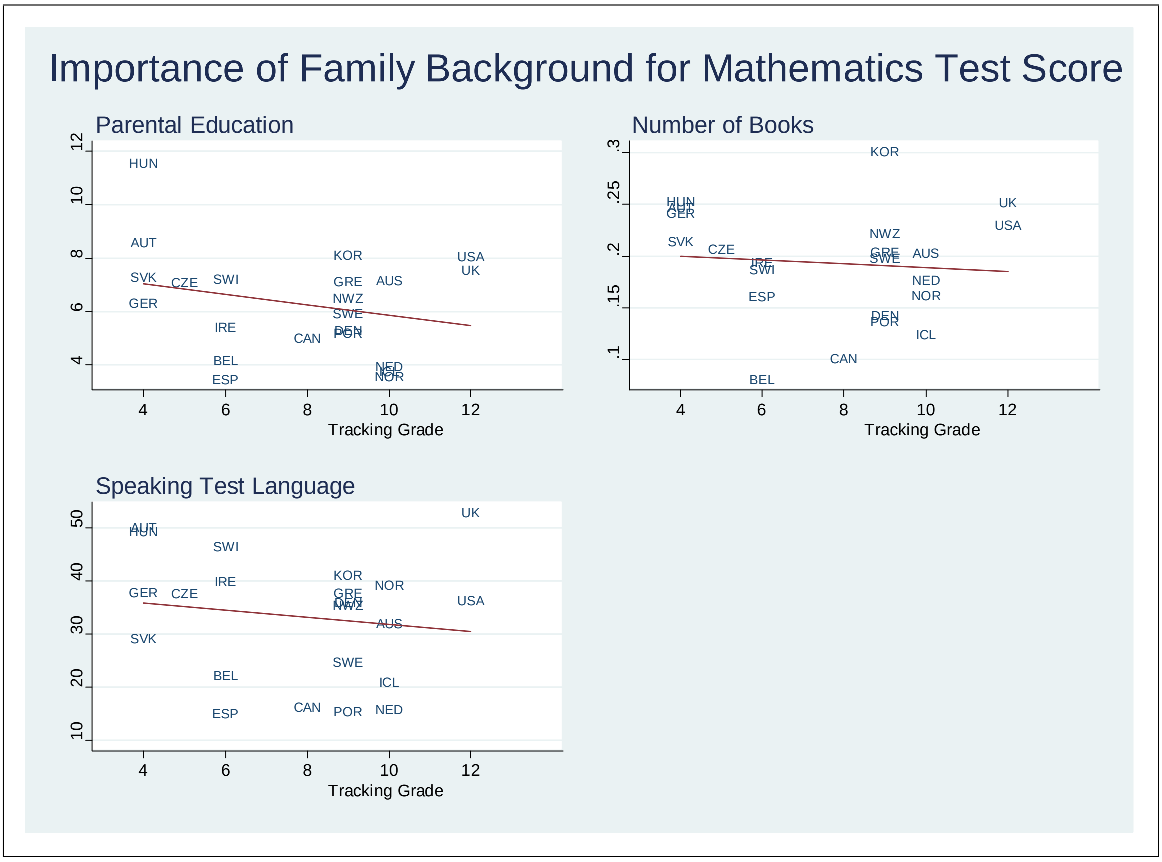

Figure 2: Family Background and Tracking Grade TIMSS 1995 (Grades 7/8)

The visual evidence from Figures 1 and 2 clearly indicate that family background is more important for secondary school students' test scores in early tracking countries. The fact that similar relationships exist for different studies (PISA and TIMSS, which are carried out by different organizations and use different methodologies) hints that this relationship is not a random feature of any particular dataset. The graphs do not, however, provide evidence about whether tracking has a causal impact on the importance of parental background on students' test scores. It may well be the case that there exist unobserved country-level factors which affect the impact of parental background on educational test scores. If these unobserved factors are correlated with a country's tracking regime, estimates using a simple cross-section of students would be biased.

A simple way to investigate the role of unobserved factors is to repeat the graphical analysis from Figures 1 and 2 using tests administered to students before tracking has taken place in any of the countries in the sample. In the absence of anticipation effects there should be no relationship between the tracking regime (which affects students after taking the test) and the importance of the students' parental background. For reading, I use data from the Progress in International Reading Literacy Study (PIRLS), which tested primary school students in grade 4. It is important to note, that none of the countries in the sample tracks students before the end of grade 4. Applying the same techniques used in the previous graphs, Figure 3 shows the relationship between the importance of family background for test scores and the tracking grade (which will eventually affect students) for this new data. Figure 4 then repeats this exercise with the grades 3 and 4 mathematics data from TIMSS.[3]

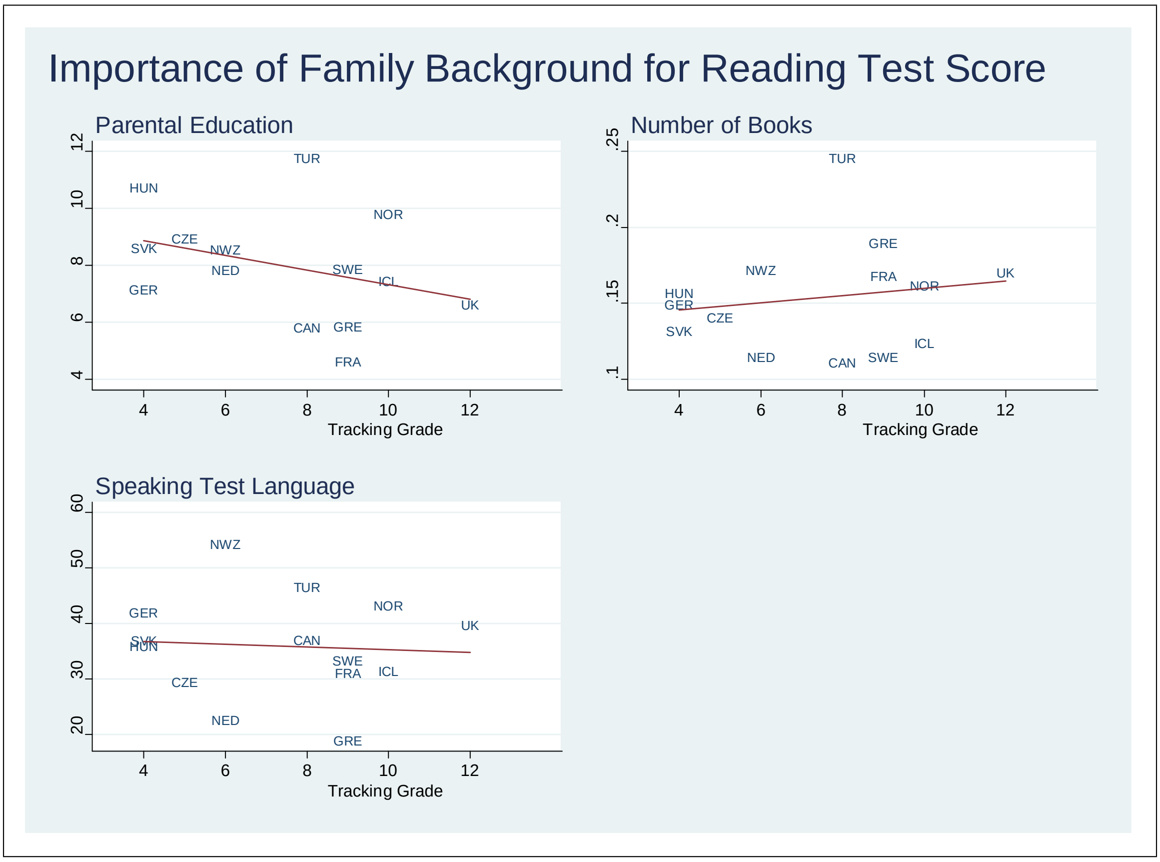

Figure 3: Family Background and Tracking Grade PIRLS 2001 (Grade 4)

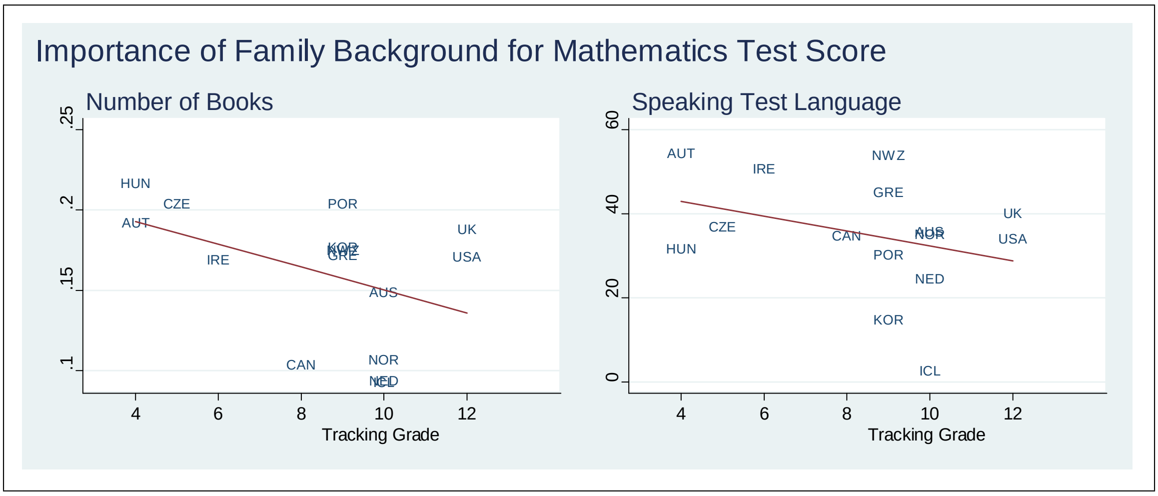

Figure 4: Family Background and Tracking Grade TIMSS 1995 (Grades 3/4)

[3]Unfortunately, the TIMSS data for the primary school grades does not include parental education. Therefore, I omit this variable from the graphical analysis for mathematics test scores of primary school students.

