"The effect of a Specialized versus a General Upper-Secondary school curriculum on students' performance and inequality. A differences-in-differences cross country analysis"

Ana Balcão Reis, Afonso Câmara Leme, Luís Catela Nunes, Josep-Oriol Escardíbul, Carmo Seabra*

ABSTRACT: Countries differ in their upper secondary school systems in a way that some require their students to choose a specialization from a set of areas - typically natural sciences, economic sciences, humanities or arts - and follow that specialization for the course of their upper secondary education years (e.g. Portugal, Spain, Sweden) whereas by contrast, others including Finland, Germany or the U.S. follow a general curriculum where students, albeit being able to choose between different classes in distinct areas, are not required to follow a single specialization and thus, receive a more general education. Because countries only follow one system or the other, a cross-country analysis is required to estimate the possible effects of these institutional differences. An international differences-in-differences approach is chosen to account for country heterogeneity and unobserved factors influencing student outcomes, by using both PISA and PIAAC data for 20 different countries. We also run quantile regressions to explore whether the existence of a specialized system has different effects on different performance quantiles. While country-level regressions results suggest that the choice of one system or the other does not account for differences across countries in either the mean performance or the inequality of students' test scores, regressions with student micro-level data suggest that there is a positive effect in Mathematics achievement for students following a specialized academic upper-secondary school system. However, these results lose their statistical significance when standard errors are clustered at the country-level.

JEL Codes: I21, I24, I28
Keywords: PISA, PIAAC, Education inequality, Tracking, Specialization, Difference- in-Differences, Curriculum

* Authors' contacts: Ana Balcão Reis (abr@novasbe.pt); Afonso Câmara Leme (afonso.scl@gmail.com); Luís Catela Nunes (lcnunes@novasbe.pt); Josep-Oriol Escardíbul (oescardibul@ub.es); Carmo Seabra (carmo.seabra@novasbe.pt)

1. Introduction

The efficiency and equity of a country's educational system is a topic of major concern and interest for policymakers, researchers, parents and students alike. Educational systems can be considered widely different across countries regarding their characteristics. Nevertheless, one distinction that can be clearly made is whether a country applies educational tracking and the timing in which it does so. Tracking can be generally described as separating students into different school classes, usually by academic ability or curriculum focus – students whose overall achievement is above average are assigned to the same class, as are students whose achievement is at an average level or below it. Usually, this separation also determines whether a student follows an academic-based track that prepares them for entrance in a university or a vocational-based track that on contrast, is designed to prepare students for entrance in the labor market following the conclusion of their school years. Furthermore, tracking can occur within schools, as is more common in the US - a form of ability-grouping also called streaming - or it can lead to a division of students between different schools, as is the norm in countries such as Germany, the Netherlands and several other European countries (Maaz et al., 2008). Some countries such as Germany and Austria track students as early as age 10, while the majority of OECD countries does so at age 15 or 16 (Woessmann, 2009). The discussion of the advantages and disadvantages of early tracking has been a topic of much heated debate for a long time (see Heath (1984) for a collection of arguments for and against ability-grouping between schools in the UK). Those in favor of it tend to argue that homogenous classes allow for a focused curriculum adapted to the level and learning pace of each student, in which the professor doesn't have to worry about losing the interest of fast learners or making the slowest ones fall behind, resulting into an optimal learning by all students. The arguments for comprehensive schools, as opposed to tracking, claim that lower-level students will be systematically disadvantaged by slower learning environments if they are separated early on by tracking, leaving them farther behind the level of the upper groups, and thus, increasing inequality – both in their educational achievement and later in their life, through lower earnings in their adult years (Pekkarinen et al., 2009). This possibility that tracking aggravates economic inequality is also related to the fear that it leads to a distribution of students through socio-economic background, perpetuating a bias against more disadvantaged students (Brunello & Chechi, 2007). Furthermore, opposers of tracking, by assuming non-linear peer effects, can argue that in heterogeneous classes, the higher ability students lose nothing while lower ability ones benefit from this interaction, giving a raise to efficiency (Benabou, 1996). On the other hand, if it is assumed that students are better off with peers of their own level, tracking could even improve the level of mean performance and reduce inequality (Dobbelsteen et al. 2002). Thus, from a theoretical point of view, the effects of tracking are widely controversial, suggesting a substantial uncertainty about its impact on both the level and distribution of students' achievement (see Betts, 2011; Meier & Schutz, 2007 for a review on the theoretical considerations on the impact of tracking). The empirical evidence on this matter is too, quite unclear. While the empirical literature tends to suggest that tracking aggravates inequality in achievement, the major issue with any empirical research on tracking is that other unmeasured factors bias the estimations of its impact. For example, studies that exploit changes in tracking policies in schools are potentially biased if other changes to schools are simultaneously made. Nonetheless, Hanushek and Woessmann (2006) analysis has given fairly robust evidence that early tracking increases inequality in achievement, without a clear impact on mean performance. To account for the unmeasured factors biasing the estimated impact of tracking, they apply an innovative international differences-in-differences approach, that compares average test scores and deviations from the mean at the country level, for different grades, before any country has introduced tracking and after tracking has been implemented in some countries. However, there is a type of tracking that has never been investigated in the literature before. This type of tracking is present at the upper secondary school level for students who are following an academic path and similarly to the other types of tracking, it is applied by some countries and absent in others. Countries differ in their upper secondary school systems in a way that some require their students to choose a specialization from a set of areas - typically natural sciences, economic sciences, humanities or arts - and follow that specialization for the course of their upper secondary education years (e.g. Portugal, Spain, Sweden) whereas by contrast, others including Finland, Denmark or the U.S. follow a general curriculum where students, albeit being able to choose between different classes in distinct areas, are not required to follow a single specialization and thus, receive a more general education. To the best of our knowledge, there has never been any paper specifically analyzing either theoretically or empirically the possible impacts of this specific type of tracking. As specialized systems can be considered somewhat less common – something which is supported by the proportion of countries with such system in our study sample – some additional information regarding its characteristics and anecdotal evidence about its possible effects on student outcomes is given.

