

Ruhose, Jens; Schwerdt, Guido

Working Paper

Does Early Educational Tracking Increase Migrant-Native Achievement Gaps? Differences-In-Differences Evidence Across Countries

CESifo Working Paper, No. 5248

Provided in Cooperation with:

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Ruhose, Jens; Schwerdt, Guido (2015) : Does Early Educational Tracking Increase Migrant-Native Achievement Gaps? Differences-In-Differences Evidence Across Countries, CESifo Working Paper, No. 5248, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:

<https://hdl.handle.net/10419/108788>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



Does Early Educational Tracking Increase Migrant-Native Achievement Gaps? Differences-In-Differences Evidence Across Countries

Jens Ruhose
Guido Schwerdt

CESIFO WORKING PAPER NO. 5248

CATEGORY 5: ECONOMICS OF EDUCATION

MARCH 2015

Presented at CESifo Area Conference on the Economics of Education, September 2014

An electronic version of the paper may be downloaded

- from the SSRN website: www.SSRN.com
- from the RePEc website: www.RePEc.org
- from the CESifo website: www.CESifo-group.org/wp

ISSN 2364-1428

Does Early Educational Tracking Increase Migrant-Native Achievement Gaps? Differences-In-Differences Evidence Across Countries

Abstract

We study whether early tracking of students based on ability increases migrant-native achievement gaps. To eliminate confounding impacts of unobserved country traits, we employ a differences-in-differences strategy that exploits international variation in the age of tracking as well as student achievement before and after potential tracking. Based on pooled data from 12 large-scale international student assessments, we show that cross-sectional estimates are likely to be downward-biased. Our differences-in-differences estimates suggest that early tracking does not significantly affect overall migrant-native achievement gaps, but we find evidence for a detrimental impact for less integrated migrants.

JEL-Code: I210, J150, I280.

Keywords: immigration, educational inequalities, educational tracking, differences-in-differences.

Jens Ruhose
Ifo Institute – Leibniz Institute for
Economic Research
at the University of Munich
Poschingerstrasse 5
Germany – 81679 Munich
ruhose@ifo.de

Guido Schwerdt
Department of Economics
University of Konstanz
Box 133
Germany - 78457 Konstanz
guido.schwerdt@uni-konstanz.de

March 2015

We are grateful to Eric Hanushek, Lant Pritchett, Ludger Woessmann as well as participants at the CESifo Economics of Education area conference in Munich, the European Society for Population Economics (ESPE) meeting in Aarhus, the European Economic Association (EEA) meeting in Gothenburg, the German Economic Association (VfS) meeting in Düsseldorf, the Second Lisbon Research Workshop on Economics, Statistics and Econometrics of Education and seminar participants in Munich and Konstanz.

1 Introduction

Migrants in major European countries are failing to assimilate economically (Algan et al., 2010). Thus, their better integration has become a priority for policymakers in the European Union. However, it is less clear what public policy can do to effectively address this situation. As equal opportunity in education is key to the successful long-term integration of immigrants and the educational achievement of migrants lags behind that of native students in almost all European countries (e.g., Ammermueller, 2007; Nielsen et al., 2003; Van Ours and Veenman, 2003; Bauer and Riphahn, 2007; Schnepf, 2007; Schneeweis, 2011), education policies are being looked at with particular interest.

One institutional feature of several school systems in Europe that has been shown to generally increase educational inequality is early tracking of students into different types of secondary school based on their ability (e.g., Hanushek and Woessmann, 2006). In the context of migrant inclusion, early tracking may have detrimental effects if migrants are more likely to be inadequately matched to educational pathways.¹ For example, this may be the case when migrants face difficulties to signal their educational potential at an early stage, either because of a lack of proficiency in the language of instruction (Akresh and Akresh, 2010) or a systematically different parental background.²

These general insights lead many to conclude that early tracking systems might be specifically detrimental to students with a migration background (e.g., Van de Werfhorst and Mijs, 2010) and it is often explicitly recommended to policymakers that educational systems be made less selective to improve opportunities for migrant students (e.g., NESSE, 2008). Direct evidence based on cross-sectional data on the relationship between complete forms of educational tracking and migrants' relative achievement indeed seems to support these conclusions (e.g., Cobb-Clark et al., 2012). However, cross-sectional estimates are plagued with endogeneity concerns arising, for example, due to potentially selective migration into countries with early tracking systems.

This paper studies the impact of ability-based early tracking of students into different types of secondary school on migrant-native test score gaps in a differences-in-differences framework, which implicitly controls for unobserved differences in relevant characteristics of the migrant

¹For example, Lüdemann and Schwerdt (2013) and Kiss (2013) show that second-generation immigrants in Germany receive worse grades and worse teacher recommendations for secondary school tracks conditional on student achievement.

²Several studies document that early educational tracking between school types increases the effects of parental background on educational outcomes (e.g., Bauer and Riphahn, 2006; Meghir and Palme, 2005; Pekkari-nen et al., 2009; Kerr et al., 2013).

and native student populations that remain constant over educational stages. In particular, our identification strategy makes use of the fact that no country tracks students in primary school. This allows us to exploit variation in migrant-native test score gaps between primary and secondary school as well as variation in the age of tracking between countries to identify the effect of early tracking. To benchmark our differences-in-differences estimates, we additionally present results based on commonly estimated cross-sectional models that rely entirely on selection-on-observable assumptions.

Our empirical analysis is based on a comprehensive analytical sample that we obtained by pooling data from all existing waves of the three largest international assessments of student achievement during primary and secondary school – PISA, TIMSS, and PIRLS – which test students in reading, math, and science. As a result, we have internationally comparable information on student achievement and background characteristics for more than one million students from 45 countries.

We find that early tracking does not significantly affect the evolution of overall migrant-native test score gaps from primary to secondary school in math and science. Nor do we find significant effects for reading, but results are somewhat less clear-cut. For all subjects, however, there are no substantial negative impacts of more than 10% of a standard deviation. The small and insignificant estimates of the overall effect conceal a detrimental effect of early tracking on the relative achievement of second-generation immigrants who do not speak the language of the testing country at home. For this subgroup of migrant students, we find a significant detrimental effect of early tracking on relative achievement in reading of about 11% of a standard deviation and similar, but insignificant, effect sizes in terms of math and science achievement. Thus, our findings show that tracking students early into different types of schools by ability does not substantially reduce relative achievement growth of all migrant children, but it does so for those who are presumably less integrated into the country's society.

The key empirical challenge for such an investigation is to distinguish accidental correlation from causation. When investigating effects of a system-level variable, such as early educational tracking, empirical research is basically forced to exploit cross-country variation in educational policies for identification.³ However, simple cross-country identification strategies based on cross-sectional data rely on strong conditional independence assumptions, basically assuming away any differences in non-ignorable unobservable country traits. It is questionable whether

³Questions regarding the school system could also be addressed by exploiting school reforms within countries. However, especially in the case of migration economics, results from a particular country with a particular migrant population are hard to generalize to other countries with other school systems and other migrant populations.

any policy conclusions can be drawn from such evidence.

As first noted by Hanushek and Woessmann (2006), the availability of test score data by country before and after tracking allows estimating effects of early educational tracking in a differences-in-differences framework. Employing this framework and identifying the effect of early tracking based on a comparison of the evolution of migrant-native achievement gaps from primary to secondary school reveals no significant effects of early tracking. The results from commonly estimated cross-sectional models would, however, imply a substantial positive association between early educational tracking and the size of the migrant-native test score gap. In all three domains – math, science, and reading – migrant-native achievement gaps in secondary school are found to be between 20 - 30% of a standard deviation larger in countries that track students before the age of 15.

However, we show that a similar relationship already exists in primary school, a period during which no country has yet tracked students according to ability. This indicates more that the association between early tracking and the migrant-native achievement gap is driven by selective migration toward early tracking countries. Among the late tracking countries, there are the anglophone countries – United States, Australia, Canada, New Zealand, and United Kingdom – that have highly selective migration policies in place, whereas early tracking countries are mainly located in Europe – Germany, Austria, the Netherlands, Belgium, and Italy – and do not select migrants as strictly as the other countries do. Stricter immigration policies create a more selected migrant population (Grogger and Hanson, 2011), which could partly explain the large cross-sectional differences in the migrant-native achievement gap. Furthermore, there could be bilateral country factors between sending and receiving countries, such as language differences, cultural differences, networks and diasporas, and religious differences, that can shape the skill structure of the migrant population. For our study, this insight is crucial because it implies that results from cross-sectional models are likely to be confounded by country-specific migrant populations and how they interact with the country-specific school system.

The remainder of the paper is structured as follows. Section 2 describes the data. In Section 3, we present our empirical strategy. Section 4 presents our main results and Section 5 contains a discussion as to their robustness. Section 6 investigates whether our main results conceal more pronounced effects for specific groups of students. Section 7 concludes.

2 Data

2.1 International Large-Scale Student Assessments

To address our research question as comprehensively as possible, we use data from all existing waves of the three largest international large-scale student assessments: the Programme for International Student Assessment (PISA), the Trends in International Mathematics and Science Study (TIMSS), and the Progress in International Reading Literacy Study (PIRLS).⁴

Each of these large-scale student assessments has a different focus with respect to the tested population and tested domains. PISA tests students at age 15 in reading, science, and math. TIMSS evaluates performance in math and science in grades 4 and 8. PIRLS tests fourth graders in reading only. Thus, for primary school students, we have test scores for reading from PIRLS and for math and science from TIMSS; for secondary school students, we have math and science test scores from both TIMSS and PISA and reading test scores from PISA.

We pool data from all cycles of these large-scale student assessments conducted between 1995 and 2012. Specifically, we use information on reading achievement in primary school from the PIRLS waves 2001 and 2006, and on math and science achievement in primary school from the TIMSS waves 1995, 2003, and 2007.⁵ Our data on achievement in secondary school come from the PISA waves 2000, 2003, 2006, 2009, and 2012 in the domains of reading, math, and science as well as from the TIMSS waves 1995, 1999, 2003, 2007, 2011 for math and science. In sum, our sample contains information on more than one million tested students in 45 countries.

2.2 Defining Early Educational Tracking

The 45 countries have differently structured school systems. One difference of particular importance to our work is that countries track students based on ability into different schools at different ages. Information on school systems for a large number of countries, including information on the age of first tracking, is provided by the UNESCO International Bureau of Education (2013). We checked this information against OECD data (OECD (2007, Table 5.2, p. 162), OECD (2010, Table A5.2, p. 147)) and information provided in other studies on educational tracking (Brunello and Checchi, 2007; Waldinger, 2007; Hanushek and Woessmann, 2006) to ensure that the age of first tracking information is accurate and consistent with what

⁴The data from all TIMSS and PIRLS surveys, along with detailed documentation of the data and the survey design, are available at <http://timss.bc.edu>. The PISA data and documentation are available at <http://www.oecd.org/pisa>.

⁵PIRLS 2011 and TIMSS 2011 in grade 4 do not provide information on whether the child and/or the parents are born in the testing country. We therefore cannot use these surveys in our analysis.

is commonly used in the literature.

Table 1 reports age and grade level of first selection in the school system by country as obtained from the UNESCO International Bureau of Education (2013). The table reveals considerable variation in the age of tracking. For example, Austria and Germany track students according to ability into different tracks of secondary school at age 10, whereas in the majority of countries students stay in comprehensive schools until age 15 or 16. We follow Hanushek and Woessmann (2006) and define early tracking as the process of being tracked for the first time into different school tracks before age 15. Table 1 reports which countries are early tracking countries based on this definition. There are 15 early tracking countries and 30 late tracking countries. Although this way of defining early educational tracking is common in the literature, it is still somewhat arbitrary. Therefore, we test the robustness of our main results using alternative definitions in Section 5.