Apart from the number of books variable for the reading test score, there is a negative relationship between the importance of parental background and the tracking regime even before tracking has taken place in any country. This suggests that there are unobserved factors affecting the impact of parental background that are correlated with a country's tracking regime. Investigating the effect of tracking therefore requires an identification strategy that controls for these pre-existing differences across early and late tracking countries. To address this problem, I use a difference-in-differences (DiD) methodology. The DiD results presented below indicate that, once the "pre-tracking" level of the family background effect on children's test scores is controlled for, tracking no longer affects the impact of family background.[4]

These results cast serious doubt on the conclusions of a number of concurrent papers, such as Hanushek and Woessmann (2006) or Ammermueller (2005), which find that tracking increases educational inequality and exacerbates the effect of family background for students' test scores. I show, however, that slight changes in tracking variables, samples, or specifications renders the results of the relevant cross-country studies insignificant.

Understanding whether ability tracking intensifies the impact of parental background is important not only for educational researchers but also for policy makers. Countries that would like to increase educational equity should delay tracking if it increases the importance of family background for educational outcomes. Delaying the tracking age, which is often suggested as a way to reduce the link between parental background and educational outcomes, has been an important aspect of school reform in a number of different countries. Since the 1960s, the UK, some Scandinavian countries, Spain, and most recently Poland delayed their tracking age. If tracking, however, does not intensify the importance of family background, any such move would induce large costs without benefit. Careful studies on the effect of tracking are therefore needed to understand its effect on educational inequality.

The remainder of this paper is organized as follows. Section 2 provides a review of the relevant literature. Section 3 presents the identification strategy used to identify the causal effect of tracking. Section 4 discusses the data coming from a number of cross-country school studies: PIRLS, PISA and TIMSS. Section 5 presents the main regression results for reading and mathematics test scores separately. Section 6 probes the sensitivity of my findings and Section 7 discusses these results and concludes.

2 Previous Research

There are a number of studies that investigate the effect of tracking on educational equity. I classify these studies into three broad categories: those using educational reforms of the tracking grade within a country, those using existing within country variation in tracking,

[4] It is worthy to note that this method does not control for anticipation effects of parents or students which may affect the importance of family background already before tracking has taken place.

and those using cross-country differences in tracking regimes to identify the effect of tracking.

A widely studied educational reform is the abolishment of ability tracking in Britain during the 1960s and 1970s (see for example Galindo-Rueda and Vignoles (2005)). In spite of the attention paid to this reform, Pischke and Manning (2006) show that no existing study has successfully solved selection effects into tracked and untracked schools during the transition period from a selective to a comprehensive schooling system.

Another strand of the literature uses existing within-country variation in tracking to identify its effect on educational equity. Using data mostly from the United States, the general finding is that there are at most small effects of tracking on educational equity.[5] Papers exploiting within-country variation of tracking, however, suffer from three main problems. First, the tracking measure is based on teachers', headteacher's or students' judgements about whether a school tracks students. It is not clear whether these measures for tracking are treated consistently across schools. A second problem with these studies is that unobserved factors that affect test scores (student motivation, for example) could be correlated with attending a tracked or an untracked school, producing biased results. A third problem is that tracking regimes have little variation within a country; there is much more variation in tracking regimes across different countries.

These problems have led a number of researchers to use cross-country variation in tracking to identify the effect of tracking. The reasoning is that data from different schooling systems may alleviate the problem of having a noisy tracking measure. Furthermore, one may deal with the second problem stated above, as it is unlikely that individual student unobservables are correlated with a country level measure for tracking. These advantages, however, come at the expense of another omitted variable bias problem. It could be the case that the cross-country variation in tracking is correlated with country-level unobservables, which influence test scores.

The increasing availability of data with internationally comparable test scores has triggered research on the effect of tracking on students' test scores in the very recent past.[6] Hanushek and Woessmann (2006) carry out an analysis which uses cross-country variation in tracking to identify its effect on the within-country variance of educational test scores. They address the problem of unobserved country level variables by using a DiD approach. Their findings "provide (...) reasonably strong support for the disequalizing effects of early tracking." One of the main problems of their study, however, is the use of country-level data of the dispersion of test scores, since this leads to extremely small sample sizes. Furthermore, I show below that using a different measure for the tracking regime renders their results insignificant. Their results are further weakened by restricting their sample to OECD countries, only.

[5] See Betts and Shkolnik (1998), Rees, Brewer and Argys (2000) and Figlio and Page (2000) for studies which try to control for sorting of students across tracked and untracked schools.

[6] Brunello and Checchi (2006) investigate the effect of tracking for outcomes later in life and find that tracking does not affect the importance of family background for reading literacy of adults. They find, however, that tracking exacerbates the effect of parental background for wages later in life.

There are two recent studies which are more closely related to the research presented here, as they are trying to investigate whether tracking affects the impact of the family background on students' test scores. Schuetz, Ursprung, and Woessmann (2005) interact a measure for a country's tracking system with the students family background to see whether tracking changes the impact of parental background for educational achievement. Using data on secondary school students tested in TIMSS they find that "the family-background effect is larger...the earlier a country tracks its students into different school types by ability." They try to control for confounding country-level factors by including some controls for other institutional features of the education system. It is virtually impossible, however, to control for all unobserved country level factors that are correlated with the tracking measure and affect the importance of parental background on students' test scores. Any unobserved factor may bias the results, casting doubt on the findings of studies that do not control for pre-tracking differences in the importance of parental background for students' test scores.

The second paper looking explicitly at the importance of the socioeconomic background is a study by Ammermueller (2005). Using PISA 2000 and PIRLS data, he finds that "(T)he social origin of students ... increases its effect on student performance in countries with a differentiated schooling system...". Ammermueller's tracking measure uses the number of tracks in a schooling system. This measure does not capture the differential timing of tracking across different countries, which should have an impact on the influence of tracking. In checking the robustness of his results, I show that using a slightly different specification I do not find an effect of tracking even using Ammermueller's tracking measure.

3 Identification

Like recent research on ability tracking, I exploit the fact that different countries have different tracking policies. I use cross-country variation in tracking to identify the effect of tracking on the importance of parental background for children's test scores. One strategy to investigate whether tracking exacerbates the impact of parental background for educational achievement would be to estimate a standard education production function adding an interaction term of the parental background variables with the tracking regime. This is the approach used by Schuetz, Ursprung, and Woessmann (2005) and Brunello and Checchi (2006). They estimate an equation similar to equation (1):

$$(1) \qquad T_{isc} = \beta_1 + \beta_2 ST_{isc} + \beta_3 F_{isc} + \gamma_1 (F_{isc} * ET_c) + \beta_4 SQ_{sc} + \beta_5 C_c + \varepsilon_{isc}$$

Where T is a test score for individual i in school s and country c. ST is a vector of individual characteristics of student i in school s and country c. F is a vector of family characteristics of student i in school s and country c. ET is a dummy variable indicating whether a country tracks students at an early stage of their student life. SQ is a vector of school quality variables of school s in country c. C is a vector of country dummies. is an error term. The main interest then lies in obtaining consistent estimates of ; the interaction of the family background measures with the tracking regime. Finding to be positive would then be evidence that parental background is more important in early tracking countries.

As in most cross-country studies, the major concern is that any country level variable (in this case the tracking measure) is correlated with unobserved country level variables. This is not an issue as long as these variables do not affect the importance of parental background for students' test scores because the regression includes country fixed effects. It is quite likely, however, that some country level unobservables affect the influence of family background on test scores, like for example pre-primary care. The coverage of pre-primary care may be correlated with the tracking regime. At the same time, pre-primary care may affect the impact of the family background on students' educational achievement and thus biasing the estimates of in equation (1). The graphical analysis presented in the introduction already indicates that unobserved variables may pose a severe problem in this context. I showed above that parental background seems to be more important in early tracking countries even before actual tracking has taken place.

To solve this problem, I use a difference-in-differences strategy, with test scores taken at two points of a child's educational career. The first point is in primary school, before tracking has taken place in any of the countries in the sample. The second is in secondary school, after tracking has occurred in some countries. I then compare the change between the early and the late test in the importance of family background in early versus late tracking countries.