One of the clearest differences from the ability grouping tracking, in which schools make strong recommendations for what type of track the student should follow based on his previous achievement, is that the choice of track placement in this case – the specialization – is solely the student's and parents' choice, at times aided by psychological advice or tests too. This decision is usually based on the academic interests of the student and his plans for his area of study at the tertiary education level, as some university programs often ask for the completion of some courses specific to a certain area (e.g. it is unlikely for a student following an arts specialization to study medicine at the tertiary level). Hence, this type of tracking does not divide students by ability but rather by area of interest and specialization and for this reason, the theoretical framework concerning the effects of having classrooms with homogenous or heterogeneous levels of student ability does not directly apply. Although the different specialization areas are not designed to have different levels of difficulty per se, as Bishop (2010) points out, the maths-science lines have a reputation for being the most difficult and prestigious, giving better chances of being accepted in varied tertiary education programs. A study done by the Portuguese Ministry of Education in 2014 shows that the biggest proportion of students with high grade point averages (classified as “Excellent” or “Merit”) is present in the maths- science track, but it is also the track that has the second highest proportion of “Fails” out of the four tracks (Cid et al., 2014) – possibly giving further confirmation that it is the specialization that attracts the best students but the one with the highest level of difficulty too. Although the degree of specialization varies across countries – as in some specialized systems, students are still able to choose quite freely from courses of other areas, while in others this choice is much more restricted – students will in any case have different focuses and thus, it is plausible to think that this will make them excel or fall behind in certain areas when compared to their peers of different specializations. One of the clearest differences between specialization areas is that the science and economics lines consistently have a higher focus on mathematics than the others, while to some lesser extent the humanities lines have a higher focus on languages, writing and reading skills. Therefore, it is possible that international tests show in some way an effect of this type of tracking, either through a change in the distribution of results of a country (e.g. an increase in educational inequality) or even a change in the level of mean performance. Given that the biggest difference between specialization areas seems to be in the extent of their mathematics curriculum, our initial intuition supposed that if any impact would be found, this would be more likely seen through an increase in the inequality of results for the mathematics exams for countries following a specialized system.

Since countries only follow one system or the other[1], this type of tracking requires a cross-country analysis to estimate the possible effects of these institutional differences. For this reason, an

[1] An exception was found for the case of Finland where, as noted by Kirjavainen (2007), around 13% of upper secondary schools offer the option of following a specialized curriculum. Nevertheless, most of them still offer the general track.

international differences-in-differences analysis is used to account for country heterogeneity and unobserved factors influencing student outcomes. The impact of having a specialized system in upper secondary school, as opposed to a general one, is identified by comparing the change in outcomes between a point just before students have started this stage of schooling – measured using PISA data - and a later point, after they have completed it – using data from the recently done PIAAC study – across countries with a specialized and general systems. Two approaches of this differences-in-differences methodology are followed: one, using country-level data, following the approach of Hanushek & Woessman (2006); the other, using student micro-level data, following the approaches of Waldinger (2007), Jakubowski (2010) and Ruhose & Schwerdt (2016).