2.3 Analytical Sample

We restrict our focus to second-generation immigrants to ensure that both native and migrant students have spent their entire school career in the testing country. We define as second-generation immigrants all students born in the testing country who have at least one parent who was born abroad. A native student is born in the testing country and has parents who are also born there. We test the robustness of our main results with respect to alternative definitions of migrant status in Section 5.

All students with missing information on their parents' country of birth or students born abroad are removed from the sample. Students with missing information on test scores, age, gender, and books at home are also excluded from our analysis. To avoid having our results largely driven by countries with very few migrant students, we follow common practice and drop all countries with a migrant ratio below 3%.⁶ We check the robustness of our main results with respect to this sample restriction in Section 5.

For our main analysis, we include only those countries for which we have information on all three domains (reading, math, and science) in at least one assessment during primary school and one assessment during secondary school. This restriction should ensure that comparisons of results across domains are not confounded by differences in the country composition of estimation samples. Table 1 lists the countries included in this baseline sample. We explore the robustness of our main results with respect to this sample restriction in Section 5.

⁶All countries listed in Table 1 meet this requirement.

To ease interpretation of regression coefficients, we standardize test scores in each domain within each survey cycle before we restrict the sample to the countries in our analysis to have an international mean of 0 and a standard deviation of 1. Thus, when using standardized test scores as regression outcomes, coefficients represent changes in standard deviations of test scores.⁷

Table 2 provides summary statistics separately for students tested in primary and secondary school on test scores by tracking and migrant status. Panels A and B of Table 2 report mean test scores by migrant status separately for late and early tracking countries. Test score averages in Panel A refer to secondary school; those reported in Panel B refer to primary school. In secondary school, natives generally outperform migrants independently of the tracking system with the exception of reading achievement in late tracking systems. However, migrant-native test score gaps are substantially larger in early tracking countries. For example, the unconditional test score gap in math is more than ten times larger in early tracking (-0.206) than in late tracking systems (-0.016).

However, this overall pattern is not very different from the general pattern in primary school, as can be seen in Panel B of Table 2. As no country has yet tracked students in primary school, this suggests that the observed correlation between early educational tracking and migrant-native test score gaps in secondary school may at least partly arise due to factors other than educational tracking. Nevertheless, early educational tracking may aggravate educational disadvantages that already exist during primary school.

This possibility is illustrated in Figure 1, which plots the average migrant-native achievement gaps in primary and secondary school reported in Table 2. The figure reveals noticeable differences between tracking systems in the evolution of migrant-native achievement gaps over educational stages. While differences in test scores between natives and migrants become smaller in all subjects between primary and secondary school in late tracking countries, unconditional test score gaps widen in early tracking countries. Whether these differences are in fact statistically significant and whether they can be explained by compositional differences in the student population over time and across countries is the focus of our empirical analysis in the following sections.

To investigate to what extent compositional differences in the student population matter for our analysis, we collect information on the demographic background and on the socioeconomic

⁷Alternative, we have standardized test scores only over the sample of countries that are included in our baseline regression. However, whether we standardize within educational stage or within each survey, the results do not change qualitatively. Results with different standardizations are not presented, but are available upon request.

background, which we proxy with the commonly used books at home variable, from the three large-scale student assessments. These background variables include age, gender, books at home (five categories), a dummy indicating whether students are tested in an OECD country, and GDP per capita of the test country. Panels A and B of Table 3 report averages of the demographic and socioeconomic background variables in primary and secondary school by tracking and migrant status and reveal some noteworthy differences. Migrants are overrepresented in OECD countries and countries with a higher GDP per capita. Students (both natives and migrants) in late tracking countries are more likely to be living in an OECD and high-income country. Finally, as expected, the information on books at home shows that migrants are on average underrepresented in the category of having more than 200 books and somewhat overrepresented in the lower categories. This relationship is stronger in the early tracking countries than in the late tracking countries.

3 Empirical Strategy

All else equal, does early educational tracking increase migrant-native achievement gaps? We begin answering this question by estimating conventional cross-sectional models using micro data and control for country fixed effects. We then investigate how results change in a differences-in-differences framework.

3.1 Cross-Sectional Model

To benchmark our differences-in-differences estimates, we first estimate commonly estimated cross-sectional regression models (e.g., Cobb-Clark et al., 2012). We estimate these models based on pooled data for the three domains of math, science, and reading using the following equation:

$$Y_{ic} = \alpha_0 + \alpha_1 ET_c \times MIG_i + \alpha_2 MIG_i + \mathbf{X}'\gamma + \mu_c + \epsilon_{ic}, \quad (1)$$

where Y_{ic} is the standardized tests score of individual i in country c , ET_c is a binary variable taking the value 1 if country c is an early tracking country and 0 otherwise, MIG_i is a binary variable taking the value 1 if student i is a second-generation immigrant and 0 otherwise, $\mathbf{X}$ is a vector of student and country background variables, μ_c is a survey times country fixed effect, and ϵ_c is an idiosyncratic error.⁸

⁸Note that due to the inclusion of country fixed effects, the parameter on an indicator variable for early tracking countries is not identified, which is why ET_c only enters Equation (1) interacted with MIG_i . Estimating versions of Equation (1) that include the single ET_c dummy, but exclude country fixed effects, leads to similar results.

The key parameter of interest is α_1 . This parameter captures any additional effect of being a migrant in an early tracking country over and above the isolated effect of migration background on test scores in a late tracking country, which is captured by α_2 . Estimating Equation (1) by OLS produces unbiased estimates for α_1 only if the observable covariates included in the model are the sole reason $ET_c \times MIG_i$ and ϵ_{ic} might be correlated.

However, it seems likely that this selection-on-observables assumption is violated in this case. For example, migrant students in early as opposed to late tracking countries might be very different with respect to unobserved characteristics that affect the level of achievement at any educational stage. One example could be the Guestworker program in Germany that has attracted a lot of Turkish workers during the 1960s and 1970s (Danzer and Yaman, 2013). Later, the workers' families were allowed to move to Germany as well. As a result, Germany is now one of the main destinations for Turkish migrants (Grogger and Hanson, 2011). Cross-sectional regressions cannot account for this historical event and would show an accidental correlation between early tracking and low educational migrant performance. Another concern is that other educational features affecting migrant-native achievement differences might systematically differ between early and late tracking countries. For example, the quality of the teaching force differs widely between countries (Hanushek et al., 2014), which is not captured in a cross-sectional design.

3.2 Differences-in-Differences Model

Longitudinal data on the evolution of test scores after tracking for the same cohort of students in a large set of countries would allow us to eliminate any confounding impact of unobserved time-invariant characteristics on the size of migrant-native test score gaps in a straightforward differences-in-differences framework. Unfortunately, such data are not available. However, as first noted by Hanushek and Woessmann (2006), information on student achievement and background from large-scale international student assessments, such as PISA, PIRLS, and TIMSS, that test and survey cross-sections of students in different countries at different educational stages permits a differences-in-differences estimation strategy.

Our identification strategy makes use of the fact that no country tracks students in primary school. Thus, we can obtain information on migrant-native achievement differences prior to tracking from the student assessments in grade 4 provided by PIRLS (reading) and TIMSS (math/science). Information on migrant-native achievement differences in secondary school, after tracking, is provided by PISA (reading/math/science), which tests students at age 15, and

the TIMSS (math/science) assessments in grade 8. The corresponding differences-in-differences model based on pooled data on students observed at two educational stages (e), where e might be primary or secondary school, is given by:

$$\begin{aligned}
Y_{ice} = & \beta_0 + \beta_1 ET_c \times SEC_e \times MIG_i + \\
& \beta_2 ET_c \times MIG_i + \beta_3 ET_c \times SEC_e + \beta_4 MIG_i \times SEC_e + \\
& \beta_5 MIG_i + \beta_6 SEC_e + \mathbf{X}'\delta + \nu_c + \eta_{ice},
\end{aligned} \tag{2}$$

which, in addition to the variables included in Equation (1), includes the dummy variable SEC_e that identifies observations in secondary school together with its interactions with $ET_c \times MIG_i$, ET_c , and MIG_i .

The key parameter of interest in Equation (2) is the parameter on the triple interaction term β_1 . This parameter measures the effect of early educational tracking on the migrant-native achievement gap. The advantage of estimating Equation (2) is that the inclusion of the interaction $ET_c \times MIG_i$ captures any confounding impact of unobserved educational-stage-invariant characteristics on the level of the migrant-native test score gap, which is represented by the parameter β_2 .

The effect β_1 is identified under the assumption that any unobserved differences between students in early and late tracking countries does not differentially affect the evolution of the migrant-native test score gap between primary and secondary school. This is the well-known "parallel trends" assumption of differences-in-differences models. We argue that this identifying assumption is more credible than the selection-on-observables assumption required to causally interpret estimates based on Equation (1). The parallel trends assumption would be violated if the migrant-native achievement gap in primary school is affected by early tracking. One example could be that native parents in early tracking countries anticipate the tracking event and therefore push their children at the end of primary school to perform well. We discuss this issue more in Section 5.4.

To estimate Equation (2), we have to match to each assessment in primary school (PIRLS for reading, and TIMSS for math and science) one assessment in secondary school (PISA for reading, math, and science and TIMSS for math and science). Because there are multiple waves of PISA, TIMSS, and PIRLS, we can construct multiple matches of survey waves (Hanushek and Woessmann, 2006). Figure 2 visualizes how we match the different survey waves. Altogether, we construct 18 different matches, containing exactly one cohort of students tested in primary and one tested in secondary school.⁹ For the final dataset, we append the student-level data for

⁹Appendix Table A-1 provides country information on the inclusion of the country in a specific survey match.

each survey match. Appending all survey matches lead to the problem that some countries are overrepresented. Therefore, the original sampling weights are adjusted such that each country carries a weight of 1. The survey matches can be grouped into 3 types: PISA-PIRLS, PISA-TIMSS, and TIMSS-TIMSS comparisons. In the baseline analysis, we merge PISA-TIMSS and TIMSS-TIMSS comparisons for math and science.¹⁰

4 Results

4.1 Results from the Cross-Sectional Model

We begin the presentation of our results by showing OLS estimates of Equation (1). Table 4 sets out regression results for math, science, and reading achievement in secondary school when educational tracking has already occurred in some countries, but not in others. Our baseline sample includes 25 countries, nine of which track students before age 15. Results of estimating Equation (1) are presented both with and without other control variables. However, all regression models include country-by-survey-cycle fixed effects. To allow for within-country correlation of error terms, we cluster standard errors at the country level. We estimate weighted regressions using original sampling weights, which we adjusted such that each country carries a weight of 1.

Estimations without further controls show that in all three domains, early educational tracking is related to lower achievement by migrant relative to native students. In late tracking countries, average achievement of migrant and native students is about the same. The regression coefficients on the migrant dummy, the estimates of α_2 in Equation (1), are small and statistically insignificant. In early tracking countries, however, migrants score on average about 20% of a standard deviation lower than native students. Estimates of α_1 in Equation (1) are statistically significant at the 10% level in reading and math and insignificant in science. Overall, these findings are in line with the descriptive evidence presented above.

To investigate to what extent these differences can be explained by compositional differences in student populations, we control for demographic and socioeconomic background variables. Results are reported in Columns 2, 4, and 6 of Table 4. The coefficients on the additional covariates have the expected signs and are comparable with other evidence on the determinants of student achievement. Number of books at home is the strongest predictor of student

As the table shows, the composition of countries changes from survey match to survey match because not every survey is carried out in the same countries. Section 5.3 provides robustness checks by changing the composition of the survey matches.

¹⁰Appendix Table A-2 provide differences-in-differences results for each comparison separately.

achievement. For example, the difference in test scores of students with more than 200 books at home compared to the reference group of students amount to almost an entire standard deviation. Boys perform significantly worse than girls in reading, but perform slightly better in math and science. Age is positively related with achievement in reading.