This is a legitimate strategy to control for unobserved country level variables under the identifying assumption that the unobserved country characteristics do not change between the primary and secondary school grades.[7] In a regression framework this difference-in-differences methodology is implemented by estimating equation (2).[8]

$$
\begin{aligned}
(2) \qquad T_{isct} = {} & \beta_1 + \beta_2 ST_{isct} + \beta_3 F_{isct} + \beta_4 SECONDARY_t + \beta_5 SQ_{sct} + \beta_6 C_c \\
& + \gamma_1 (F_{isct} * ET_c) + \gamma_2 (F_{isct} * SECONDARY_t) \\
& + \gamma_3 (F_{isct} * ET_c * SECONDARY_t) + \varepsilon_{isct}
\end{aligned}
$$

The abbreviations are the same as for equation (1). The subscript t indicates the two cohorts: primary school students and secondary school cohorts. Compared to the first specification this specification now adds the dummy variable SECONDARY, which indicates that an observation is taken from the late test examining secondary school students. This variable controls for any systematic difference between the two test scores.[9] The main in-

[7] It is worthy to note that the differences are not taken for the same individual. Therefore, a further identifying assumption is that the random assignment of the studies ensured that they tested a representative sample.

[8] In results not reported here I also allow the student characteristics to vary across the two different studies. This does not affect the findings on γ_3.

[9] As both test scores may not have tested the same skills one has to assume that systematic differences

terest now lies in identifying the coefficient . If turns out to be positive family background becomes relatively more important in early tracking countries between primary and secondary school. If was negative family background becomes relatively less important in early tracking countries after actual tracking has taken place. This model is estimated using the pooled data from a study testing primary school students and a study testing secondary school students. To allow for arbitrary correlations of the error term for students within one country all standard errors are clustered at the country level. The following part describes the data I use to estimate equation (2).

4 Data

4.1 Data on Test Scores, Student Characteristics, Family Characteristics, and School Quality Variables

The data on test scores, student characteristics and family background originates from the microdatasets of three school studies. The data on reading skills comes from PISA and PIRLS, while the data on mathematics comes from TIMSS. I only use data from OECD countries because I want to compare countries with a similar development of the educational sector. Table 1 gives an overview of the countries in the different estimation samples.

Data for the Reading Results

To implement the DiD strategy, I use two reading test scores taken at different points in time of the students' school career. Firstly, I use test scores from PIRLS 2001, a study testing students before tracking has occurred in any country. This study was carried out by the International Association for the Evaluation of Educational Achievement (IEA) in the year 2001. PIRLS tested students in grade 4 of primary school, before tracking has taken place in any country of the sample. Each student obtained a test score that was scaled to have an international mean of 500 and a standard deviation of 100. In addition to testing the reading skills of primary school students, the students' parents were asked to provide information on the student and on family background characteristics. Every participating school had to provide information on class sizes and other measures of school quality.

Secondly, I use a test score from a study testing students after tracking occurred in some countries. For the reading results I use data from PISA in the years 2000 and 2003. Administered by the OECD, this study evaluated reading skills of 15-year-old students, usually attending grades 9 or 10 of secondary school. Like PIRLS, the PISA study was scaled to have an international mean of 500 and a standard deviation of 100.[10] The information on

in the two tests are uncorrelated with the family background measures and the tracking regime to obtain consistent estimates of γ_3.

[10]The scores obtainable in the database are plausible values, which are not exactly test scores. "They

the parental background variables and the student characteristics is taken from the student questionnaire; the information on school quality variables from a questionnaire which had to be answered by each participating school.

The DiD estimation requires a primary school and a secondary school test score for each country. Therefore I can only use observations from countries that participated in both the PISA 2003 and the PIRLS studies in the main specification for reading.[11] The second column of Table 1 lists all countries in this specification. Column 3 lists the countries used in a robustness check with data from PIRLS and PISA 2000.

Data for the Mathematics Results

To estimate the specification for the mathematics results I use data from TIMSS. Carried out by the IEA, TIMSS 1995 tested mathematics and science skills of two different age cohorts (one cohort of primary school students and one of secondary school students) across different countries. I use the test results of the primary school cohort (attending grades 3 or 4) as an early test score before tracking has occurred in any of the participating countries. TIMSS also tested the mathematics skills of secondary school students attending grades 7 and 8, which I use as the later test score after which tracking has taken place in the early tracking countries. Again, I complement the data on test scores with data from student questionnaires and from the answers to the school questionnaires. The countries in this sample are shown in column 4 of Table 1.

In 1999 the primary cohort was retested when they were in grades 7 or 8. As a robustness check for the mathematics results, I combine the TIMSS data from grades 3/4 of 1995 and grades 7/8 from 1999 as can be seen from the last column of Table 1.

Parental Background Variables Used for Estimation Purposes

For my estimation I rely on the parental background data available in the above datasets. An important parental background factor for children's educational attainment is parental education. More educated parents may be better informed about good parenting practices and may create a home environment, which stimulates the learning of their children. In the tables reported below, I use highest parental education, which is measured as the years or education of the parent with more education. To save space I do not report the specifications where I use father's and mother's education separately. The results are very similar and do not affect any of the conclusions drawn below.

are random numbers drawn from the distribution of scores that could be reasonably assigned to each individual...". Refer to the Technical Reports of PISA for more information.

[11] In PIRLS England and Scotland are sampled independently. PISA, however, did not sample them separately. To estimate my results I combine the information of these two countries. Furthermore, in PIRLS Canada is represented only by the provinces Ontario and Quebec whereas PISA sampled students from all Canadian provinces. In the reported results for the difference-in-differences specification I include the combined data on Great Britain and data on Canada. Discarding all observations from these two countries does only have a small effect on the estimated coefficients and does not affect the significance or insignificance of the results.

Furthermore, I use the number of books in a student's home as an alternative measure for parental background. This variable is frequently used by educational researchers, and may capture parents' valuation of education or serve as a proxy for income as books are consumption goods.[12] Finally, I also use a variable indicating whether children speak the language of school instruction at home as another measure of parental background.

4.2 Data on Tracking

I complement the data on test scores, student characteristics, parental background, and school quality variables with a country level measure of the tracking regime. Some countries educate students in the same type of school up to the end of secondary school. Other countries, however, separate students at early stages of their schooling career. Figure 5 demonstrates the tracking policy of a representative late tracking country, Finland, and a representative early tracking country, Austria. Whereas Finland does not track students up to the end of grade nine, students in Austria are placed in different school tracks after four years of schooling.

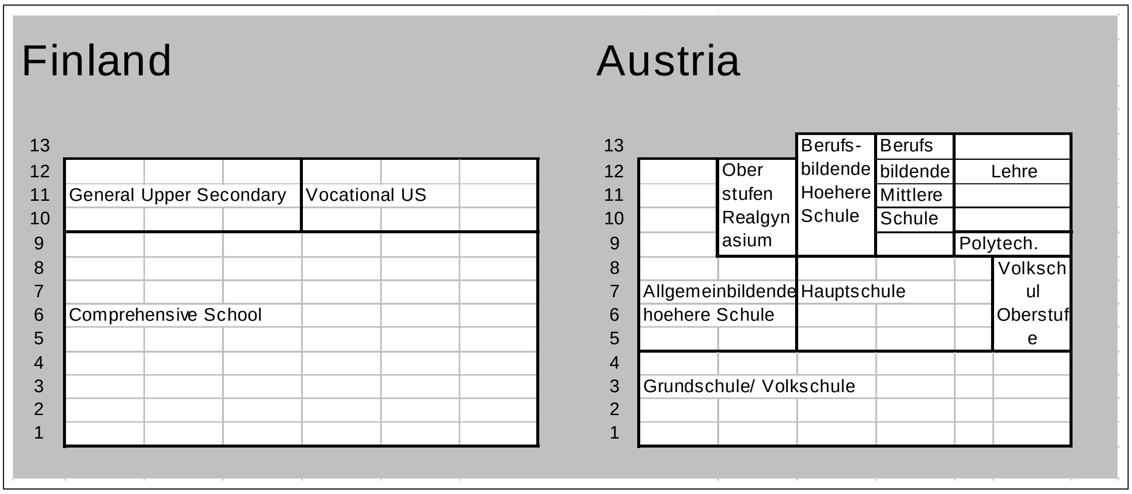

Figure 5: Tracking Policies in Finland and Austria.

I exploit the cross-country variation in tracking policies to identify the effect of tracking on the importance of family background for students' test scores. I construct a measure that indicates after how many grades tracking occurs in each of the countries of my sample. For the purposes of this paper I define tracking as educating students in different types of schools. This constitutes the strongest form of tracking. Therefore, I do not consider specialization tracks for certain subjects within a school as tracking. Instead, I focus on a very strong definition because the effects of tracking are potentially largest for this stark form of tracking. Furthermore, other definitions would result in a much more noisy tracking

[12]See Schuetz et al. (2005) for a discussion on using the number of books as a measure for parental background.

measure.[13] Table 2 gives an overview of the grade after which tracking takes place in the countries of my sample.