The regression results of this first analysis suggest that the choice of one system or the other does not account for differences across countries in either the mean performance or the inequality of students' test scores.

This paper is organized as follows. Section 2 presents the econometric strategy followed and Section 3 describes the data used. Section 4 presents the regression results and Section 5 concludes.

2. Econometric Strategy

The common method to estimate the determinants of a student's achievement is to consider an education production function, where the achievement of a student - usually measured by test scores – depends on various factors: personal characteristics of the student such as gender, innate ability and previous educational trajectories; family characteristics that usually consider the highest level of education of a student's parents, income and physical resources available at home such as books, computers or the access to Internet; school characteristics such as its infrastructures, teacher's quality, class size, the curriculum taught, peer effects and even specific characteristics of a country's educational system.

Evidently, a student's achievement depends on an immense amount of factors - some of them, very hard to observe and measure. Still, one could consider a model like the one shown in equation (1) to estimate the impact of a county following a specialized system, as opposed to a general one, on a student's achievement:

$$A_{icg} = \alpha_c + \gamma SPEC_{igc} + X_{igc}\beta + \varepsilon_{igc} \qquad (1)$$

where the achievement of student i in grade g and country c (A_{icg}) is determined by a country specific intercept (α), several characteristics of the student, his family and school (vector X) and the existence of a specialized system in upper secondary school ($SPEC$). However, two main issues arise with using this methodology to estimate the impact of our variable of interest, $SPEC$. First of all, as Hanushek (2003) has shown, we cannot be confident with any estimates of the β, due to insufficient data on all the characteristics of a student's family background, innate ability, school characteristics and peer effects, not only at the contemporaneous level but also regarding their past influence. Not taking into account all of these hard-to-measure determinants to achievement will lead to the standard problem of omitted variables bias. Secondly, regarding the influence of having a specialized educational system in a cross-country analysis, if in a country every upper secondary school student in an academic track follows a specialization, $SPEC$ will be a country-fixed effect (a constant) and as such, we cannot estimate its impact on achievement.

Hanushek and Woessmann (2006) face the same problem when trying to estimate the impact of early tracking on achievement. To solve this issue, they effectively apply an international differences-in-differences approach, by comparing the average achievement gain in countries with early tracking to that of countries without it - as no country applies tracking in the early primary grades, it is possible to compare the level and distribution of student performance in these grades by using international assessment programs such as TIMSS and PIRLS, to these levels in secondary school where some countries have separated students into differing-ability schools and others have not, by looking at the results of PISA and TIMMS for secondary school. As such, their model regresses secondary-school outcomes on primary-school outcomes plus an indicator for the existence of tracking.

Given the similarity of their analysis to the one of this paper, the same methodology is followed in a first step. In this way, the level and distribution of performance of younger students, at grades before upper secondary school, is compared with those of older students, after they've been through a specialized or general educational system in upper secondary schooling. The impact of having a specialized system will then be estimated by:

$$\gamma = \overline{\overline{\Delta A}}_{spec} - \overline{\overline{\Delta A}}_{gen} + (\bar{\bar{v}}_{spec} - \bar{\bar{v}}_{gen}) \qquad (2)$$

where the double bar denotes the average achievement gain in each group of countries, the ones with a specialized educational system in upper secondary school (*spec*) and the ones with a general system (*gen*) and $\bar{\bar{v}}$ represent the expected composite errors. Thus, by taking a double-difference, a differences-in-differences approach is applied to estimate the effects of a specialized system. This identification strategy is therefore valid if the expected composite errors are uncorrelated with the existence of a specialized system. Conditions in which this could be violated are discussed further below.