To account for selective migration into more economically developed destination countries, we also include two country characteristics—GDP per capita and OECD membership—interacted with migrant status as covariates.¹¹ The interaction with OECD membership is highly significant and negative, indicating that migrants in non-OECD countries—relative to their native peers—score substantially better than migrants in OECD countries.

Turning to our main parameters of interest, we observe in all three domains that estimates of α_1 in Equation (1), the coefficients on the interaction between early tracking and migrant status, become smaller in absolute terms, suggesting that migrant-native achievement gaps are in fact partly attributable to demographic and socioeconomic background variables. Compared to the unconditional regression results, estimated effects of early educational tracking shrink in absolute terms to 13% of a standard deviation in reading, 11% of a standard deviation in math, and 10% of a standard deviation (insignificant) in science.¹² Interestingly, unconditional results for the estimates of α_2 in Equation (1) indicate that the average migrant-native achievement gap is mainly driven by early tracking countries. In late tracking countries, migrants even slightly outperform natives (on average) in reading (but not significantly so). Conditional on all covariates, migrants in non-OECD countries with mean GDP per capita and late tracking school systems significantly outperform their native peers in all subjects (see estimates of α_2 in Equation (1)). The coefficients on the OECD $\times$ migrant interactions reveal that the migrant-native achievement gap is mainly an issue in OECD countries. There, we observe that migrants perform significantly worse than natives.

As argued above, the estimates of α_1 and α_2 presented in Table 4 may not have a causal interpretation if unobserved differences between early and late tracking countries systematically affect the relative performance of migrants with respect to their native peers. In this case, the differences in secondary school performance between the two groups of countries may arise even if the age of educational tracking by ability were the same in all countries. While we cannot observe this counterfactual scenario, the availability of test score data on primary

¹¹Note that GDP per capita is de-measured to provide the interpretation in terms of deviations from a country with mean international income.

¹²Appendix Table A-2 reproduces the estimates found in Table 4, but estimates specifications for math and science separately for PISA and TIMSS data. Results are qualitatively comparable to the results presented in Table 4.

school performance allows a reasonable robustness analysis. Whether a country tracks students early should have no effect on migrant-native achievement gaps before tracking has actually occurred.¹³ Thus, we expect to find no significant estimate of α_1 when estimating Equation (1) based on primary school data.

Table 5, however, confirms that there is already a migrant-native achievement gap in early tracking countries in primary school. Coefficients on the interaction terms between early educational tracking and migrant status are negative and statistically significantly different from 0 in all three domains. This suggests that differences between early and late tracking countries in the performance of migrants relative to their native peers already exist in primary school. Unconditionally, estimates of α_1 in primary school are around 14% to 15% of a standard deviation in all three domains—about 5 percentage points lower than in secondary school. With further controls, estimates are 8% of a standard deviation in reading and 9% of a standard deviation in math and science. Overall, these estimates are just slightly lower in absolute terms than the estimates in secondary school, with a somewhat larger difference in reading. This clearly suggests that the estimates of the impact of early educational tracking on secondary school performance of migrants relative to their native peers presented in Table 4 are biased upward in absolute terms. However, there may be small effects of early educational tracking on migrant-native achievement gaps that result in an increase in achievement gaps over educational stages. Comparing coefficient estimates in Tables 4 and 5 suggest that this may particularly be the case for reading achievement.

4.2 Results from the Differences-in-Differences Model

The previous subsection showed that migrant-native achievement gaps are present in primary and secondary school. In this section, we combine both results to investigate whether early tracking has an effect on the change in migrant-native achievement gaps over educational stages by estimating the differences-in-differences model in Equation (2). Table 6 presents the results from estimating the differences-in-differences model once with and without further controls. All regressions include country-by-survey-cycle fixed effects. As before, we estimate regression models separately for each domain.

Estimates of the impact of educational-stage-invariant factors on migrant-native test score gaps (the parameter β_2 in Equation (2)) are set out in the second row of Table 6. The coefficient estimates on the early tracking times migrant interaction basically capture any differences in

¹³We discuss this hypothesis, which is related to the parallel trends assumption and important for the causal interpretation of our results, further in Section 5.4.

migrant-native test score gaps between early and late tracking countries that already exist in primary school. Consequently, the estimated coefficients in the second row are very similar to the estimates of α_1 in Table 5.

Our key parameter of interest, β_1 , in the differences-in-differences model is given by the estimates in the first row of Table 6. Only Column 1 shows a significant coefficient, suggesting that early educational tracking increases migrant-native test score gaps in reading by 7% of a standard deviation. However, when we additionally control for observable demographic and socioeconomic background characteristics, the coefficient estimate is one-third lower in absolute terms and statistically insignificant. For math and science, none of the coefficients on the triple interaction is significant. Moreover, point estimates in the specifications with further controls are small at -0.04 in math and -0.02 in science. Due to the clustering of standard errors at the country level, the precision of our estimates is somewhat low. Standard errors on the triple interaction coefficients in the estimations with further controls range from 0.03 to 0.06. We therefore cannot rule out the possibility that early educational tracking might have some small effects.

Overall, based on the findings, not much of a case can be made for a substantial negative impact of early educational tracking on overall migrant-native test score gaps. At most, our findings indicate that early educational tracking may have some detrimental impact on immigrants' relative achievement in reading. Although effect sizes below 0.1 cannot be rejected, our differences-in-differences estimates give rise to concern that estimates based on cross-sectional models, such as the one presented in Table 4, should not be interpreted causally.

However, the relatively low statistical precision of these estimates demands a closer inspection of the data before drawing firm policy conclusions. In particular, the small negative point estimates may be a result of true zero effects for some groups of migrant students and substantial negative effects for others. We explore the possibility of substantial effect heterogeneities in Section 6, after conducting several robustness checks with respect to the choice of our preferred empirical specification in the next section.

5 Robustness Checks

In this paper we argue that, at a minimum, our differences-in-differences approach identifies the effect of early educational tracking under more credible assumptions than those used in making cross-sectional estimates. However, unobserved factors affecting the evolution over educational stages of migrant-native test score gaps that are correlated with age at tracking

may bias our main results. In this section, we conduct a wide range of sensitivity checks to assess the robustness of our main results.

5.1 Different Country Samples

Table 7 presents a set of robustness checks and is organized similarly to Table 6. Each cell shows estimates of our differences-in-differences model in Equation (2). Outcomes are reading scores in Columns 1 and 2, math scores in Columns 3 and 4, and science scores in Columns 5 and 6. For each outcome, we estimate Equation (2) once with and without other controls. However, Table 7 only reports estimates of our key parameter of interest, β_1 . Each row represents a different robustness check, the type of which is specified in the first column.

We begin by estimating Equation (2) based on an extended set of countries. In this analysis, we include all 45 countries listed in Table 1. Results are reported in the first row of Table 7. The overall pattern of results is virtually the same. We obtain similar results again when restricting our sample to a set of countries for which we have data for all potential survey matches, that is, matches for reading in PIRLS–PISA, math in PISA–TIMSS and TIMSS–TIMSS, and science in PISA–TIMSS and TIMSS–TIMSS (see also Table 1). This restriction reduces the sample to 20 countries, six of which are early tracking countries. The results from this variation can be found in the second row of Table 7.

A country’s level of economic development may matter for our analysis because of selective migration into more or less developed countries. Following Cobb-Clark et al. (2012), we limit our sample to OECD countries (see Row 3 of Table 7). The estimates are slightly larger in absolute terms in all domains, but only the estimates for reading remain significant once further controls are included. In Row 4, we restrict the sample to include only main destination countries as defined in Grogger and Hanson (2011).¹⁴ This restriction results in even larger effect sizes in absolute terms, which are now significant in reading and math, and almost significant in science. The results indicate that especially the three countries Austria, Germany, and the Netherlands drive the small effects in the baseline regressions, including all countries.

The estimates change in a similar but more pronounced way when we exclude countries with a migrant ratio below 10% (instead of 3% as in the baseline sample). However, as shown in Row 5 of Table 7, estimates in math and science remain insignificant once further controls

¹⁴These countries include the early tracking countries Austria, Germany, and the Netherlands, and the late tracking countries Canada, Denmark, New Zealand, Norway, Sweden, the UK, and the United States, which accounted – together with Australia, Finland, France, Ireland, Spain, which are excluded due to data availability – for 87.6% of all immigrants hosted by OECD countries in 2000 (Grogger and Hanson, 2011).

are included.¹⁵

Restricting our sample to subsets of countries (Rows 2 to 5 of Table 7) revealed that the selection of countries has some impact on our estimates. While some degree of country heterogeneity can be expected, a key concern is that our overall finding is entirely driven by the evolution of the relative achievement of migrants in a specific country. To investigate this possibility, we re-estimate our preferred specification several times and exclude each time a different country. Table 8 presents the results from this piecewise deletion exercise. The results are very similar and show that our main result is not driven by a specific country.

5.2 Different Tracking Definitions

As discussed in Section 2, we followed Hanushek and Woessmann (2006) in defining early tracking systems as education systems that track students for the first time into different school tracks before age 15. Rows 6 to 10 of Table 7 show the robustness of our main results to alternative definitions of early educational tracking.

We first employ a more restrictive definition of early tracking (see Row 6 of Table 7) and include as early tracking countries only those that track their students before age 12 and before grade 6 (see also Table 1). In Row 7, we check whether results are driven by the very early tracking countries and exclude countries that track students before grade 6. Using a dummy for categorizing countries according to their tracking status could lead to false classifications. In Row 8, we avoid this issue and use the actual tracking age (as reported in Table 1) instead of the early tracking dummy variable. The findings of Waldinger (2007) indicate that results may partly depend on whether age or grade of first selection is used to define early tracking systems. Thus, in Row 9, we define as early tracking countries those that track students before grade 8. Finally, in Row 10, we use the actual tracking grade (as reported in Table 1) instead of the early tracking dummy variable.

The results of these robustness checks indicate that our finding of no substantial impact of early educational tracking on migrant-native test score gaps is in no way driven by how "early tracking country" is defined.

¹⁵This restriction excludes the early tracking countries Czech Republic, Hungary, Italy, and Slovak Republic, and the late tracking countries Greece, Iceland, Lithuania, and Taipei. We also exclude Norway, Russia, and the United Kingdom in some survey matches. Thus, we have almost the same set of early tracking countries as in the main destination countries sample before, that is, Austria, Germany, the Netherlands, plus Singapore and Belgium.

5.3 Different Survey Matches Definitions

Our identification strategy may be influenced by potentially confounding cohort and calendar time effects as we basically compare different cohorts tested in different years. The existence of general cohort or calendar time effects is of no concern, but cohort or calendar time effects that differentially affect the relative performance of migrant students in early and late tracking countries might bias our main estimates.

To address this concern, we limit our attention to combinations of surveys (or test pairs) that either keep the cohort or the calendar year constant (see Figure 2). For example, one of these test pairs pools information on all fourth and eighth graders tested in TIMSS in 2007. In this specific test pair, calendar time effects should play no role as all students are tested at the same time, whereas cohort effects remain a worry. In another test pair, we pool information on all fourth graders tested in TIMSS 2003 and all eighth graders in 2007. In this case, cohort effects might be less of a concern as both samples should be drawn from the same cohort of students. In Row 11 of Table 7, we use only data from combinations of surveys that sample from the same cohort; in Row 12 we use only data from combinations of surveys that are conducted in the same calendar year. Once again the results are very similar to our baseline results.

In sum, the robustness checks generally support the validity of our main findings. In particular, none of the robustness checks that alter the definition of early tracking or that try to control for cohort or calendar time effects show significant negative effects. Only the restriction to main destination countries reveals significant effects in reading and math. However, this is more likely to reflect some true effect heterogeneity rather than being an indication for a bias in our results.