My tracking measure is very similar to the one developed by Hanushek and Woessmann (2006).[14] Their measure captures the age at which tracking takes place. Mine, on the other hand, relates to the grade after which tracking takes place, which will capture differences in the school starting age across countries. They classify some countries differently to my classification, yet a comparison of the two measures indicates that they are quite similar, which is indicated by the correlation of .81 between the two measures as reported in Table 3. I show, in fact, that my results hold up when I use the Hanushek and Woessmann tracking measure.

Alternatively, Ammermueller (2005) uses a tracking measure that counts the number of school tracks of a country in lower secondary education. This does not capture how many years a student is exposed to ability tracking, which is a disadvantage as it is likely that possible effects of tracking are stronger the longer students are educated in different tracks. Nevertheless, this measure has an advantage over measures using tracking grade or tracking age, as it captures how many school tracks exist in each country. By construction, higher values of the number of school track index indicate more tracking. According to my measure, however, a higher tracking grade indicates less tracking. Therefore the two measures should be negatively correlated. Table 3 confirms a negative correlation of the two measures of -0.75. I show below that my results hold even if I use the number of tracks measure.

To estimate equation (2), I define an early tracking threshold that indicates whether a country tracks students early or late. I vary the early tracking threshold between tracking after grade 4 and tracking after grade 8. I show below that my conclusions do not vary if I use different threshold levels for early tracking, indicating that the results are not driven by an arbitrary choice of the early tracking threshold.

For the countries with later tracking it is sometimes problematic to define the correct tracking grade. To overcome this problem, I use a dummy variable early versus late tracking. This provides a good measure because it is not affected by whether a later tracking country tracks at grade 9 or grade 10. The measurement problems are much less severe for the early tracking countries, as most of these have very clear tracking rules. In particular, the grade 4 to grade 6 thresholds provide measures of tracking, which should be free from major measurement error. I use these tracking measures to identify the effect of tracking on the importance of parental background for educational attainment, with results presented in the following section.

[13]Some countries exhibit within-country variation of tracking; with some regions tracking students at different grades than others. Unfortunately, I cannot distinguish different regions in the PISA, PIRLS and TIMSS databases. I therefore assign the grade at which most regions track as the tracking measure for these countries.

[14]I thank Ludger Woessmann for kindly providing me with their measure of tracking.

5 Main Results

5.1 Reading Results

I use data from PIRLS as the early test score (grade 4 students) and data from PISA 2003 (grade 9 or 10 students) as test score of secondary school students to estimate the reading results. Pooling the data from these two educational studies, I estimate equation (2), and present results in Table 4.[15] The specification used to generate these results takes grade 5 as the early tracking threshold: countries which track at the end of grade 5 or before are classified as early tracking countries and countries tracking at later stages of a student's career are classified as late tracking countries. I show below that the results are robust to different definitions of the early tracking thresholds, such as grade 4, grade 6 or grade 8.

To summarize the results, the coefficients on the student characteristics have the expected signs and are all highly significant in all specifications. Females do significantly better than males in reading, and native pupils do better than immigrants. Not surprisingly, children living with only one of their parents or those who live without their parents (because they live with other guardians) do significantly worse than students who live with both parents.

In column (1) I report the specification that uses highest parental education as the sole parental background variable. Students whose better-educated parent has one more year of education score about 6 points better in reading. Interestingly, the importance of parental education for reading is as important for primary school students as it is for secondary school students. This is verified by inspecting the very small and insignificant coefficient on the interaction of parental education with a dummy for the secondary school test. The positive and significant interaction of parental education and an indicator for early tracking countries, on the other hand, indicates that parental education is more important in early tracking countries. The point estimate shows that an increase of one year of parental education (for the better educated parent) increases the student's reading test score by about 5 points more in early tracking countries compared to countries with later tracking grades. This coefficient just indicates that parental education is more important in early tracking countries. Whether early tracking itself is causing parental education to be more important or whether other unobserved country level factors, that are correlated with the tracking regime, play a role has to be investigated in greater detail. To assess this directly, I include the triple interaction of parental education, an indicator for early tracking, and a dummy for the secondary school test into equation (2). To make this variable more visible in the tables it is reported in bold characters. The negative and insignificant coefficient on this interaction indicates that the importance of parental education for students' test scores does not increase between the grades tested in PIRLS and PISA in early tracking

[15] All the specifications reported in the table include school quality variables. Excluding school quality increases the sample in each country because of fewer missing values and also allows to keep France in the estimation sample. Results from these regressions are reported in Table A1 of the appendix.

countries. This suggests that parental background does not become more important after actual tracking has taken place.

Column (2) of Table 4 reports the results from a specification using the number of books in the students' home as the relevant family background variable. The coefficient on the number of books variable indicates that an increase of 100 books in the student's home increases her test score by 13 points.[16] The results on the interaction of the number of books variable with the secondary school dummy indicates that the number of books in a student's home is more important for secondary school students than for primary school students. The positive and significant interaction between the number of books variable and the indicator for early tracking countries demonstrates that the number of books in a student's home is more important in early tracking countries. The fact that the triple interaction of the number of books, the indicator for early tracking, and the dummy for the secondary school test is negative and insignificant shows that the number of books in a student's home do not become more important in early tracking countries after actual tracking has taken place.

In column (3) I report the specification using an indicator whether the student speaks the test language with his parents. Speaking the test language at home is a significant and important factor of doing well in reading. The interaction of the language indicator with the secondary school dummy shows that this factor is less important in secondary school compared to primary school. This is not surprising given that secondary school students have many more opportunities to speak the language of the country they live in with people other than their parents than primary school students. The positive and significant interaction of speaking the test language and the indicator for early tracking countries shows that speaking the test language is more important in these countries compared to countries which track later. Once again, however, the triple interaction of speaking the test language, the early tracking dummy, and an indicator for the secondary school test is negative and insignificant.

Including all parental background measures in the same specification does not affect the conclusions drawn before. The results reported in column (4) indicate that the importance of parental background does not increase in early tracking countries after actual tracking has taken place. These results therefore cast serious doubt on interpreting the correlation of tracking and the importance of the family background as causal.

[16] The number of books variable is coded in 6 categories. I assigned each student the category midpoint as the number of books at her home. This imposes some restrictions on the functional form of the relationship between the number of books and test scores. Reestimating this specification by using dummy variables for each book category gives similar results, namely that the number of books is significantly more important in early tracking countries. It does not become more important after actual tracking has taken place, however. For the sake of clear exposition I only report the results from the continuous number of books measure.

5.2 Mathematics Results

To further investigate the role of tracking in intensifying the importance of parental background, I present results with mathematics scores as the outcome. Unfortunately, TIMSS 1995 did not assess parental education for the students in the primary school cohort. Therefore, I only report results for the number of books and the language measures of parental background.[17]

Table 5 presents the results from estimating equation (2) using the TIMSS data. The coefficients on the student characteristics all have the expected signs and are similar to the ones obtained for the reading test score. The main differences are that being native matters less for mathematics, and that girls do worse in mathematics.

In the first column I use the number of books in a pupil's home as the relevant parental background measure, which appears to be an important factor affecting student performance. Increasing the number of books by 100 increases a student's test score by about 15 points. The positive and significant coefficient on the interaction of the number of books variable and an indicator for the secondary school cohort indicates that the number of books are more important in later stages of a student's career. The coefficient of interest, the triple interaction, is negative and marginally significant, indicating that family background becomes less important in early tracking countries after tracking has taken place.

The specifications reported in column (2) of Table 5 use an indicator for speaking the test language at home as the relevant parental background measure. Speaking the test language at home has a strong and significant effect on mathematics test scores, but seems to be less important in secondary school. The coefficient on the interaction of the language indicator and the dummy for the secondary school cohort, however, is not significant. Speaking the test language at home seems to be more important in early tracking countries but also the coefficient on the interaction of the language dummy with the indicator for early tracking countries is not significantly different from 0. The coefficient on the triple interaction again is not significantly different from 0 and has a point estimate which is very close to 0. Speaking the test language at home does not become more important in early tracking countries after actual tracking has taken place.

Including both family background measures at the same time does not affect any of the above conclusions as can be seen be inspecting the results reported in column (3) of Table 5.

The results for mathematics confirm the reading results presented before. There is clearly no support for the hypothesis that tracking exacerbates the importance of family background after actual tracking has taken place.

[17] Again these results are reported for the sample including the school quality controls. The results without school quality data which result in a larger sample with more countries are reported in Table A2 in the appendix. As can be seen, the results do not differ in any significant way.