The equation used in this country-level differences-in-differences approach analyzes both mean performance and the inequality of test scores, as measured by the within-country standard deviation and also the difference between different performance percentiles – the test score difference between the student performing at the 95th percentile and the student performing at the 5th percentile in each country and likewise for the 75th and 25th percentile. The identification will thus, follow the below form:

$$Y_c^2 = \beta_0 + \beta_1 Y_c^1 + \gamma SPEC_c + \varepsilon_c \quad (3)$$

where Y_c^2 represents the various ways of measuring students' performance and its distribution (mean performance and inequality) at the country-level at a point posterior to the conclusion of upper secondary education, which depends on a constant term (β_0); the same measure of students' performance at the country-level used for the dependent variable but at a point just before upper-secondary school (A_c^1); a dummy variable indicating whether students in a country follow a specialized system in upper secondary school ($SPEC_c$); and an error term (ε_c). Additional control variables such as the GDP per capita and a country's cumulative expenditure per student by age 15 years old are added as robustness checks.

As argued by Jakubowski (2010), there are two essential assumptions behind this country-level differences-in-differences. The first one is that the sample of students at t=1 and t=2 (i.e. before and after students have been through a specialized or general system) are fully comparable and representative of the population of students in each country. That is, students sampled in t=1 have on average similar characteristics to those sampled in t=2 in each of the countries considered in the analysis. If this is the case, then any unobserved time-invariant characteristics of the sample of students in each country is cancelled out. The second assumption is that there are "same time effects", also known as the "parallel-trends" assumption of differences-in-differences models: achievement would change by the same magnitude in both types of countries with the two different systems if a specialization tracking would not take place at all.

Still, there are two key limitations to this country-level differences-in-differences approach. First of all, since data is aggregated at the country level, this leads to extremely small sample sizes of around 20 countries. Furthermore, for this same reason, the number of additional control variables that can be added in the regressions is importantly limited, as they have to be considered at country-level.

The second methodological line followed in this paper is to consider the same differences-in-differences model but with student micro-level data, following the approaches of Waldinger (2007), Jakubowski (2010) and Ruhose & Schwerdt (2016). The model used to estimate the effect of a specialized system on students' achievement is then the following:

$$Y_{ict} = \beta_0 + X'_{ict}.\beta_1 + \beta_2 UPSEC_t + \beta_3 SPEC_c + \gamma UPSEC_t * \ SPEC_c + \epsilon_{ict} \qquad (4)$$

where Y_{ict} is the standardized test score of student *i*, in country *c* and time *t*; X_{ict} is a vector of student background variables that includes the number of books at home, the highest level of parental education, the migration status and whether the student speaks the language of the test at home; $UPSEC_t$ is a dummy variable that takes the value 1 if an observation is taken from a student who has finished upper secondary school and the value 0 if from a student in an educational stage just before upper-secondary schooling; $SPEC_c$ is a dummy variable that identifies whether an observation comes from a country with a specialized system in upper-secondary school and ϵ_{ict} is the error term. The key parameter of interest is therefore γ, which under the identifying assumptions of this differences-in-differences methodology discussed above, estimates the effect of a specialized system, in contrast with a general system, on a student's achievement. It is worthwhile noting that since we do not observe the same individuals in the two time periods (before and after upper-secondary schooling), a further identifying assumption is that the stratified sampling nature of the surveys used in this analysis guarantees that a representative sample is tested.

A final approach is to run Quantile regressions with these micro-data. Rather than estimating the conditional mean of the dependent variable, the objective is to identify if there are different effects on specific quantiles, by estimating the conditional quantile of the dependent variable. From a policy perspective this seems to be a relevant question, as it could be that a Specialized system does not have an effect on average but does have an effect for lower or higher performing students.

3. Data

Since around the late 1950's international testing of students began to be undertaken with the objective of comparing the performance of students across countries using the same evaluation criteria. Although these early studies faced some problems of sampling and within-country selectivity, since then, they have improved substantially and are currently widely used by researchers to evaluate educational matters. One of the most famous and used studies (if perhaps not the most) is the triennial PISA test by the OECD that started in 2000, with the intent of improving education policies and outcomes. This assessment program tested around 510,000 students in the last round of 2012 over 65 economies, in the key subjects of mathematics, reading and science. Students who take part in this exam are 15 years old of age (or have turned 16 very recently), an age which is chosen due to the fact in many countries, this coincides with the end of compulsory education – and for the interest of this paper, with a point just before the entrance to upper secondary schooling too.