5.4 Anticipation Effects

Finally, a potential threat to our identification strategy would be if the migrant-native test score gap in primary school is affected by educational tracking itself. This could be the case, for example, when native parents put more effort than migrant parents in supporting their children in anticipation of the upcoming educational tracking. The possibility of such anticipation effects is a common and known threat to all studies with this identification strategy (e.g., Hanushek and Woessmann, 2006; Waldinger, 2007).

While we have no direct way of testing for anticipation effects, a widening of the migrant-native test score gap during primary school as students get closer to the tracking decision may be some indication for the relevance of anticipation effects. The fact that TIMSS has

administered the test in primary school in 1995 within the same country to children in two different grades in the early tracking countries Singapore, Austria, Czech Republic, and the Netherlands allows us to document the evolution of migrant-native achievement gaps in these countries primary school grades. We do so by estimating Equation (1) separately for different grades in primary school.

We find that migrant-native achievement gaps increase only very little between the earlier and the later grade in the four countries (Appendix Table A-3).¹⁶ Thus, this admittedly crude test reveals no indication for anticipation effects.¹⁷ The finding is further supported by evidence presented in Waldinger (2007), who also concludes that anticipation effects are probably not very important.

6 Effect Heterogeneity

Overall, we find no strong evidence for a significant universal effect of early educational tracking on the relative achievement of migrant students, but results are also not very clear-cut and small effect sizes cannot be excluded. This may conceal a substantial detrimental effect of early tracking for specific groups of migrant students, while other migrant students indeed do not suffer from an earlier tracking age. Thus, it is less clear why early tracking should have any detrimental effect on relative achievement on children of migrants who are highly integrated into the host country's society. However, migrant students from less integrated families may have difficulties to signal their educational potential at an early stage of compulsory schooling and thus suffer more from a system that aims to match students to different educational pathways at an early age.

Proficiency in the national language is an obvious requirement for full participation in a country's society and consequently a good proxy for the degree of integration of migrants. Language proficiency itself is also essential to childrens educational success. Previous studies already showed that second-generation migrants perform substantially better when they are more proficient in the national language (e.g., Dustmann et al., 2012; Schneeweis, 2011). Moreover, Akresh and Akresh (2010) provide experimental evidence that foreign-born students have better test results when taking a test in their native language instead of the language of the host country.

¹⁶For Singapore, Austria, and Czech Republic *earlier grade* refers to grade 3 and *later grade* refers to grade 4. For the Netherlands, *earlier grade* refers to grade 5 and *later grade* refers to grade 6.

¹⁷Naturally, anticipation effects may still be relevant even in the absence of changes in migrant-native achievement gaps in later grades if native parents support the learning efforts of their children more than migrant parents in anticipation of the tracking decision already in first grade (or kindergarten).

Thus, proficiency in the test language may be a key source of effect heterogeneity. We can create a rough binary measure of test language proficiency by using self-reported information on whether students frequently speak the test language at home. We then extend the regression model in Equation (2) by adding interaction terms between the migrant indicator and the dummy variable that identifies migrant students who either only sometimes or never speak the test language at home. Table 9 presents the results of estimating this extended model.

Overall, the results confirm previous evidence that migrant students who either sometimes or never speak the test language at home perform worse in all three domains than migrant students who almost always speak the test language at home. More importantly, however, estimates in Rows 1 and 2 of Table 9 suggest that the effect of early educational tracking in fact depends crucially on the proficiency in the language of the test. For the baseline category of migrant students who speak the test language at home, the coefficients on the triple interaction are now insignificant, close to zero and even have positive signs. For migrants who do not speak the test language at home, we instead find that early tracking reduces their relative achievement by 11% of a standard deviation in reading. For math and science we find similar effects, which are, however, not significant because of large standard errors.

Table 10 tests for other effect heterogeneities. In this table, we estimate Equation (2) separately for subgroups of the student population. The type of subgroup analysis is indicated in the first column of Table 10. In Rows 1 and 2, we replicate the previous subgroup analysis along the language spoken at home divide in this more flexible way to provide a better benchmark. Other subgroups include boys, girls, and students within specific books-at-home categories. Additionally, we assess the robustness of our main results with respect to an alternative definition of migrant students. Specifically, we limit our attention to second-generation immigrants with both parents born abroad, excluding all other second-generation immigrants from the analysis.

In contrast to the effect heterogeneity that we observe with respect to the language spoken at home, we find very similar results for boys and girls as well as for students from different family backgrounds. The latter finding is particularly important as it shows that the detrimental effect of early tracking for migrant students who do not speak the test language at home is not a mere byproduct of a correlation between parental educational background and language proficiency. Overall, these results underline the importance of national language proficiency and integration as the key source of effect heterogeneity.

7 Conclusion

In this paper we investigate whether early ability-based tracking of students into different types of secondary school systematically increases migrant-native test score gaps. Based on individual test score data from 12 large-scale international tests, we show that estimates from cross-sectional regression models suggest large positive effects on migrant-native test score gaps of tracking students before the age of 15. However, estimates from differences-in-differences models that exploit variation in migrant-native test score gaps between primary and secondary school as well as variation in the age at tracking between countries reveal no significant effects. Although these estimates are not precise enough to rule small negative effects of early tracking, overall effect sizes suggested by cross-sectional estimates are clearly overstated.

A large set of robustness checks confirms our main findings. In particular, our results are not driven by a subset of countries or by our definition of early tracking. Yet, in terms of reading achievement, there is some weak indication for a detrimental effect of early tracking.

A closer inspection of the data, however, reveals an important effect heterogeneity with respect to how frequently migrant students speak the language of the test at home. For migrant students who almost never speak the test language at home, we find a significant detrimental effect of early tracking on relative achievement in reading of about 11% of a standard deviation and similar, but insignificant, effect sizes in terms of math and science achievement.

In the context of migrant inclusion, proficiency in the national language has been shown to be a key determinant of educational success for children of immigrants (e.g., Dustmann et al., 2012; Schneeweis, 2011). This study adds to this literature by providing first evidence on an important interaction between migrants' proficiency in the national language and the age of first tracking into different types of secondary schools by ability. Theoretically, such an interaction effect is expected if limited language proficiency decreases the ability of students to signal their true educational potential (Akresh and Akresh, 2010), which arguably matters more for the skill development of students in educational systems that aim to match students to different educational tracks at an early age.

A key implication of our findings is that tracking at later stages may not substantially reduce overall migrant-native test score gaps everywhere, but has the most scope for improving educational opportunities of children from less integrated families, who lack proficiency in the national language. Thus, calls for a more comprehensive school system to improve opportunities for migrant students appear to be particularly justified in settings in which proficiency in the national language constitutes a barrier to integration for the majority of migrant families.

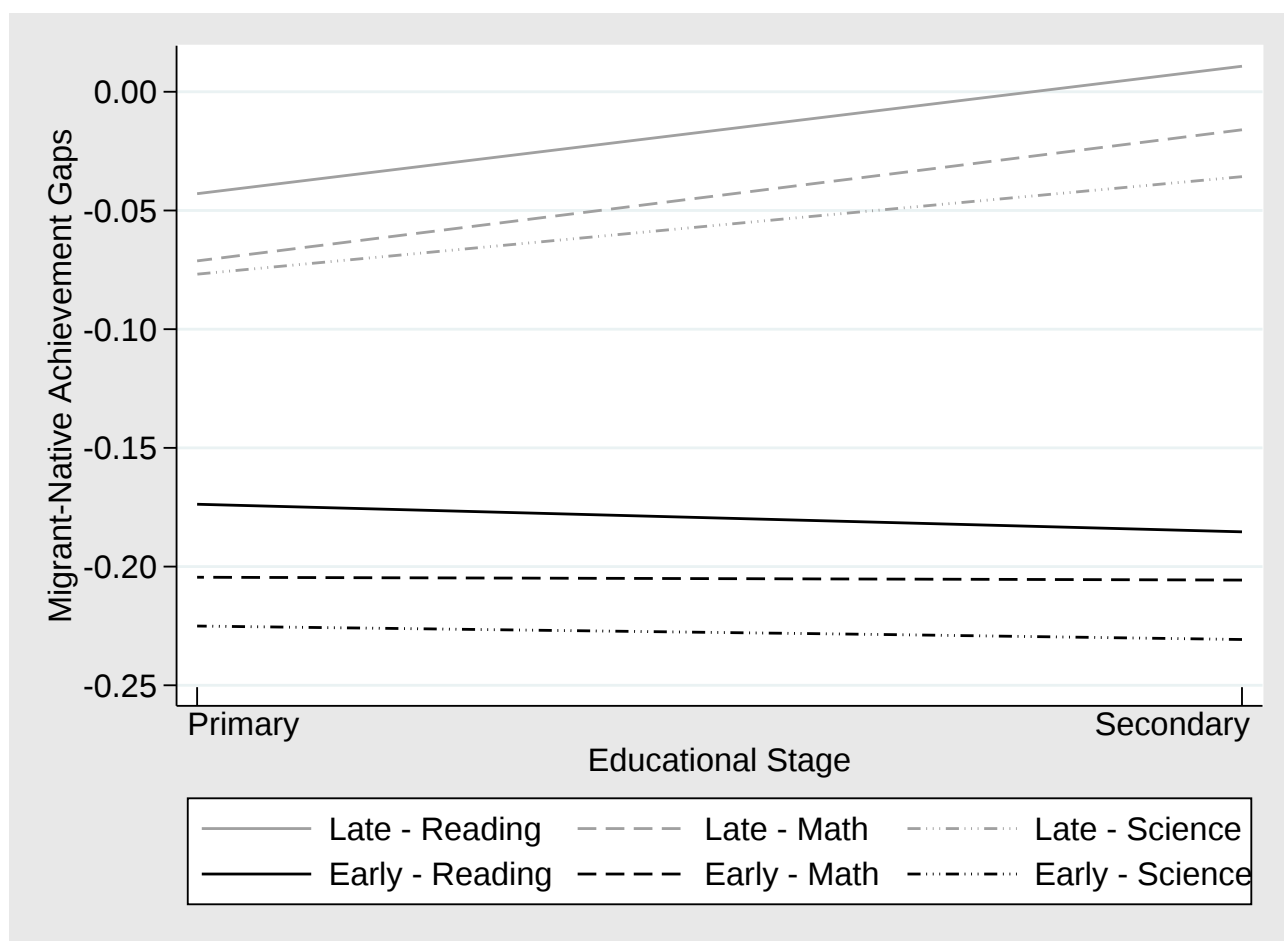
References

- Akresh, R. and Akresh, I. R. (2010). Using achievement tests to measure language assimilation and language bias among the children of immigrants. *Journal of Human Resources*, 46(3):647–667.
- Algan, Y., Dustmann, C., Glitz, A., and Manning, A. (2010). The Economic Situation of First- and Second-Generation Immigrants in France, Germany, and the UK. *Economic Journal*, 120(542):F4 – F30.
- Ammermueller, A. (2007). Poor Background or Low Returns? Why Immigrant Students in Germany Perform so Poorly in PISA. *Education Economics*, 15(2):215–230.
- Bauer, P. and Riphahn, R. T. (2006). Timing of school tracking as a determinant of intergenerational transmission of education. *Economics Letters*, 91(1):90–97.
- Bauer, P. and Riphahn, R. T. (2007). Heterogeneity in the intergenerational transmission of educational attainment: Evidence from Switzerland on natives and second-generation immigrants. *Journal of Population Economics*, 20(1):121–148.
- Brunello, G. and Checchi, D. (2007). Does school tracking affect equality of opportunity? New international evidence. *Economic Policy*, 22(52):781–861.
- Cobb-Clark, D. A., Sinning, M., and Stillman, S. (2012). Migrant Youths’ Educational Achievement: The Role of Institutions. *ANNALS of the American Academy of Political and Social Science*, 643(1):18–45.
- Danzer, A. M. and Yaman, F. (2013). Do Ethnic Enclaves Impede Immigrants’ Integration? Evidence from a Quasi-experimental Social-interaction Approach. *Review of International Economics*, 21(2):311–325.
- Dustmann, C., Frattini, T., and Lanza, G. (2012). Educational Achievement of Second Generation Immigrants: An International Comparison. *Economic Policy*, 27(69):143–185.
- Grogger, J. and Hanson, G. H. (2011). Income maximization and the selection and sorting of international migrants. *Journal of Development Economics*, 95(1):42–57.
- Hanushek, E. A., Piopiunik, M., and Wiederhold, S. (2014). The Value of Smarter Teachers: International Evidence on Teacher Cognitive Skills and Student Performance. Munich Discussion Papers No. 2014-51.