6 Sensitivity Analysis

These results run contrary to the findings of the existing studies using cross-country data to investigate the impact of tracking on educational equality. This section therefore investigates the robustness of my findings. My first test uses different thresholds for the grade identifying early versus late tracking countries. The results of these regressions are reported in Tables 6 and 7 for the reading and mathematics test scores respectively. To save space I only report the results for the specification using all parental background variables in the same regression.[18] For the reading test score this is the equivalent to the specification reported in column (4) of Table 4. Table 6 shows that the coefficient on the triple interaction term of family background, early tracking, and the dummy indicating whether the observation comes from the secondary school (PISA) sample is hardly ever significantly different from zero. The only exception is the specification using grade 4 as the early tracking threshold. In this case the coefficient on the triple interaction for parental education is marginally significant but with a negative coefficient. This would indicate that parental education becomes less important in early tracking countries after actual tracking has taken place. The results reported in Table 6 do not support the hypothesis that tracking exacerbates the importance of family background for students' test scores.

Table 7 reports the results for the mathematics test score. Using the end of grade 4 as the threshold for early tracking, the coefficients of the triple interactions indicate that family background becomes more important in early tracking countries between grades 3/4 and grades 7/8. Using any other grade as the threshold to divide the sample into early versus late tracking countries I find no evidence that the importance of family background increases between grades 3/4 and grades 7/8.

As mentioned above, there is more than one way to measure the extent of tracking. Hanushek & Woessmann and Schuetz et al., for example, use the age at which tracking takes place as their measure. Ammermueller defines tracking slightly differently by the number of school tracks that exist during secondary school in a certain country. Table 8 compares the results from using my tracking measure (grade after which tracking occurs), and the other two defined in this paragraph. The triple interactions of the family background variable, the tracking measure, and the dummy indicating whether the observation is from the secondary grades are mostly not significantly different from zero, independent of the tracking measure used in the estimation. The results presented in column 2 of Table 8 show that the number of books at home is relatively less important in countries with later tracking ages in grades 9/10 compared to grade 4. This result, however, is only significant at the 10 percent level. Speaking the test language seems to be less important in later grades in countries with more tracks (thus more tracking). Again, this coefficient is only significant at the 10 percent level. These two marginally significant triple interactions propose conflicting conclusions on whether more tracking increases or reduces the importance of family background for

[18] Using the other specifications does not affect the conclusions drawn from this exercise.

children's reading test scores. All other coefficients on the triple interaction terms are not significantly different from zero and indicate no effect of tracking after tracking has taken place.

For the mathematics test score, with results reported in Table 9, none of the coefficients on the triple interaction terms is significantly different from zero. Furthermore, the point estimate of these coefficients is always close to 0. There is no evidence that tracking influences the importance of parental background for educational attainment after tracking has taken place.

Another worry may be that factors varying at the country level and affecting the importance of the family background are not constant over time, and would thus not be controlled using the fixed effects strategy. Using tests administered in alternative years may address this problem. Because PISA 2003 and PIRLS 2001 were not carried out in the same year, there may have been changes in the schooling systems affecting the importance of the family background on students' test scores between 2001 and 2003. Unfortunately there are no comparable international pupil ability tests testing reading skills of different cohorts in the same year. Nonetheless, using the PISA 2000 data instead of the 2003 wave of PISA may be a suitable way to deal with this problem. This is because changes between 2001 and 2003 will not have had an effect on the PISA 2000 results.

Table 10 reports the results from the same specifications as reported in Table 4 but using PISA 2000 and PIRLS 2001 as the data sources. The coefficients on the triple interactions show that there is no evidence that the importance of family background increases relatively more in early tracking countries between grade 4 and grades 9/10.

The DiD results for the mathematics test score presented above were estimated using the 1995 wave of TIMSS, which tested students of both the grades 3/4 and grades 7/8 cohorts. Nonetheless, these results may be problematic because the factors affecting the importance of family background for mathematics test scores may have changed in a year shortly before 1995. Students from the two cohorts may have been exposed to this changing environment at a different age. I use the 1999 wave of TIMSS which tested students in grades 7 or 8 to address this potential problem. Table 11 reports the results from this exercise. Note that the number of countries which can be used for estimation purposes is very small, so any conclusion drawn from this exercise should be handled with caution.[19] Again, the triple interactions are almost always insignificant. The only exception is the interaction involving

[19] I use a different set of student characteristics compared to all other regressions presented in this paper to estimate the specification reported in Table 11. In all other regressions I was using the same set of student characteristics: a gender dummy, age, a dummy whether the student was born in the country and dummies indicating whether the student lives with a singleparent or without parents. If I did include the dummies indicating the family structure in this estimation the sample would be reduced by about 30 percent and I would lose 3 out of 8 countries because 3 countries did not report the family structure variables. This affects the results in a non-random way. Therefore I omit the two dummies indicating whether the student lives with a singleparent or without parents. Table A3 shows that omitting these 2 variables does not have an impact on the results obtained from estimating the difference-in-differences results for the original TIMSS 1995 sample as can be seen from comparing the results presented in Table A3 and Table 5. This makes me confident that omitting these two variables is the preferred option for having a more representative sample.

the language spoken at home if considered as the only family background variable which is reported in column 2. In this case, the coefficient indicates that family background becomes relatively more important in early tracking countries between grades 3/4 and grades 7/8.

7 Discussion and Conclusion

The results obtained from estimating equation (2) show that family background is more important in early tracking countries. I find, however, that family background does not become more important after tracking occurs in the early tracking countries. In the absence of anticipation effects this is evidence that tracking does not causally affect the importance of family background for educational attainment. While this result is interesting in itself, it is even more interesting that at the same time, two other papers find that tracking exacerbates educational inequalities using the some of the datasets I use in this paper. One study was carried out my Hanushek and Woessmann (2006) and one by Ammermueller (2005). I discuss the likely reasons for obtaining different results in turn.

Hanushek and Woessmann do not look at the specific effect of family background on test scores but look at whether tracking increases inequality measured by the within country standard deviation, the 75th-25th or the 95th-5th percentile differences of test scores. To address the concern of bias due to unobservables, they employ a similar DiD strategy to the one in this paper: they look at the change in inequality between a primary school test score (no tracking in any country) and a test score for secondary school students (after tracking in some countries). In their main specification they use data on reading from the same data sources I use for my main specification on reading test scores. In order to assess the reasons of obtaining different findings, I replicate their results and show that these results are not stable if one considers a different measure for tracking and a slightly different sample. Column (1) of Table 12 reports the second column of their main table (Table 2 in Hanushek and Woessmann (2006)).[20] They regress the within country standard deviation (appropriately normalized) for secondary school students on an indicator for early tracking. As a measure for tracking they use a threshold of tracking before the age of 15. To control for pre-existing levels of inequality they also include the within-country standard deviation of test scores for primary school students. They conclude that inequality is higher for early tracking countries based on a positive and significant coefficient on the early tracking dummy as reported in the first column of Table 12. In column (2) of Table 12, I show, however, that using the same sample but a different tracking measure (here tracking at the end of grade 5 or before) the results are no longer significantly different from 0. Furthermore, using a sample of OECD countries only, which reduces the sample by only three countries, further weakens their results. Using small variations in their specification I can, therefore, no longer

[20] I thank Ludger Woessmann for providing me with details on the way they estimate their results.

replicate their findings and obtain results which are in line with the results presented in this paper. Namely, that tracking does not increase educational inequality.

My results are also slightly different from the ones obtained by Ammermueller (2005) even though the identification strategies used are very similar and he uses data from PISA and PIRLS as well. He finds that the number of books at a students home do not become more important in tracking systems with more tracks after actual tracking has taken place which is the same result as I find. He finds that speaking the test language becomes less important in countries with more tracks (and thus more tracking). This is similar to the finding on speaking the test language in this paper. On the other hand, he finds that students with native parents do better in tracked countries after actual tracking has taken place. Furthermore, he finds that the importance of parental education increases in countries with more tracking. Therefore 2 of his 4 results indicate that tracking does not affect or reduces the importance of family background for students' test scores. The results on the other 2 family background variables indicate that tracking exacerbates the importance of family background.

How can I reconcile this with my findings? Ammermueller uses a different definition for parental education, namely having a parent with a university degree whereas I use a linear years of education measure. Furthermore, he drops 2 countries (Canada and England) because not all regions in those countries were sampled. Reestimating my results with this reduced sample does not change the conclusion from my paper: tracking does not intensify the role of family background for students' education. The factor which probably explains a big part of the difference between our papers is that I cluster my standard errors at the country level, while he is clustering the standard errors at the school level. Clustering at the country level allows for any arbitrary correlation of the error terms of students within one country. This is the level of clustering which is appropriate in this setting. Clustering at the school level leads to much smaller standard errors, and thus to significant coefficients which may well be insignificant if one clusters at the country level.