To asses whether specialization in upper secondary school has an impact on student's performance, we would additionally ideally use a test such as PISA, but that instead, evaluates students just after they have concluded this last stage of secondary education. In 2012, the OECD released the results of the first edition of PIAAC, a study that aims at evaluating adults from 16 to 65 years old in the areas of literacy, numeracy and problem-solving. Around 155,000 individuals were tested, over 24 countries in this first round of the study. By restricting the sample of this study to individuals who are 18 or 19 years old and have recently finished academic upper-secondary schooling, it is then possible to get a sample of interest for the analysis proposed.

The countries that are present on both PISA and PIAAC studies are the following: Australia, Austria, Belgium (Flanders only), Canada, Czech Republic, Denmark, Estonia, Finland, France, Germany, Ireland, Italy, Japan, Korea, Netherlands, Norway, Poland, Russian Federation, Slovak Republic, Spain, Sweden, United Kingdom (England and Northern Ireland only) and United States. The subjects that are common to both tests are the reading and mathematics parts of the exams.

As the methodology of this paper proposes to compare the performance of students and its distribution at a point before and after they've been through an academically- oriented education in upper secondary school (following a specialized or general system), several removals had to be done in the original PIAAC database and to a lesser extent, in the PISA database as well. Starting with PIAAC, this database has information on the highest level of education obtained by each examinee according to the ISCED 97 classification (UNESCO, 1997) – an international classification of different levels of education created by UNESCO to facilitate comparisons of education statistics and indicators across countries, given their wide variety in terms of structure and curricular content. This not only tells us whether an individual has completed at least upper secondary education - the basic condition to be eligible for this analysis - but also if this education had an academic or vocational orientation. Once again, as we are only interested in analyzing students that followed an academic education in upper secondary school (since they are the only ones subject to either a specialized or general system), students with a vocational education at this level were removed from the study. These individuals that were removed are classified as having a highest level of education ISCED 3C, a decision which is complemented with a section of the PIAAC database that has information specifically on whether this level of highest education was vocationally oriented. Some individuals did not report their highest level of education according to the ISCDED classification and for those from the countries of Canada and Estonia, this information is not available at all – therefore, they were also removed. Australia was also not considered in the study as the data for this country had to be paid for. Furthermore, PIAAC has information on the area of study for the highest level of education obtained – with this, individuals who followed an area which was clearly not academic, were removed too (e.g. agriculture, social services, manufacturing).

Concerning the age group of individuals in the PIAAC database, we chose to include only individuals with a maximum age of 19 years old to minimize the chances of having other confounding factors such as the quality of post-secondary and tertiary education influencing a student's achievement. Still, it would be relevant to see, in case any effects are found from the two different systems, if they would remain in an older population. In any case, extending the age group in the PIAAC sample did not change our main results.

Concerning the PISA database, removals were made mostly in countries that have tracking between academic and vocational schools before or at the age when students take this exam. Once again, since our methodology proposes to compare individuals that have been through an academic upper secondary education following either a specialized or general system with their equivalents at a point before this stage of education, students who are already following a vocational track at the age of 15 are not eligible for this analysis. In this way, in countries with no early tracking between vocational and academic tracks such as Norway, Finland or United States no removals needed to be made, whereas in countries with early tracking such as Austria or Germany, several removals were done. These removals were then made for students in a vocational or pre-vocational program (ISCED 3 or

ISCED 2) and as in the PIAAC database, for students with missing information on the type of programme they were attending.

Regarding the PISA cohorts used to run the differences-in-differences estimation, two possibilities exist: we can follow the PISA 2009 cohort, as 15 year old students were approximately 18 or 19 years old in PIAAC 2012, or focus on contemporaneous measures of performance at the two different educational stages by using the PISA 2012 and PIAAC 2012. According to Hanushek and Woessman (2006), the former possibility has the advantage of minimizing the possibilities of having confounding factors from any inherent variation in family background and peer characteristics, while the latter has the benefit of assessing the two different cohorts in a stable educational system. As a mean to provide a further robustness check to our results, we test with the two combinations in the country-level regressions and follow one cohort in the micro-level regressions.

The control-variables used in these micro-level regressions have to be available in both surveys. We chose to include the number of books at home, the highest level of parental education, the migrant status and a variable indicating whether the individual speaks the language of the test at home. In the PIAAC survey, participants are asked how many books at home when they were 16 years old, making this variable comparable with the information provided in PISA.