- Hanushek, E. A. and Woessmann, L. (2006). Does Educational Tracking affect Performance and Inequality? Differences-in-Differences Evidence across Countries. *Economic Journal*, 116(510):63–76.
- Kerr, S. P., Pekkarinen, T., and Uusitalo, R. (2013). School tracking and development of cognitive skills. *Journal of Labor Economics*, 31(3):577–602.
- Kiss, D. (2013). Are immigrants and girls graded worse? Results of a matching approach. *Education Economics*, 21(5):447–463.
- Lüdemann, E. and Schwerdt, G. (2013). Migration background and educational tracking: Is there a double disadvantage for second-generation immigrants? *Journal of Population Economics*, 26(2):455–481.
- Meghir, C. and Palme, M. (2005). Educational Reform, Ability, and Family Background. *American Economic Review*, 95(1):414–424.
- NESSE (2008). Education and Migration: Strategies for integrating migrant children in European schools and societies. Independent expert report submitted to the European Commission by the NESSE network. ISBN: 978-92-79-12804-2.
- Nielsen, H. S., Rosholm, M., Smith, N., and Husted, L. (2003). The school-to-work transition of 2nd generation immigrants in Denmark. *Journal of Population Economics*, 16(4):755–786.
- OECD (2007). *PISA 2006 - Volume 2: Data / Données*. OECD Publishing.
- OECD (2010). *PISA 2009 Results: What Makes a School Successful? Resources, Policies and Practices - Volume IV*. OECD Publishing.
- Pekkarinen, T., Uusitalo, R., and Kerr, S. (2009). School tracking and intergenerational income mobility: Evidence from the Finnish comprehensive school reform. *Journal of Public Economics*, 93(7-8):965–973.
- Schneeweis, N. (2011). Educational institutions and the integration of migrants. *Journal of Population Economics*, 24(4):1281–1308.
- Schnepf, S. V. (2007). Immigrants’ educational disadvantage: an examination across ten countries and three surveys. *Journal of Population Economics*, 20(3):527–545.

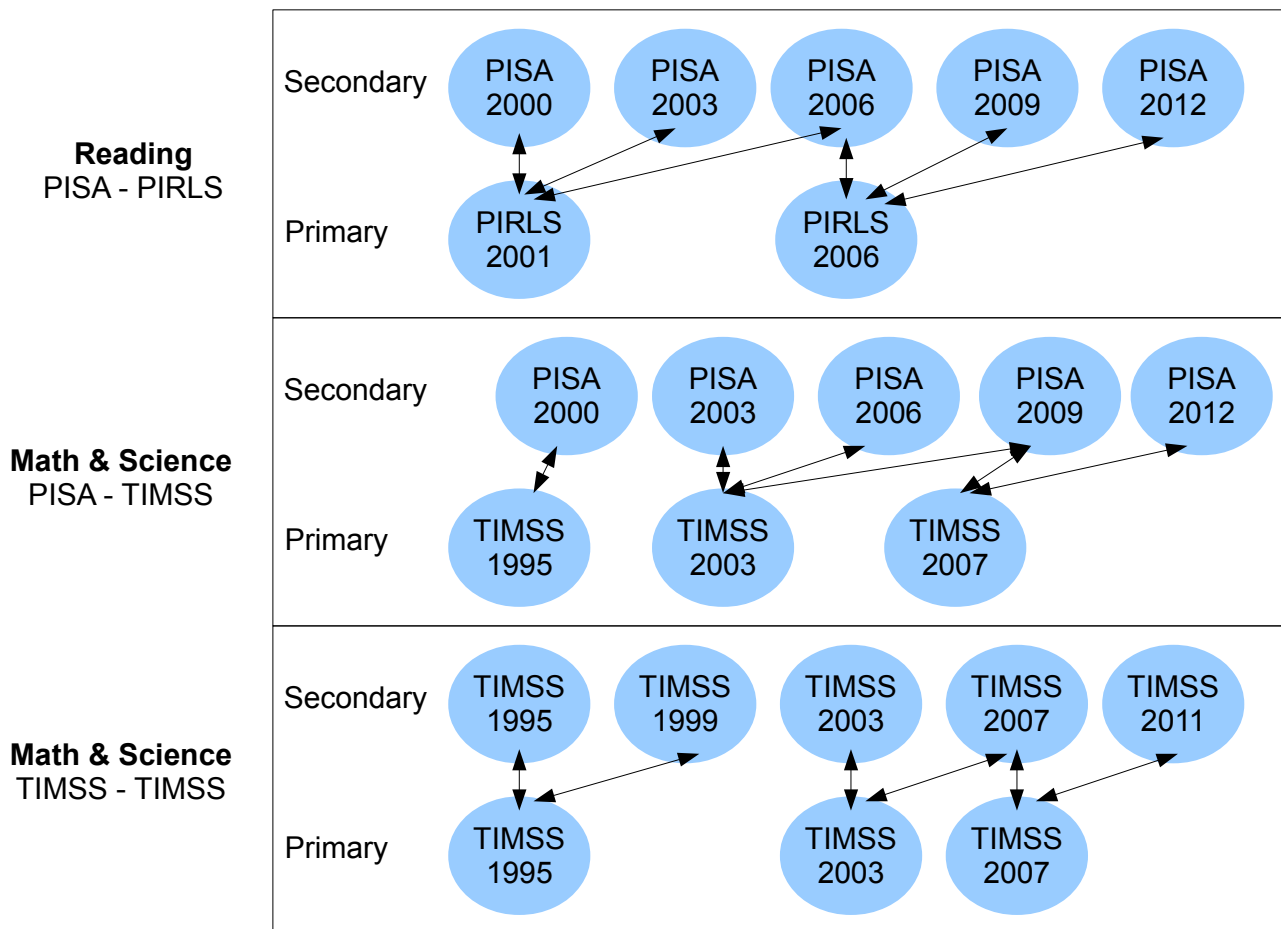
- UNESCO International Bureau of Education (2013). Country Dossiers. <http://www.ibe.unesco.org/en/worldwide.html>.
- Van de Werfhorst, H. G. and Mijs, J. J. (2010). Achievement Inequality and the Institutional Structure of Educational Systems: A Comparative Perspective. *Annual Review of Sociology*, 36(1):407–428.
- Van Ours, J. C. and Veenman, J. (2003). The educational attainment of second-generation immigrants in the Netherlands. *Journal of Population Economics*, 16(4):739–753.
- Waldinger, F. (2007). Does Ability Tracking Exacerbate the Role of Family Background for Students' Test Scores? Mimeo. University of Warwick.

Figure 1: Average Migrant-Native Achievement Gaps in Primary and Secondary School



Notes: The figure shows average migrant-native achievement gaps by subtracting the unconditional mean average migrant test score from the unconditional mean native test score (see summary statistics in Table 2). Means of the variables and are compiled by using all waves (all years) of TIMSS, PISA and PIRLS separately by primary and secondary school for the set of countries in the pooled application, i.e. corresponding to the baseline results. Student weights are used for weighting.

Figure 2: Matching of Survey Waves



Notes: The figure shows the matching of survey waves for the differences-in-differences model. Author's own depiction.

Table 1: *Tracking Status of Countries*

Late tracking				Early tracking			
Country	Baseline	Tracking		Country	Baseline	Tracking	
	sample	Age	Grade		sample	Age	Grade
Canada	yes	16	8	Austria	yes	10	4
Denmark	yes	16	10	Belgium	yes	12	7
Greece	yes	15	9	Czech Republic	yes	11	5
Hong Kong	yes	16	11	Germany	yes	10	4
Iceland	yes	16	10	Hungary	yes	11	4
Latvia	yes	16	9	Italy	yes	14	8
Lithuania	yes	15	8	Netherlands	yes	12	6
New Zealand	yes	16	12	Singapore	yes	12	6
norway	yes	16	10	Slovak Republic	yes	11	4
Qatar	yes	15	9	Ireland	no	12	6
Russia	yes	15	9	Luxembourg	no	13	6
Slovenia	yes	15	9	Macedonia	no	14	9
Sweden	yes	16	9	Mongolia	no	11	4
Taipei	yes	15	9	Trinidad and Tobago	no	11	6
United Kingdom	yes	16	12	United Arab Emirates	no	14	9
United States	yes	16	12				
Argentina	no	15	9				
Australia	no	16	10				
Cyprus	no	15	10				
El Salvador	no	16	9				
France	no	15	9				
Georgia	no	15	9				
Israel	no	15	10				
Kazakhstan	no	15	9				
Kuwait	no	17	12				
Moldova	no	15	10				
Morocco	no	16	12				
Portugal	no	15	9				
Spain	no	16	10				
Tunisia	no	16	10				

Notes: *Early tracking* is defined as the age of first selection before age 15. The age and grade of tracking describes the first selection in the school system. The UNESCO International Bureau of Education (2013) provides information about a large number of school systems around the world. We inferred the tracking status of a country from these information. The data was double-checked with OECD data (OECD (2007, Table 5.2, p. 162) and OECD (2010, Table A5.2, p. 147)) and with data from Brunello and Checchi (2007) and Waldinger (2007).

Table 2: *Summary Statistics – Test Scores*

<i>Panel A: Test scores in secondary school</i>										
Variable	Late tracking countries					Early tracking countries				
	Migrants		Natives		Difference	Migrants		Natives		Difference
	Mean	Obs	Mean	Obs		Mean	Obs	Mean	Obs	
Reading score	0.009 (0.879)	72,910	-0.002 (0.859)	327,765	0.011*** [0.004]	-0.161 (0.951)	24,725	0.024 (0.878)	199,072	-0.185*** [0.006]
Math score	-0.013 (0.809)	116,018	0.003 (0.795)	479,924	-0.016*** [0.003]	-0.179 (0.887)	31,939	0.026 (0.831)	225,690	-0.206*** [0.005]
Science score	-0.029 (0.869)	116,036	0.006 (0.820)	479,972	-0.036*** [0.003]	-0.201 (0.940)	31,933	0.03 (0.851)	225,737	-0.231*** [0.005]
<i>Panel B: Test scores in primary school</i>										
Variable	Late tracking countries					Early tracking countries				
	Migrants		Natives		Difference	Migrants		Natives		Difference
	Mean	Obs	Mean	Obs		Mean	Obs	Mean	Obs	
Reading score	-0.034 (0.716)	25,064	0.009 (0.673)	91,123	-0.043*** [0.005]	-0.147 (0.653)	8,986	0.026 (0.612)	46,827	-0.174*** [0.007]
Math score	-0.058 (0.699)	37,860	0.014 (0.678)	132,594	-0.071*** [0.004]	-0.174 (0.693)	13,213	0.031 (0.654)	62,803	-0.204*** [0.006]
Science score	-0.062 (0.749)	37,860	0.015 (0.695)	132,594	-0.077*** [0.004]	-0.191 (0.699)	13,213	0.034 (0.629)	62,803	-0.225*** [0.006]