My results clearly indicate that the estimates in the existing literature are not stable to using slightly different tracking measures, samples, and specifications. This further increases the confidence in my finding that tracking does not exacerbate the importance of parental background for educational test scores. The question remains why tracking does not affect the importance of family background. As mentioned above, parents have the possibility to self select according to socioeconomic status even in untracked systems, so one path may be through residential segregation. This is exemplified by the literature on the effect of school quality on house prices which mainly uses evidence from the US and the UK - both untracked countries.[21]

Another way of improving a child's education even in untracked systems is to enroll the child in a better and possibly expensive private school. Furthermore, choosing certain

[21] See Black (1999) and Figlio and Lucas (2004) for careful studies of the effect of school quality on house prices.

subjects (e.g. learning ancient languages) may lead to better peers and thus better learning even in a system which has no official tracking policy. Thus even untracked systems give parents and students opportunities to select into better schools. Further research is needed to pin down the factors which enable parents to improve their child's education and may thus create inequalities in opportunity for children from different backgrounds. Nonetheless, my findings could be consistent with the view that tracking has a causal impact on the importance of family background for children's test scores if the above mentioned effects of tracking happened before actual tracking takes place. I use studies which tested students in grades 3 or 4 as a baseline level of the role of parental background. This is just a short time before some of the early tracking countries sort students according to ability. Parents may anticipate the effect of tracking and motivate their children to do better in school already before tracking takes place. My DiD methodology may fail to pick up these anticipatory effects of tracking. Given that my results are independent of the tracking threshold I choose may be an indication that anticipation effects are not very important. Research strategies which circumvent this problem may shed more light on the debate of the effect of tracking. I conclude, however, that the cross-country evidence to date, if carried out carefully, does not suggest that tracking exacerbates the importance of family background for students' test scores.

References

[1] **Ammermueller, Andreas (2005)** "Educational Opportunities and the Role of Institutions", *ZEW Discussion Paper*, No. 05-44

[2] **Argys, Laura, Daniel Rees and Dominic Brewer (1996)**, "Detracking America's Schools: Equity at Zero Cost?", *Journal of Policy Analysis and Management*, Vol. 15, No. 4, 623-645

[3] **Betts, Julian and Jamie Sholnik (1998)** "The Effects of Ability Grouping on Student Math Achievement and Resource Allocation in Secondary Schools.", *University of San Diego Working Paper Series*, Economics Working Paper, 96-25R

[4] **Black, Sandra (1999)** "Do Better School Matter? Parental Valuation of Elementary Ecucation", *The Quarterly Journal of Economics*, Vol. 114, No. 2

[5] **Brunello, Giorgio and Daniele Checchi (2006)**, "Does School Tracking Affect Equality of Opportunity? New International Evidence", *IZA Discussion Paper,* No. 2348

[6] **Figlio, David and Maurice Lucas (2004)**, "What's in a Grade? School Report Cards and the Housing Market", *The American Economic Review*, Vol. 94, No. 3

[7] **Figlio, David and Marianne Page (2000)** "School Choice and the Distributional Effects of Ability Tracking: Does Separation Increase Equality?", *NBER Working Paper 8055*

[8] **Galindo-Rueda, Fernando and Anna Vignoles (2005)** "The heterogeneous effect of selection in secondary schools: Understanding the changing role of ability", *CEE Discussion Paper*, No. 52

[9] **Hanushek, Eric and Ludger Woessmann (2006)** "Does Educational Tracking Affect Performance and Inequality? Differences-In-Differences Evidence Across Countries", *Economic Journal,* Vol. 116

[10] **Henderson, Vernon, Mieszkowski, Peter and Sauvageau, Yvon (1978)** "Peer Group Effects and Educational Production Functions", *Journal of Public Economics*, 10

[11] **IEA (2003)** "PIRLS 2001 Technical Report", *IEA*

[12] **Pischke, Jörn-Steffen and Alan Manning (2006)** "Comprehensive versus Selective Schooling in England in Wales: What Do We Know?", *NBER Working Paper*, No. W12176

[13] **OECD (2004)** "Learning for Tomorrow's World - First Results from PISA 2003", *OECD*

[14] **OECD (2005)** "PISA 2003 Technical Report", *OECD Publishing*

[15] **Rees, Daniel, Dominic Brewer and Laura Argys (2000)**, "How should we measure the effect of ability grouping on student performance", *Economics of Education Review*, Vol. 19, 17-20

[16] **Schuetz, Gabriela, Heinrich Ursprung and Ludger Woessmann (2005)**, "Education Policy and Equality of Opportunity", CESIFO Working Paper No. 1518

[17] **Summers, Anita and Barbara Wolfe (1977)** "Do Schools Make a Difference?", *The American Economic Review*, Vol. 67, No. 4

[18] **UNESCO (1997)** "International Standard Standard Classification of Education", *UNESCO*

[19] **Vandenberghe, Vincent (2005)** "Achievement Effectiveness and Equity - The Role of Tracking, Grade Repetition and Inter-school Segregation", *forthcoming Applied Economic Letters*

8 Tables

Table 1: Countries in the different Estimation Samples

Secondary School Test Primary School Test	PISA 2003 PIRLS[1]	PISA 2000 PIRLS	TIMSS 1995 7/8[2,3] + TIMSS 1995 3/4[4]	TIMSS 1999 7/8 + TIMSS 1995 3/4
	Main Specification Reading	Robustness Check Reading	Main Specification Mathematics	Robustness Check Mathematics
Australia			■□	■
Austria			□	
Canada	■□	■	■□	■
Czech Republic	■□	■	■□	■
France	□	■		
Germany	■□	■		
Greece	■□	■	□	
Hungary	■□		■□	■
Iceland	■□	■	■□	
Ireland			■□	
Italy	■□	■		
Korea			■□	■
Netherlands	■□	■	□	
New Zealand	■□	■	■□	■
Norway	■□	■	□	
Portugal			■□	
Slovakia	■□			
Sweden	■□	■		
Turkey	■□			
United Kingdom	■□	■	■□	■
USA			■□	■
Number of Countries	14/15	12	11/15	8

■ including school quality □ without school quality

[1]The US drops out because the parental background information is missing for all students in PIRLS. [2]Also France and Japan participated in TIMSS. As the information on being born in the country is missing for all observations they are not in my sample. [3]School quality is missing for all observations from Austria, Greece and Norway. [4]The information on school quality for the Netherlands is missing for the Grade 4 wave.

Table 2: Countries in the PISA 2003 and TIMSS 95 (Secondary) Samples, Tracking Grade, and Means of Relevant Variables

Tracking Grade	Country	Mean Test Score Reading PISA 2003	Mean Parental Education PISA 2003	Mean Number of Books PISA 2003	Speaking Test Language (%) PISA 2003	Mean Test Score Maths TIMSS 95-8	Mean Parental Education TIMSS 95-8	Mean Number of Books TIMSS 95-8	Speaking Test Language (%) TIMSS 95-8
4	Austria	497.4	13.1	170.2	91.5	535.3	12.4	136.9	88.5
4	Germany	497.5	13.1	206.7	92.8	498.5	10.6	140.3	88.4
4	Hungary	480.7	12.9	243.7	99.3	522.0	12.5	178.4	98.7
4	Slovakia	475.2	13.8	161.3	97.0	527.8	12.1	124.7	88.9
5	Czech Republic	505.4	13.9	251.9	99.1	538.7	11.8	164.3	98.5
6	Ireland	517.6	12.5	158.4	97.4	515.3	11.5	124.4	97.5
6	Netherlands	518.0	12.8	181.1	85.4	534.8	12.0	129.8	90.2
6	New Zealand	523.7	13.5	205.7	90.7	493.3	12.7	178.7	90.8
8	Canada	516.1	14.3	209.0	89.3	505.2	13.2	155.1	82.9
8	Italy	500.3	12.7	178.1	73.9				
8	Turkey	444.0	9.0	82.9	97.3				
9	France	500.1	12.0	162.9	93.9	518.1	11.5	118.0	94.7
9	Greece	467.2	13.0	140.3	96.3	464.0	10.7	95.7	95.5
9	Korea	533.4	12.4	172.1	99.9	592.4	11.8	127.7	95.3
9	Portugal	476.1	9.1	128.8	98.6	440.9	8.2	98.3	97.0
9	Sweden	513.2	13.5	236.6	92.7	513.7	12.9	177.2	91.3
10	Australia	524.1	13.1	232.6	91.9	524.3	12.3	181.7	92.2
10	Iceland	491.7	14.5	251.8	98.3	469.8	12.0	168.3	95.5
10	Norway	499.6	14.6	251.1	94.8	485.9	13.0	184.9	94.3
12	UK	512.0	12.9	175.5	96.9	485.0	12.0	133.8	94.2
12	USA	493.8	13.5	160.1	90.9	483.1	13.4	136.3	87.2

The differences in the means of the family background variables across PISA and TIMSS originate from the fact that the tests were carried out at different times, for students in different grades, and by different organizations with different questionnaires. The particularly big differences for the number of books variable can be explained by the fact that for PISA the top category for books is "more than 500 books", whereas for TIMSS the top category is "more than 200 books".