Finally, regarding the information about whether a country follows a specialized or general system in upper secondary school, this data was gathered from different sources ranging from OECD documents, information available on the websites of the Ministry of Education of each country and other articles about the curriculum and educational system of a country. Although it was not possible to have the exact information on when such systems were introduced in each country, we are confident that they have remained the same way since at least 2008 (the period when the cohorts considered are expected to have started upper secondary education) and assume there have been no changes since then, which we were able to verify for the vast majority of countries in our sample. In the group of 20 countries analyzed in this study, 14 of them follow a general system and 6 a specialized one. The general system group is constituted by Austria, Belgium, Czech Republic, Denmark, Finland, Germany, Ireland, Japan, Korea, Poland, Russian Federation[2] Slovak Republic,

---

[2] OECD considers the results from the Russian Federation in the PIAAC survey to be preliminary. Furthermore, they do not include the region of Moscow which is not possible to remove independently from the PISA study. Although we decide to keep this country in our regression analysis, removing it from the sample did not change the results.

United Kingdom and United States. The specialized system group is formed by France, Italy, Norway, Netherlands, Spain and Sweden.

Test scores for each individual in both PISA and PIAAC are estimated with plausible values based on a multiple imputation technique (see OECD, 2009; and Pokropek & Jakubowski 2012, 2012 for more on how plausible values and skills are estimated for PISA and PIAAC respectively). As PISA and PIAAC have different scales (PISA from 200- 800 and PIAAC from 0-500), the variables analyzed in the next section are standardized to have a cross-country mean of zero and a standard deviation of 1.

4. Results

4.1. Country-level regressions

The country-level regression analysis applies a differences-in-differences estimation, as described in section 2 of this paper, to estimate the impact of a country following a specialized system in upper secondary school. The regressions following equation (3) consider two variables as dependent: on the one hand, the country-level mean performance in the subjects of mathematics and reading shown in PIAAC; on the other hand, inequality in those subjects. The analysis is developed separately for the PISA 2009 – PIAAC 2012 and PISA 2012-PIAAC 2012 combinations.

Table 1 and Table 2 report the regression results for Mathematics and Reading achievement, respectively, in a country-level differences-in-differences estimation. Concerning our variable of interest, the estimates of the dummy variable indicating whether a country has a specialized system in upper-secondary school are consistently non-significant across measures of achievement (mean performance and inequality), in the two subjects of mathematics and reading, and in combinations of PISA2009-PIAAC2012 and PISA2012-PIAAC2012. Experimenting with different measures of inequality such as the test-score difference between the 75th and the 25th performance percentile or the difference between the 95th and 5th percentile does not change the results. Moreover, adding further control variables such as the GDP per capita (in purchasing power parities, reported by The World Bank) and/or a country's cumulative educational expenditure per student by age 15 (reported by OECD, 2012) does not change the statistical significance of the specialized system variable either.

Regarding how mean achievement results in PIAAC relate to those of PISA, the point estimates of around 0.5 imply that countries tend to reduce their mean performance and inequality from PISA to

PIAAC. This result is consistent across test-pairs with statistically significant estimations in both reading and mathematics mean performance achievement. However, in none of the regressions that consider measures of inequality in performance in PIAAC as the dependent variable, are these related to the inequality of performance existent in PISA. Jakubowksi (2010) had already obtained this result when analysing the dispersion of results in PISA and PIRLS or TIMSS (4th grade students) in this same country-level differences-in-differences setting, arguing that it makes this country-level methodology questionable to analyse the inequality of results in different surveys.

Table 1

Mathematics Achievement

Survey combinations	**PISA 09 - PIAAC 12**		**PISA 12 - PIAAC 12**	
Measure of achievment	Mean	Standard Deviation	Mean	Standard Deviation
Achievement in PISA	0.612**	-0.159	0.510**	- 0.161
	(0.167)	(0.207)	(0.215)	(0.239)
Specialized system	0.139	- 0.312	0.276	- 0.338
	(0.448)	(0.399)	(0.458)	(0.402)
Constant	- 0.042	0.094	- 0.082	0.101
	(0.211)	(0.309)	(0.247)	(0.309)
N	20			
R^2	0.373	0.043	0.249	0.043

Dependent variables: Country-level mean performance and inequality (measured by the standard deviation of test scores) in Mathematics achievement. Number of Countries: 20. Robust standard errors in parentheses. Significance levels: ***1%, **5%, *10%