Notes: Summary statistics show means of the variables and are compiled by using all waves (all years) of TIMSS, PISA and PIRLS separately by primary and secondary school for the set of countries in the pooled application, that is, corresponding to the baseline results. Test scores are de-meaned and standardized to have an international mean of 0 and a standard deviation of 1. Student weights are used for weighting. Each observation is weighted such that each country carries a weight of one. Standard deviations in round and standard errors in rectangular parentheses. A migrant is defined as second generation migrant: The student is born in the testing country but at least one parent is born outside this country. Significance levels for testing whether the difference is equal to zero or not: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3: Summary Statistics - Demographics and Socioeconomic Background

Panel A: Demographics and socioeconomic background in secondary school						
	Late tracking countries			Early tracking countries		
Variable	Migrants	Natives	Difference	Migrants	Natives	Difference
Age	15.062 (0.964)	15.097 (0.948)	-0.035 0.003	15.327 (0.816)	15.374 (0.784)	-0.048*** [0.004]
Male	0.492	0.497	-0.005*** [0.002]	0.497	0.503	-0.006** [0.003]
0–10 books	0.139	0.092	0.047*** [0.001]	0.156	0.091	0.065*** [0.002]
11–25 books	0.195	0.164	0.031*** [0.001]	0.194	0.157	0.037*** [0.002]
26–100 books	0.289	0.305	-0.016*** [0.001]	0.303	0.318	-0.014*** [0.003]
101–200 books	0.168	0.197	-0.029*** [0.001]	0.159	0.200	-0.042*** [0.002]
> 200 books	0.207	0.240	-0.033*** [0.001]	0.186	0.233	-0.046*** [0.002]
OECD	0.540	0.644	-0.104*** [0.002]	0.774	0.905	-0.131*** [0.002]
GDP per capita	35,817 (22,983)	30,867 (22,264)	4,950*** [70]	32,800 (12,132)	26,642 (12,575)	6,158*** [69]
Observations	126,325	543,206		37,674	276,946	
Panel B: Demographics and socioeconomic background in primary school						
	Late tracking countries			Early tracking countries		
Variable	Migrants	Natives	Difference	Migrants	Natives	Difference
Age	10.098 (0.672)	10.173 (0.687)	-0.075*** [0.003]	10.247 (0.561)	10.203 (0.537)	0.044*** [0.004]
Male	0.487	0.492	-0.006*** [0.002]	0.504	0.500	0.004 [0.004]
0–10 books	0.131	0.081	0.050*** [0.001]	0.138	0.074	0.064*** [0.002]
11–25 books	0.209	0.190	0.019*** [0.002]	0.256	0.216	0.040*** [0.003]
26–100 books	0.312	0.342	-0.031*** [0.002]	0.325	0.371	-0.046*** [0.004]
101–200 books	0.164	0.194	-0.030*** [0.002]	0.146	0.182	-0.036*** [0.003]
> 200 books	0.184	0.193	-0.009*** [0.002]	0.135	0.157	-0.022*** [0.003]
OECD	0.592	0.633	-0.041*** [0.002]	0.784	0.907	-0.123*** [0.002]
GDP per capita	31,579 (19,601)	27,602 (20,696)	3,977*** [92]	30,181 (10,870)	24,900 (12,374)	5,281*** [89]
Observations	62,924	223,717		22,199	109,630	

Notes: Summary statistics show means of the variables and are compiled by using all waves (all years) of TIMSS, PISA and PIRLS separately by primary and secondary school for the set of countries in the pooled application, that is, corresponding to the baseline results. The table shows the result of pooling samples by educational stage, that is, pooling PISA and TIMSS for secondary school and PIRLS and TIMSS for primary school. Numbers of observations between test score observations in Panel A of Table 2 do not add up compared to Panel A of this table because math and science test scores are reported both by PISA and TIMSS and PISA contains missing test scores. This summary statistic, however, reports background information for all students with at least one valid test score in one of the three subjects. In Panel B of both tables, number of observations add up because reading test scores are reported by PIRLS only and math and science test scores by TIMSS only. Neither of the surveys contains missing test scores. Student weights are used for weighting. Each observation is weighted such that each country carries a weight of one. Standard deviations in round and standard errors in rectangular parentheses. A migrant is defined as second generation migrant: The student is born in the testing country but at least one parent is born outside this country. Significance levels for testing whether the difference is equal to zero or not: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4: *OLS Estimation – Secondary School*

	(1)	(2)	(3)	(4)	(5)	(6)
	Reading		Math		Science	
	PISA		PISA, TIMSS		PISA, TIMSS	
Early tracking $\times$ migrant	−0.207*	−0.128**	−0.198*	−0.109*	−0.203	−0.097
	(0.105)	(0.058)	(0.103)	(0.056)	(0.120)	(0.057)
Migrant	0.012	0.103**	−0.018	0.086***	−0.039	0.107**
	(0.054)	(0.042)	(0.053)	(0.030)	(0.066)	(0.039)
<i>Further controls</i>						
GDP per capita $\times$ migrant		0.004**		0.003**		0.003*
		(0.002)		(0.001)		(0.002)
OECD $\times$ migrant		−0.180***		−0.201***		−0.272***
		(0.055)		(0.046)		(0.058)
Male		−0.326***		0.086***		0.076***
		(0.018)		(0.015)		(0.020)
Age		0.151***		0.008		0.042
		(0.010)		(0.027)		(0.029)
11-25 books		0.284***		0.265***		0.295***
		(0.025)		(0.027)		(0.026)
26-100 books		0.556***		0.542***		0.581***
		(0.035)		(0.037)		(0.035)
101-200 books		0.776***		0.747***		0.799***
		(0.043)		(0.042)		(0.041)
> 200 books		0.990***		0.936***		1.001***
		(0.055)		(0.053)		(0.053)
Constant	0.538***	−2.081***	0.864***	0.289	0.696***	−0.439
	(0.029)	(0.173)	(0.029)	(0.433)	(0.035)	(0.460)
Survey $\times$ country fixed effects	yes	yes	yes	yes	yes	yes
R^2	0.146	0.289	0.232	0.328	0.179	0.288
Individual observations	624,472	624,472	853,571	853,571	853,678	853,678
Distinct country observations	25	25	25	25	25	25
<i>Early tracking countries</i>	9	9	9	9	9	9

Notes: The samples are restricted to the set of countries used in the baseline DiD models. Dependent variables are the original test scores which are standardized within each survey and year to international mean 0 and sd 1. Robust standard errors in parentheses and clustered at the country level. Original students weights are used for weighting. Each observation is weighted such that each country carries a weight of one. *Early tracking* is 1 for countries with tracking age below 15 and 0 otherwise. The reference category for *books* is *0-10 books*. *GDP per capita* is yearly de-measured to provide deviations from a country with mean GDP per capita. A migrant is defined as second generation migrant: The student is born in the testing country but at least one parent is born outside this country. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 5: *OLS Estimation – Primary School*

	(1)	(2)	(3)	(4)	(5)	(6)
	Reading		Math		Science	
	PIRLS		TIMSS		TIMSS	
Early tracking $\times$ migrant	−0.138*	−0.077**	−0.136**	−0.092***	−0.151*	−0.091**
	(0.074)	(0.036)	(0.056)	(0.029)	(0.076)	(0.040)
Migrant	−0.046	0.081**	−0.077**	0.013	−0.083	0.039
	(0.039)	(0.033)	(0.034)	(0.026)	(0.051)	(0.032)
<i>Further controls</i>						
GDP per capita $\times$ migrant		0.003*		0.003***		0.004**
		(0.002)		(0.001)		(0.001)
OECD $\times$ migrant		−0.191***		−0.137***		−0.205***
		(0.042)		(0.032)		(0.045)
Male		−0.128***		0.060***		0.067***
		(0.013)		(0.008)		(0.010)
Age		−0.047*		0.132***		0.129***
		(0.025)		(0.046)		(0.042)
11-25 books		0.256***		0.230***		0.229***
		(0.021)		(0.029)		(0.028)
26-100 books		0.470***		0.443***		0.432***
		(0.033)		(0.037)		(0.039)
101-200 books		0.602***		0.556***		0.553***
		(0.043)		(0.045)		(0.047)
> 200 books		0.608***		0.571***		0.603***
		(0.049)		(0.054)		(0.058)
Constant	0.361***	0.470*	0.692***	−0.964**	0.222***	−1.432***
	(0.019)	(0.243)	(0.017)	(0.451)	(0.026)	(0.413)
Survey $\times$ country fixed effects	yes	yes	yes	yes	yes	yes
R^2	0.238	0.308	0.317	0.367	0.238	0.296
Individual observations	172,000	172,000	246,470	246,470	246,470	246,470
Distinct country observations	25	25	25	25	25	25
<i>Early tracking countries</i>	9	9	9	9	9	9

Notes: The samples are restricted to the set of countries used in the baseline DiD models. Dependent variables are the original test scores which are standardized within each survey and year to international mean 0 and sd 1. Standard errors clustered at the country level in parentheses. Original students weights are used for weighting. Each observation is weighted such that each country carries a weight of one. *Early tracking* is 1 for countries with tracking age below 15 and 0 otherwise. The reference category for *books* is *0-10 books*. *GDP per capita* is yearly de-meaned to provide deviations from a country with mean GDP per capita. A migrant is defined as second generation migrant: The student is born in the testing country but at least one parent is born outside this country. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 6: *Differences-In-Differences Estimation – Baseline Results*

	(1)	(2)	(3)	(4)	(5)	(6)
	Reading		Math		Science	
Secondary: Primary:	PISA PIRLS		PISA, TIMSS TIMSS		PISA, TIMSS TIMSS	
Early tracking × secondary × migrant	−0.072* (0.039)	−0.048 (0.032)	−0.062 (0.057)	−0.037 (0.045)	−0.050 (0.064)	−0.020 (0.052)
Early tracking × migrant	−0.139* (0.074)	−0.082** (0.037)	−0.136** (0.056)	−0.080** (0.029)	−0.153* (0.076)	−0.084** (0.040)
Migrant × secondary	0.059** (0.025)	0.037* (0.019)	0.055* (0.027)	0.038** (0.015)	0.039* (0.022)	0.019 (0.012)
Migrant	−0.046 (0.039)	0.074** (0.032)	−0.077** (0.034)	0.030 (0.025)	−0.081 (0.051)	0.062* (0.035)
<i>Further controls</i>						
GDP per capita × migrant		0.003* (0.002)		0.003*** (0.001)		0.004** (0.002)
OECD × migrant		−0.186*** (0.047)		−0.170*** (0.036)		−0.236*** (0.050)
Male		−0.227*** (0.013)		0.074*** (0.011)		0.069*** (0.013)
Age		0.024 (0.019)		0.068* (0.035)		0.082** (0.034)
Secondary × 11-25 books		0.033 (0.019)		0.038 (0.030)		0.071*** (0.021)
Secondary × 26-100 books		0.094*** (0.023)		0.109*** (0.034)		0.161*** (0.025)
Secondary × 101-200 books		0.186*** (0.025)		0.204*** (0.039)		0.260*** (0.030)
Secondary × > 200 books		0.388*** (0.030)		0.379*** (0.045)		0.417*** (0.037)
11-25 books		0.253*** (0.022)		0.226*** (0.029)		0.224*** (0.029)
26-100 books		0.467*** (0.034)		0.435*** (0.037)		0.423*** (0.039)
101-200 books		0.599*** (0.044)		0.547*** (0.044)		0.541*** (0.046)
> 200 books		0.610*** (0.049)		0.561*** (0.054)		0.588*** (0.058)
Constant	0.538*** (0.029)	−0.131 (0.295)	0.865*** (0.029)	−0.637 (0.548)	0.697*** (0.035)	−1.058* (0.534)
Survey × country fixed effects	yes	yes	yes	yes	yes	yes
R^2	0.186	0.295	0.271	0.346	0.207	0.294
Individual observations	1,209,793	1,209,793	1,832,197	1,832,197	1,832,304	1,832,304
Distinct country observations	25	25	25	25	25	25
Early tracking countries	9	9	9	9	9	9