Table 3: Correlations of Tracking Measures

	Tracking Grade	Tracking Age	Number of Tracks
Tracking Grade	1		
Tracking Age	0.81	1	
Number of Tracks in Secondary School	-0.75	-0.69	1

Table 4: Difference-in-Differences Reading (PIRLS 2001 and PISA 2003)

Dependent Variable: Reading Test Score

	(1)	(2)	(3)	(4)
Highest Parental Education (years)	6.16			4.68
	(0.81)***			(0.62)***
HPE*PISA	0.14			-0.39
	(1.05)			(0.82)
HPE*Earlytrack	4.78			3.44
	(1.27)***			(1.29)**
HPE*Earlytrack*Secondary	-1.28			-1.24
	(0.73)			(1.25)
Number of Books		0.13		0.10
		(0.02)***		(0.01)***
NoB*PISA		0.06		0.06
		(0.02)**		(0.02)**
NoB*Earlytrack		0.05		-0.01
		(0.02)**		(0.02)
NoB*Earlytrack*Secondary		-0.03		0.04
		(0.02)		(0.03)
Language			38.35	33.51
			(4.23)***	(3.76)***
Language*PISA			-18.74	-20.98
			(6.72)**	(7.37)**
Language*Earlytrack			19.02	7.51
			(8.22)**	(5.37)
Language*Earlytrack*Secondary			-10.64	-6.93
			(7.73)	(11.19)
Female	28.33	25.49	26.44	25.97
	(1.27)***	(1.16)***	(1.28)***	(1.33)***
Age of Student	6.71	5.03	3.83	6.89
	(3.24)*	(3.29)	(3.39)	(2.90)**
Native Student	26.14	22.09	20.75	17.39
	(3.95)***	(4.25)***	(4.78)***	(4.18)***
Living with Singleparent	-11.55	-9.50	-14.08	-8.87
	(2.44)***	(2.37)***	(2.30)***	(2.19)***
Living without Parents	-39.84	-39.27	-46.56	-36.34
	(6.98)***	(5.32)***	(5.69)***	(6.11)***
Secondary School Test Dummy	✓	✓	✓	✓
School Quality	✓	✓	✓	✓
Country FE	✓	✓	✓	✓
Number of Countries	14	14	14	14
Observations	116366	118690	124826	113018
R-squared	0.18	0.20	0.14	0.22

Robust standard errors in parenthesis (clustered at the country level)
* significant at 10%   ** significant at 5%   *** significant at 1%

Table 5: Difference-in-Differences Mathematics
(TIMSS 1995 Grades 3/4 and TIMSS 1995 Grades 7/8)

Dependent Variable: Mathematics Test Score

	(1)	(2)	(3)
Number of Books	0.15		0.14
	(0.02)***		(0.02)***
NoB*Secondary	0.05		0.05
	(0.01)***		(0.01)***
NoB*Earlytrack	0.07		0.05
	(0.02)**		(0.02)**
NoB*Earlytrack*Secondary	-0.04		-0.01
	(0.02)*		(0.04)
Language		33.89	30.67
		(4.77)***	(4.23)***
Language*Secondary		-7.91	-12.00
		(7.56)	(6.86)
Language*Earlytrack		6.78	3.69
		(4.61)	(4.84)
Language*Earlytrack*Secondary		0.86	-1.54
		(8.50)	(5.04)
Female	-4.98	-4.81	-5.19
	(1.36)***	(1.53)**	(1.36)***
Age	14.5	13.51	14.82
	(3.56)***	(3.72)***	(3.39)***
Native	7.75	5.69	2.8
	(4.96)	(5.54)	(5.07)
Living with Singleparent	-12.31	-16.17	-12.47
	(2.01)***	(1.96)***	(1.97)***
Living without Parents	-27.61	-31.19	-25.82
	(4.66)***	(4.80)***	(4.47)***
Secondary School Test Dummy	✓	✓	✓
School Quality	✓	✓	✓
Country FE	✓	✓	✓
Number of Countries	11	11	11
Observations	102245	96873	95616
R-squared	0.20	0.16	0.21

Robust standard errors in parenthesis (clustered at the country level)
* significant at 10% ** significant at 5% *** significant at 1%

Table 6: Sensitivity Analysis 1: Early Tracking Thresholds (Reading)

Dependent Variable: Reading Test Score

	Grade 4	Grade 5	Grade 6	Grade 8
Highest Parental	4.88	4.68	4.73	5.57
Education (years)	(0.61)***	(0.62)***	(0.62)***	(0.48)***
HPE*Secondary	-0.40	-0.39	-0.45	-1.34
	(0.83)	(0.82)	(0.80)	(0.94)
HPE*Earlytrack	3.33	3.44	2.34	-0.36
	(1.74)*	(1.29)**	(1.15)*	(0.76)
HPE*Earlytrack*Secondary	-2.07	-1.24	-0.81	0.99
	(1.15)*	(1.25)	(0.99)	(0.84)
Number of Books	0.10	0.10	0.10	0.07
	(0.01)***	(0.01)***	(0.01)***	(0.01)***
NoB*Secondary	0.06	0.06	0.05	0.11
	(0.02)***	(0.02)**	(0.02)**	(0.02)***
NoB*Earlytrack	-0.02	-0.01	-0.00	0.04
	(0.02)	(0.02)	(0.02)	(0.02)**
NoB*Earlytrack*Secondary	0.05	0.04	0.04	-0.07
	(0.03)	(0.03)	(0.03)	(0.02)**
Language	34.30	33.51	34.10	37.15
	(3.48)***	(3.76)***	(3.63)***	(6.98)***
Language*Secondary	-22.61	-20.98	-22.26	-26.54
	(6.99)***	(7.37)**	(7.63)**	(10.68)**
Language*Earlytrack	7.26	7.51	4.59	-2.36
	(6.71)	(5.37)	(5.92)	(7.23)
Language*Earlytrack*Secondary	2.38	-6.93	-2.90	4.23
	(7.98)	(11.19)	(9.04)	(12.27)
Secondary School Test Dummy	✓	✓	✓	✓
Student Characteristics	✓	✓	✓	✓
School Quality	✓	✓	✓	✓
Country FE	✓	✓	✓	✓
Number of Countries	14	14	14	14
Observations	113018	113018	113018	113018
R-squared	0.22	0.22	0.22	0.22

Table 7: Sensitivity Analysis 1: Early Tracking Thresholds (Mathematics)

Dependent Variable: Mathematics Test Score

	Grade 4	Grade 5	Grade 6	Grade 8
Number of Books	0.15	0.14	0.14	0.17
	(0.02)***	(0.02)***	(0.02)***	(0.02)***
NoB*Secondary	0.04	0.05	0.04	0.05
	(0.01)**	(0.01)***	(0.01)***	(0.01)***
NoB*Earlytrack	0.03	0.05	0.03	-0.03
	(0.02)	(0.02)**	(0.03)	(0.03)
NoB*Earlytrack*Secondary	0.05	-0.01	0.01	-0.01
	(0.01)***	(0.04)	(0.03)	(0.02)
Language	31.20	30.67	29.46	34.12
	(3.90)***	(4.23)***	(4.45)***	(7.48)***
Language*Secondary	-12.85	-12	-12.33	-15.74
	(6.68)*	(6.86)	(7.07)	(10.3)
Language*Earlytrack	3.25	3.69	8.56	-5.91
	(5.99)	(4.84)	(5.92)	(7.80)
Language*Earlytrack*Secondary	20.15	-1.54	1.33	6.75
	(5.90)***	(5.04)	(6.15)	(7.97)
Secondary School Test Dummy	✓	✓	✓	✓
Student Characteristics	✓	✓	✓	✓
School Quality	✓	✓	✓	✓
Country FE	✓	✓	✓	✓
Number of Countries	11	11	11	11
Observations	95616	95616	95616	95616
R-squared	0.21	0.21	0.21	0.21

Robust standard errors in parenthesis (clustered at the country level)
* significant at 10% ** significant at 5% *** significant at 1%

Table 8: Sensitivity Analysis 2: Tracking Measures (Reading)

Dependent Variable: Reading Test Score

	Grade of Tracking	Age of Tracking	Number of Tracks
Highest Parental	4.68	9.71	5.25
Education (years)	(0.62)***	(2.77)***	(1.26)***
HPE*Secondary	-0.39	-2.36	-0.42
	(0.82)	(2.28)	(1.13)
HPE*Tracking Measure (TM)	3.44	-0.28	0.11
	(1.29)**	(0.17)	(0.72)
HPE*TM*Secondary	-1.24	0.11	-0.19
	(1.25)	(0.14)	(0.49)
Number of Books	0.10	0.06	0.10
	(0.01)***	(0.05)	(0.03)***
NoB*Secondary	0.06	0.18	0.02
	(0.02)**	(0.05)***	(0.04)
NoB*Tracking Measure (TM)	-0.01	0.00	0.00
	(0.02)	(0.00)	(0.01)
NoB*TM*Secondary	0.04	-0.01	0.02
	(0.03)	(0.00)*	(0.02)
Language	33.51	37.09	34.91
	(3.76)***	(14.38)**	(5.83)***
Language*Secondary	-20.98	-66.34	-3.12
	(7.37)**	(31.90)*	(10.53)
Language*TM	7.51	-0.17	-0.16
	(5.37)	(0.88)	(3.78)
Language*TM*Secondary	-6.93	2.80	-10.38
	(11.19)	(1.85)	(5.45)*
Secondary School Test Dummy	✓	✓	✓
Student Characteristics	✓	✓	✓
School Quality	✓	✓	✓
Country FE	✓	✓	✓
Number of Countries	14	14	14
Observations	113018	113018	113018
R-squared	0.22	0.22	0.22

Table 9: Sensitivity Analysis 2: Tracking Measures (Mathematics)

Dependent Variable: Mathematics Test Score

	Grade of Tracking	Age of Tracking	Number of Tracks
Number of Books	0.14	0.28	0.11
	(0.02)***	(0.05)***	(0.05)**
NoB*Secondary	0.05	0.04	0.04
	(0.01)***	(0.08)	(0.04)
NoB*Tracking Measure (TM)	0.05	-0.01	0.03
	(0.02)**	(0.00)**	(0.03)
NoB*TM*Secondary	-0.01	0.00	0.01
	(0.04)	(0.00)	(0.03)
Language	30.67	39.81	20.9
	(4.23)***	(16.09)**	(9.44)*
Language*Secondary	-12.00	-7.51	-13.66
	(6.86)	(22.45)	(10.54)
Language*Tracking Measure (TM)	3.69	-0.51	8.56
	(4.84)	(0.88)	(5.92)
Language*TM*Secondary	-1.54	-0.26	1.33
	(5.04)	(1.11)	(6.15)
Secondary School Test Dummy	✓	✓	✓
Student Characteristics	✓	✓	✓
School Quality	✓	✓	✓
Country FE	✓	✓	✓
Number of Countries	11	11	11
Observations	95616	95616	95616
R-squared	0.21	0.21	0.21

Robust standard errors in parenthesis (clustered at the country level)
* significant at 10% ** significant at 5% *** significant at 1%

Table 10: Sensitivity Analysis 3: Different Data
Reading: PISA 2000 and PIRLS 2001

Dependent Variable: Reading Test Score

	(1)	(3)	(4)	(5)
Highest Parental	5.99			4.67
Education (years)	(0.40)***			(0.35)***
HPE*Secondary	2.52			0.96
	(0.76)***			-0.74
HPE*Earlytrack	4.06			1.74
	(0.64)***			(0.26)***
HPE*Earlytrack*Secondary	0.13			1.69
	(0.56)			(1.18)
Number of Books		0.13		0.09
		(0.01)***		(0.01)***
NoB*Secondary		0.09		0.08
		(0.01)***		(0.02)***
NoB*Earlytrack		0.04		0.01
		(0.02)**		(0.01)*
NoB*Earlytrack*Secondary		-0.03		0.00
		(0.02)		(0.02)
Language			36.35	29.93
			(3.05)***	(3.37)***
Language*Secondary			6.17	-2.23
			(5.29)	(5.15)
Language*Earlytrack			14.11	11.67
			(11.25)	(7.18)
Language*Earlytrack*Secondary			-7.72	-28.73
			(5.87)	(14.49)*
Secondary School Test Dummy	✓	✓	✓	✓
Student Characteristics	✓	✓	✓	✓
School Quality	✓	✓	✓	✓
Country FE	✓	✓	✓	✓
Number of Countries	12	12	12	12
Observations	73108	76323	82500	71242
R-squared	0.18	0.19	0.12	0.23

Robust standard errors in parenthesis (clustered at the country level)
* significant at 10% ** significant at 5% *** significant at 1%

Table 11: Sensitivity Analysis 3: Different Data
Mathematics (TIMSS 1995 3/4 and TIMSS 1999 7/8)

Dependent Variable: Mathematics Test Score

	(1)	(2)	(3)
Number of Books	0.16		0.15
	(0.02)***		(0.02)***
NoB*Secondary	0.05		0.06
	(0.01)***		(0.01)***
NoB*Earlytrack	0.06		0.05
	(0.03)		(0.02)**
NoB*Earlytrack*Secondary	0.01		0.03
	(0.02)		(0.03)
Language		35.22	30.90
		(4.28)***	(3.71)***
Language*Secondary		-14.90	-17.40
		(10.77)	(9.74)
Language*Earlytrack		4.55	5.24
		(5.58)	(6.80)
Language*Earlytrack*Secondary		20.44	6.22
		(6.41)**	(5.14)
Secondary School Test Dummy	✓	✓	✓
Student Characteristics	✓	✓	✓
School Quality	✓	✓	✓
Country Dummies	✓	✓	✓
Number of Countries	8	8	8
Observations	60228	56436	55756
R-squared	0.19	0.15	0.21

Robust standard errors in parenthesis (clustered at the country level)
* significant at 10% ** significant at 5% *** significant at 1%

Table 12: Robustness of Hanushek and Woessmann Results
Dependent Variable: Country-Level Standard Deviation of Reading Test Score

	(1)[1]	(2)	(3)
Early Tracking Measure	Age of Tracking	Grade 5 or before	Grade 5 or before
Sample	Full Sample	Full Sample	OECD
Early Tracking	0.25	0.17	0.12
	(0.11)**	(0.14)	(0.16)
Inequality in Primary School	0.59	0.48	0.40
	(0.13)***	(0.16)***	(0.20)**
Number of Countries	18	18	15
R-squared	0.48	0.38	0.26

* significant at 10% ** significant at 5% *** significant at 1%
[1]Column (1) is the same as column (1) of Table 2 in Hanushek and Woessmann (2006)

9 Appendix

Table A1: Difference-in-Differences Reading (PISA 2003 and PIRLS 2001)
Without School Quality

Dependent Variable: Reading Test Score

	(1)	(2)	(3)	(4)
Highest Parental	6.27			4.66
Education (years)	(0.69)***			(0.52)***
HPE*PISA	0.01			-0.50
	(1.08)			(0.80)
HPE*Earlytrack	4.99			3.67
	(1.31)***			(1.04)***
HPE*Earlytrack*PISA	-1.25			-1.27
	(0.67)*			(1.08)
Number of Books		0.14		0.10
		(0.02)***		(0.01)***
NoB*PISA		0.06		0.06
		(0.02)***		(0.02)***
NoB*Earlytrack		0.05		-0.01
		(0.02)**		(0.02)
NoB*Earlytrack*PISA		-0.03		0.04
		(0.03)		(0.03)
Language			35.58	31.07
			(3.83)***	(3.26)***
Language*PISA			-10.98	-14.58
			(7.49)	(7.89)*
Language*Earlytrack			20.35	9.59
			(8.49)**	(3.85)**
Language*Earlytrack*PISA			-12.06	-7.86
			(8.25)	(13.54)
Secondary School Test Dummy	✓	✓	✓	✓
Student Characteristics	✓	✓	✓	✓
School Quality				
Country FE	✓	✓	✓	✓
Number of Countries	15	15	15	15
Observations	146631	150348	158682	143957
R-squared	0.18	0.20	0.14	0.19

Table A2: Difference-in-Differences Mathematics (No School Quality)
TIMSS 1995 (Grades 3/4) and TIMSS 1995 (Grades 7/8)
Without School Quality

Dependent Variable: Mathematics Test Score

	(1)	(2)	(3)
Number of Books	0.15		0.13
	(0.01)***		(0.01)***
NoB*Secondary	0.05		0.06
	(0.01)***		(0.01)***
NoB*Earlytrack	0.05		0.06
	(0.02)**		(0.02)***
NoB*Earlytrack*Secondary	0.00		-0.02
	(0.03)		(0.03)
Language		35.74	31.96
		(2.69)***	(2.54)***
Language*Secondary		-7.75	-11.71
		(5.61)	(5.09)**
Language*Earlytrack		17.26	11.74
		(5.60)***	(4.58)**
Language*Earlytrack*Secondary		11.38	9.97
		(6.51)	(4.76)*
Student Characteristics	✓	✓	✓
School Quality			
Country FE	✓	✓	✓
Number of Countries	15	15	15
Observations	210224	189915	185947
R-squared	0.18	0.15	0.19

Robust standard errors in parenthesis (clustered at the country level)
* significant at 10% ** significant at 5% *** significant at 1%

Table A3: Difference-in-Differences Mathematics
TIMSS 1995 (Grades 3/4) and TIMSS 1995 (Grades 7/8)
Omitting the Dummies for Family Structure

Dependent Variable: Mathematics Test Score

	(1)	(2)	(3)
Number of Books	0.15		0.14
	(0.02)***		(0.02)***
NoB*Secondary	0.05		0.05
	(0.01)***		(0.01)***
NoB*Earlytracking	0.07		0.05
	(0.02)**		(0.02)**
NoB*Earlytracking*Secondary	-0.04		-0.02
	(0.02)*		(0.04)
Language		34.60	31.17
		(4.89)***	(4.33)***
Language*Secondary		-8.10	-12.30
		(7.61)	(6.90)
Language*Earlytrack		7.03	3.94
		(4.71)	(4.90)
Language*Earlytrack*Secondary		0.34	-1.55
		(8.54)	(5.06)
Student Characteristics	✓	✓	✓
School Quality	✓	✓	✓
Country FE	✓	✓	✓
Number of Countries	11	11	11
Observations	102522	97310	95864
R-squared	0.20	0.16	0.20

Robust standard errors in parenthesis (clustered at the country level)
* significant at 10% ** significant at 5% *** significant at 1%