Notes: Dependent variables are the original test scores which are standardized within each survey and year to international mean 0 and sd 1. Standard errors clustered at the country level in parentheses. Original students weights are used for weighting. Each observation is weighted such that each country carries a weight of one within each educational stage. *Early tracking* is 1 for countries with tracking age below 15 and 0 otherwise. *Secondary* is equal to 1 for testing in secondary school at age 14/15 (grade 8/9) and 0 for testing in primary school at age 9/10 (grade 4). A migrant is defined as second generation migrant: The student is born in the testing country but at least one parent is born outside this country. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 7: *Robustness Checks*

	(1)	(2)	(3)	(4)	(5)	(6)
	Reading		Math		Science	
Secondary: Primary:	PISA PIRLS		PISA, TIMSS TIMSS		PISA, TIMSS TIMSS	
[1] All countries	−0.072* (0.036)	−0.049 (0.032)	−0.056 (0.068)	−0.035 (0.061)	−0.037 (0.069)	−0.019 (0.059)
[2] Common set of countries	−0.080* (0.041)	−0.064* (0.033)	−0.090 (0.063)	−0.064 (0.052)	−0.088 (0.079)	−0.056 (0.068)
[3] OECD countries	−0.091** (0.034)	−0.066** (0.028)	−0.043 (0.058)	−0.038 (0.056)	−0.039 (0.060)	−0.032 (0.058)
[4] Main destination countries	−0.131*** (0.032)	−0.078*** (0.024)	−0.094*** (0.028)	−0.066** (0.023)	−0.074** (0.023)	−0.040 (0.022)
[5] Migrant ratio > 10%	−0.099** (0.043)	−0.077** (0.031)	−0.107* (0.060)	−0.081 (0.049)	−0.095 (0.071)	−0.062 (0.062)
[6] Tracking grade 6	−0.062 (0.041)	−0.037 (0.028)	−0.035 (0.056)	−0.010 (0.036)	−0.012 (0.057)	0.017 (0.035)
[7] Tracking after grade 5	−0.076 (0.051)	−0.059 (0.040)	−0.072 (0.080)	−0.051 (0.067)	−0.078 (0.105)	−0.053 (0.090)
[8] Continuous tracking age	0.015* (0.008)	0.010 (0.006)	0.012 (0.011)	0.007 (0.008)	0.009 (0.011)	0.003 (0.009)
[9] Tracking grade	−0.057 (0.040)	−0.048 (0.032)	−0.037 (0.053)	−0.029 (0.039)	−0.043 (0.054)	−0.032 (0.043)
[10] Continuous tracking grade	0.010 (0.006)	0.008 (0.005)	0.005 (0.009)	0.004 (0.006)	0.004 (0.009)	0.002 (0.006)
[11] Same cohorts	−0.078* (0.039)	−0.049 (0.029)	−0.076 (0.062)	−0.050 (0.049)	−0.058 (0.066)	−0.026 (0.054)
[12] Same time	−0.101* (0.050)	−0.069 (0.050)	−0.042 (0.059)	−0.024 (0.054)	−0.048 (0.077)	−0.024 (0.067)
Further controls	–	yes	–	yes	–	yes
Survey × country fixed effects	yes	yes	yes	yes	yes	yes

Notes: All entries show the coefficient on the triple interaction between Early Tracking × Secondary × Migrant. Dependent variables are the original test scores which are standardized within each survey and year to international mean 0 and sd 1. Standard errors clustered at the country level in parentheses. Original students weights are used for weighting. Each observation is weighted such that each country carries a weight of one within each educational stage. *Further controls:* early tracking × migrant, migrant × secondary, migrant, GDP per capita × migrant, OECD × migrant, male, age, secondary × 11–25 books, secondary × 26–100 books, secondary × 101–200 books, secondary × > 200 books, dummies for books at home. *Early tracking* is 1 for countries with tracking age below 15 and 0 otherwise. *Secondary* is equal to 1 for testing in secondary school at age 14/15 (grade 8/9) and 0 for testing in primary school at age 9/10 (grade 4). For continuous tracking age, tracking grade, and continuous tracking grade, the early tracking dummy is replaced by continuous tracking age, a dummy for tracking before grade 9, or continuous tracking grade respectively. A migrant is defined as second generation migrant: The student is born in the testing country but at least one parent is born outside this country. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 8: *Piecewise Deletion of Countries*

	(1)		(2)		(3)	
	Reading		Math		Science	
	PISA PIRLS		PISA, TIMSS TIMSS		PISA, TIMSS TIMSS	
	<i>Coefficient</i>	<i>Std. error</i>	<i>Coefficient</i>	<i>Std. error</i>	<i>Coefficient</i>	<i>Std. error</i>
<i>Excluded country</i>						
Baseline	-0.048	(0.032)	-0.037	(0.045)	-0.020	(0.052)
<i>Panel A: Early tracking countries</i>						
Austria	-0.043	(0.034)	-0.033	(0.051)	-0.019	(0.060)
Belgium	-0.028	(0.030)	-0.000	(0.036)	0.027	(0.034)
Czech Republic	-0.051	(0.033)	-0.037	(0.048)	-0.022	(0.055)
Germany	-0.040	(0.034)	-0.027	(0.052)	-0.016	(0.063)
Hungary	-0.056*	(0.031)	-0.046	(0.044)	-0.027	(0.053)
Italy	-0.056*	(0.031)	-0.048	(0.046)	-0.031	(0.054)
Netherlands	-0.044	(0.035)	-0.035	(0.051)	-0.011	(0.059)
Singapore	-0.057*	(0.033)	-0.055	(0.048)	-0.045	(0.055)
Slovak Republic	-0.053	(0.032)	-0.051	(0.043)	-0.033	(0.052)
<i>Panel B: Late tracking countries</i>						
Canada	-0.044	(0.033)	-0.037	(0.046)	-0.024	(0.052)
Denmark	-0.054*	(0.031)	-0.043	(0.045)	-0.027	(0.052)
Greece	-0.047	(0.032)	-0.035	(0.045)	-0.019	(0.052)
Hong Kong	-0.056*	(0.032)	-0.042	(0.046)	-0.020	(0.053)
Iceland	-0.045	(0.032)	-0.038	(0.045)	-0.020	(0.052)
Latvia	-0.054	(0.032)	-0.040	(0.045)	-0.022	(0.052)
Lithuania	-0.054*	(0.031)	-0.037	(0.045)	-0.022	(0.052)
New Zealand	-0.051	(0.032)	-0.041	(0.045)	-0.022	(0.052)
Norway	-0.052	(0.032)	-0.040	(0.045)	-0.024	(0.052)
Qatar	-0.029	(0.031)	-0.021	(0.044)	-0.011	(0.054)
Russia	-0.045	(0.032)	-0.035	(0.045)	-0.017	(0.052)
Slovenia	-0.045	(0.032)	-0.040	(0.045)	-0.022	(0.052)
Sweden	-0.048	(0.033)	-0.043	(0.045)	-0.023	(0.052)
Taipei	-0.043	(0.032)	-0.036	(0.046)	-0.015	(0.052)
United Kingdom	-0.045	(0.032)	-0.033	(0.045)	-0.016	(0.052)
United States	-0.040	(0.032)	-0.035	(0.045)	-0.016	(0.052)

Notes: All entries show the coefficient on the triple interaction between Early Tracking $\times$ Secondary $\times$ Migrant. Dependent variables are the original test scores which are standardized within each survey and year to international mean 0 and sd 1. Standard errors clustered at the country level in parentheses. Original students weights are used for weighting. Each observation is weighted such that each country carries a weight of one within each educational stage. All regressions include survey $\times$ country fixed effects and all further controls. *Further controls:* early tracking $\times$ migrant, migrant $\times$ secondary, migrant, GDP per capita $\times$ migrant, OECD $\times$ migrant, male, age, secondary $\times$ 11–25 books, secondary $\times$ 26–100 books, secondary $\times$ 101–200 books, secondary $\times$ > 200 books, dummies for books at home. *Early tracking* is 1 for countries with tracking age below 15 and 0 otherwise. *Secondary* is equal to 1 for testing in secondary school at age 14/15 (grade 8/9) and 0 for testing in primary school at age 9/10 (grade 4). For continuous tracking age, tracking grade, and continuous tracking grade, the early tracking dummy is replaced by continuous tracking age, a dummy for tracking before grade 9, or continuous tracking grade respectively. A migrant is defined as second generation migrant: The student is born in the testing country but at least one parent is born outside this country. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 9: *The Role of Language Spoken at Home*

	(1)	(2)	(3)
	Reading	Math	Science
Secondary:	PISA	PISA, TIMSS	PISA, TIMSS
Primary:	PIRLS	TIMSS	TIMSS
Early tracking $\times$ secondary $\times$ migrant	0.003 (0.036)	0.017 (0.040)	0.047 (0.049)
Early tracking $\times$ secondary $\times$ migrant $\times$ foreign language	-0.109** (0.048)	-0.109 (0.082)	-0.142 (0.085)
Early tracking $\times$ migrant	-0.028 (0.045)	-0.048 (0.039)	-0.049 (0.048)
Migrant $\times$ secondary	0.027 (0.020)	0.027** (0.011)	0.006 (0.016)
Migrant	0.131*** (0.032)	0.065** (0.030)	0.128*** (0.045)
Early tracking $\times$ migrant $\times$ foreign language	-0.047 (0.049)	-0.045 (0.035)	0.011 (0.061)
Migrant $\times$ secondary $\times$ foreign language	-0.037 (0.042)	0.102 (0.060)	0.101 (0.061)
Migrant $\times$ foreign language	-0.124*** (0.034)	-0.165*** (0.026)	-0.308*** (0.061)
Constant	-0.149 (0.298)	-0.608 (0.555)	-1.049* (0.542)
Further controls	yes	yes	yes
Survey $\times$ country fixed effects	yes	yes	yes
R^2	0.296	0.347	0.297
Individual observations	1,198,521	1,809,465	1,809,568
Distinct country observations	25	25	25
<i>Early tracking countries</i>	9	9	9

Notes: Dependent variables are the original test scores which are standardized within each survey and year to international mean 0 and sd 1. Standard errors clustered at the country level in parentheses. Original students weights are used for weighting. Each observation is weighted such that each country carries a weight of one within each educational stage. *Early tracking* is 1 for countries with tracking age below 15 and 0 otherwise. *Secondary* is equal to 1 for testing in secondary school at age 14/15 (grade 8/9) and 0 for testing in primary school at age 9/10 (grade 4). A migrant is defined as second generation migrant: The student is born in the testing country but at least one parent is born outside this country. *Foreign language* is 1 when the student does not speak (or speaks only sometimes) the language of the test country at home. We recode the foreign language variable to zero for natives who report that they do not speak the language of the test country at home. *Further controls:* GDP per capita $\times$ migrant, OECD $\times$ migrant, male, age, secondary $\times$ 11-25 books, secondary $\times$ 26-100 books, secondary $\times$ 101-200 books, secondary $\times$ > 200 books, 11-25 books, 26-100 books, 101-200 books, and > 200 books. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 10: *Subgroup Analysis*

	(1)	(2)	(3)	(4)	(5)	(6)
	Reading		Math		Science	
Secondary: Primary:	PISA PIRLS		PISA, TIMSS TIMSS		PISA, TIMSS TIMSS	
[1] Language sometimes or never spoken	−0.217* (0.110)	−0.174*** (0.057)	−0.186 (0.158)	−0.100 (0.104)	−0.203 (0.170)	−0.126 (0.128)
[2] Language (almost) always spoken	−0.045 (0.039)	−0.020 (0.029)	−0.020 (0.045)	−0.008 (0.036)	−0.005 (0.052)	0.011 (0.043)
[3] Books at home: 0–25 books	−0.045 (0.044)	−0.042 (0.042)	−0.030 (0.053)	−0.022 (0.050)	−0.017 (0.062)	−0.007 (0.058)
[4] Books at home: 26–100 books	−0.055 (0.036)	−0.051 (0.033)	−0.043 (0.060)	−0.036 (0.055)	−0.024 (0.068)	−0.015 (0.063)
[5] Books at home: 101–200 books	−0.044 (0.049)	−0.047 (0.044)	−0.067 (0.048)	−0.057 (0.044)	−0.053 (0.052)	−0.042 (0.048)
[6] Books at home: > 200 books	−0.013 (0.045)	−0.012 (0.045)	−0.040 (0.048)	−0.029 (0.047)	−0.012 (0.051)	−0.000 (0.049)
[7] Male	−0.084* (0.045)	−0.058 (0.040)	−0.073 (0.065)	−0.055 (0.051)	−0.056 (0.066)	−0.031 (0.053)
[8] Female	−0.062 (0.043)	−0.040 (0.038)	−0.050 (0.057)	−0.021 (0.049)	−0.043 (0.067)	−0.010 (0.057)
[9] Both parents born abroad	−0.136** (0.061)	−0.070 (0.044)	−0.149* (0.086)	−0.073 (0.068)	−0.149 (0.091)	−0.058 (0.078)
Further controls	–	yes	–	yes	–	yes
Survey × country fixed effects	yes	yes	yes	yes	yes	yes

Notes: All entries show the coefficient on the triple interaction between Early Tracking × Secondary × Migrant. Dependent variables are the original test scores which are standardized within each survey and year to international mean 0 and sd 1. Standard errors clustered at the country level in parentheses. Original students weights are used for weighting. Each observation is weighted such that each country carries a weight of one within each educational stage. *Further controls:* early tracking × migrant, migrant × secondary, migrant, GDP per capita × migrant, OECD × migrant, male, age, secondary × 11–25 books, secondary × 26–100 books, secondary × 101–200 books, secondary × > 200 books, dummies for books at home. *Early tracking* is 1 for countries with tracking age below 15 and 0 otherwise. *Secondary* is equal to 1 for testing in secondary school at age 14/15 (grade 8/9) and 0 for testing in primary school at age 9/10 (grade 4). Except in Row 9, a migrant is defined as second generation migrant: The student is born in the testing country but at least one parent is born outside this country. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table A-1: *Country List by Matched Survey Waves*

Country	Baseline sample	PISA-PIRLS						TIMSS-TIMSS						PISA-TIMSS					
		S2003 P2001	S2000 P2001	S2006 P2001	S2006 P2006	S2009 P2006	S2012 P2006	S1995 P1995	S2003 P2003	S2007 P2007	S1999 P1995	S2007 P2003	S2011 P2007	S2000 P1995	S2003 P2003	S2006 P2003	S2009 P2003	S2009 P2007	S2012 P2007
<i>Early tracking</i>																			
Austria	yes				yes	yes	yes	yes						yes				yes	yes
Belgium	yes				yes	yes	yes		yes						yes	yes	yes		
Czech Republic	yes	yes	yes	yes				yes		yes	yes			yes				yes	yes
Germany	yes	yes	yes	yes	yes	yes	yes											yes	yes
Hungary	yes					yes	yes			yes							yes	yes	yes
Italy	yes	yes	yes	yes	yes	yes	yes					yes	yes				yes	yes	yes
Netherlands	yes	yes	yes	yes	yes	yes	yes	yes	yes		yes			yes	yes	yes	yes	yes	yes
Singapore	yes					yes	yes	yes	yes	yes			yes	yes	yes		yes	yes	yes
Slovak Republic	yes	yes		yes	yes	yes	yes			yes								yes	yes
Ireland	no							yes						yes					yes
Luxembourg	no				yes	yes	yes												
Macedonia	no		yes																
Mongolia	no									yes									
Trinidad and Tobago	no					yes													
United Arab Emirates	no																	yes	yes
<i>Late tracking</i>																			
Canada	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Denmark	yes				yes	yes	yes											yes	yes
Greece	yes	yes	yes	yes				yes						yes					
Hong Kong	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Iceland	yes	yes	yes	yes	yes	yes	yes	yes						yes					
Latvia	yes	yes	yes	yes	yes	yes	yes	yes	yes		yes			yes	yes	yes	yes	yes	yes
Lithuania	yes			yes	yes	yes	yes		yes	yes			yes			yes	yes	yes	yes
New Zealand	yes	yes	yes	yes	yes	yes	yes	yes	yes		yes		yes	yes	yes	yes	yes	yes	yes
norway	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes			yes	yes	yes	yes	yes	yes	yes
Qatar	yes				yes	yes	yes			yes			yes					yes	yes
Russia	yes	yes	yes	yes	yes	yes	yes		yes	yes		yes	yes		yes	yes	yes	yes	yes
Slovenia	yes			yes	yes	yes	yes	yes	yes	yes	yes		yes	yes		yes	yes	yes	yes
Sweden	yes	yes	yes	yes	yes	yes	yes			yes			yes					yes	yes
Taipei	yes						yes						yes						yes
United Kingdom	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
United States	yes	yes	yes			yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Argentina	no		yes	yes															
Australia	no							yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Cyprus	no							yes	yes		yes	yes							
El Salvador	no									yes									
France	no	yes	yes	yes	yes	yes	yes												
Georgia	no									yes			yes						
Israel	no		yes	yes	yes	yes	yes												
Kazakhstan	no													yes				yes	yes
Kuwait	no							yes		yes									
Moldova	no								yes										
Morocco	no								yes	yes	yes		yes						
Portugal	no							yes						yes					
Spain	no				yes	yes	yes												
Tunisia	no									yes			yes						
45		17	19	20	23	27	27	19	17	20	12	13	18	15	12	14	16	24	25

Notes: *Early tracking* is defined as the age of first selection before age 15. The letter in front of the survey year indicates testing in primary school (P) or secondary school (S).

Table A-2: OLS Estimation – Secondary School by Survey

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Reading		Math				Science			
	PISA		PISA		TIMSS		PISA		TIMSS	
Early tracking $\times$ migrant	−0.207*	−0.128**	−0.193*	−0.100	−0.161	−0.098*	−0.205*	−0.101	−0.159	−0.088
	(0.105)	(0.058)	(0.110)	(0.062)	(0.108)	(0.053)	(0.120)	(0.061)	(0.136)	(0.056)
Migrant	0.012	0.103**	−0.029	0.080**	0.013	0.087***	−0.043	0.095**	−0.016	0.114***
	(0.054)	(0.042)	(0.056)	(0.037)	(0.050)	(0.021)	(0.063)	(0.044)	(0.072)	(0.033)
<i>Further controls</i>										
GDP per capita $\times$ migrant		0.004**		0.003**		0.004***		0.003*		0.004**
		(0.002)		(0.002)		(0.001)		(0.002)		(0.002)
OECD $\times$ migrant		−0.180***		−0.202***		−0.186***		−0.247***		−0.287***
		(0.055)		(0.053)		(0.037)		(0.059)		(0.054)
Male		−0.326***		0.114***		0.041**		0.037*		0.122***
		(0.018)		(0.015)		(0.015)		(0.018)		(0.030)
Age		0.151***		0.149***		−0.066*		0.154***		−0.022
		(0.010)		(0.014)		(0.036)		(0.008)		(0.036)
11-25 books		0.284***		0.264***		0.235***		0.294***		0.262***
		(0.025)		(0.025)		(0.033)		(0.025)		(0.027)
26-100 books		0.556***		0.548***		0.485***		0.579***		0.525***
		(0.035)		(0.035)		(0.043)		(0.034)		(0.040)
101-200 books		0.776***		0.767***		0.662***		0.805***		0.719***
		(0.043)		(0.041)		(0.050)		(0.041)		(0.049)
> 200 books		0.990***		1.003***		0.775***		1.041***		0.863***
		(0.055)		(0.054)		(0.056)		(0.055)		(0.056)
Constant	0.538***	−2.081***	0.870***	−1.939***	0.725***	1.174**	0.697***	−2.185***	0.108**	−0.149
	(0.029)	(0.173)	(0.031)	(0.233)	(0.027)	(0.493)	(0.034)	(0.138)	(0.039)	(0.509)
Country $\times$ Survey fixed effects	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
R^2	0.146	0.289	0.178	0.294	0.322	0.391	0.140	0.262	0.245	0.330
Individual observations	624,472	624,472	521,593	521,593	331,978	331,978	521,700	521,700	331,978	331,978
Distinct country observations	25	25	25	25	21	21	25	25	21	21
Early tracking countries	9	9	9	9	6	6	9	9	6	6

Notes: The samples are restricted to the set of countries used in the baseline DiD models. Dependent variables are the original test scores which are standardized within each survey and year to international mean 0 and sd 1. Standard errors clustered at the country level in parentheses. Original students weights are used for weighting. Each observation is weighted such that each country carries a weight of one. *Early tracking* is 1 for countries with tracking age below 15 and 0 otherwise. The reference category for *books* is 0-10 books. *GDP per capita* is yearly demeaned to provide deviations from a country with mean GDP per capita. A migrant is defined as second generation migrant: The student is born in the testing country but at least one parent is born outside this country. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table A-3: *Migrant Achievement Gaps in Primary School in TIMSS 1995*

	(1)	(2)	(3)	(4)	(5)	(6)
	Math			Science		
	Both Grades	Earlier grade	Later grade	Both Grades	Earlier grade	Later grade
Migrant	−0.120*** (0.016)	−0.099*** (0.021)	−0.116*** (0.022)	−0.131*** (0.016)	−0.103*** (0.022)	−0.138*** (0.022)
Male	0.042*** (0.012)	0.069*** (0.015)	0.067*** (0.016)	0.132*** (0.011)	0.127*** (0.016)	0.176*** (0.015)
Age	0.383*** (0.010)	−0.089*** (0.020)	−0.150*** (0.020)	0.315*** (0.009)	−0.046** (0.020)	−0.067*** (0.019)
11-25 books	0.324*** (0.029)	0.309*** (0.034)	0.280*** (0.040)	0.340*** (0.028)	0.323*** (0.037)	0.312*** (0.037)
26-100 books	0.692*** (0.027)	0.606*** (0.032)	0.617*** (0.037)	0.672*** (0.026)	0.630*** (0.034)	0.599*** (0.034)
101-200 books	0.871*** (0.028)	0.760*** (0.033)	0.813*** (0.039)	0.868*** (0.027)	0.798*** (0.036)	0.817*** (0.036)
> 200 books	0.908*** (0.028)	0.795*** (0.033)	0.888*** (0.040)	0.949*** (0.027)	0.866*** (0.035)	0.937*** (0.036)
Constant	−4.149*** (0.100)	0.183 (0.189)	1.589*** (0.208)	−3.371*** (0.096)	−0.044 (0.191)	0.726*** (0.201)
Country fixed effects	yes	yes	yes	yes	yes	yes
R^2	0.208	0.167	0.164	0.168	0.114	0.127
Individual observations	26,629	13,131	13,498	26,629	13,131	13,498
Distinct country observations	4	4	4	4	4	4

Notes: The sample is restricted to early tracking countries in TIMSS in primary schooling: Netherlands, Singapore, Austria, and Czech Republic. For Singapore, Austria, and Czech Republic *earlier grade* refers to grade 3 and *later grade* refers to grade 4. For the Netherlands, *earlier grade* refers to grade 5 and *later grade* refers to grade 6. Dependent variables are the original test scores which are standardized within each survey and year to international mean 0 and sd 1. Robust standard errors in parentheses. Original students weights are used for weighting. Each observation is weighted such that each country carries a weight of one. The reference category for *books* is *0-10 books*. A migrant is defined as second generation migrant: The student is born in the testing country but at least one parent is born outside this country. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.