Table 2

Reading Achievement

Survey combinations	PISA 09 - PIAAC 12		PISA 12 - PIAAC 12	
Measure of achievment	Mean	Standard Deviation	Mean	Standard Deviation
Achievement in PISA	0.455**	0.128	0.434**	0.134
	(0.196)	(0.200)	(0.184)	(0.171)
Specialized system	- 0.169	0.016	- 0.025	- 0.070
	(0.519)	(0.391)	(0.555)	(0.463)
Constant	0.051	0.005	0.007	0.021
	(0.215)	(0.310)	(0.221)	(0.324)
N		20		
R^2	0.209	0.016	0.190	0.017

Dependent variables: Country-level mean performance and inequality (measured by the standard deviation of test scores) in Reading achievement. Number of Countries: 20. Robust standard errors in parentheses. Significance levels: ***1%, **5%, *10%

Table 4 reports the regression results of the micro-level regressions for Mathematics and Reading achievement, following the specification of equation (4) in Section 2. These regressions consider the PISA 2009 – PIAAC 2012 cohorts. Columns 1 and 3 report the regression results with robust standard errors, whereas columns 2 and 4 report the regression results with standard errors clustered at country level. This latter approach is the one followed by Waldinger (2007) and Ruhose & Schwerdt (2016) and allows for arbitrary correlations of the error term for students within one country. The other possibility in this type of analysis is to cluster the standard errors at the school level, which in an analogous way, allows for arbitrary correlations of the error term for students within one school and typically leads to much smaller standard errors, as noted by Waldinger (2007). However, as the PIAAC survey does not include information about schools, it's not possible to follow this approach in our analysis.

Still, the statistical significance of the results for the variable of interest – the interaction between a specialized system and for results taken after upper secondary school (i.e. from PIAAC) – in Mathematics achievement change considerably depending on whether we consider robust standard errors or standard errors clustered at the country level, as the estimate is no longer statistically significant with the latter approach. These estimates suggest that there is a positive effect associated with a specialized system for Mathematics mean results of 0.148 standard deviations, however not statistical significant when standard errors are clustered at the country-level. Estimates for the variable of interest in the mean performance of Reading test scores are never statistically significant, regardless of the approach chosen to estimate the standard errors.

Regarding the other variables typically included in a differences-in-differences estimation, a variable for the post-treatment time – the Upper Secondary variable – and a variable for the treatment group – the Specialized system variable – these show statistical significant impacts with robust standard error estimates. As for the Upper Secondary variable, which indicates that an observation is taken from a student who has recently completed academic upper secondary school (i.e. from the PIAAC survey), the estimates confirm what was seen in the country-level analysis that mean results in PIAAC tend to be lower than in PISA, regardless of the different upper-secondary school systems. Considering the Specialized system variable, which indicates that an observation is taken from a student in a country with a specialized upper-secondary school system, the estimates suggest that countries with such system show lower results in both Mathematics and Reading, even before students go through upper-secondary school.

Several control variables are included to ensure that the PIAAC and PISA samples are comparable in terms of individual and family characteristics. These include the student's gender, the number of books at home, the highest level of parental education, whether the student is a first or second generation immigrant, whether the student migrated to the country of the test after age 10 and whether the students speaks the language of the test at home. The number of books variables are coded as groups of dummy variables with a reference category of 0-10 books; the highest level of parental education has a reference category students whose parents completed at most lower-secondary school; the immigration status (2^{nd} and 1^{st} generation immigrants) has a reference category native students; and the Late immigrant variable indicates that the student migrated to the test country after age 10. These control variables have their expected sign and are almost all statistical significant: boys perform better in Mathematics, whereas girls perform better in reading; students with a higher level of parental background (as measured by the number of books at home or the highest level of parental education) perform better and students with migrant status or that speak a different language at home perform lower.

Table 3
Micro-data differences-in-differences estimation: PISA09-PIAAC12

	(1)	(2)	(3)	(4)
	Reading		Mathematics	
Specialized system	- 0.035***	0.035	- 0.090***	- 0.090
	(0.005)	(0.133)	(0.006)	(0.111)
Upper Secondary	- 0.046*	- 0.046	-0.085***	-0.085
	(0.027)	(0.084)	(0.028)	(0.082)
Specialized system x Upper Secondary	-0.008	- 0.008	0.148**	0.148
	(0.058)	(0.189)	(0.060)	(0.163)
Further Controls				
Male	- 0.336***	- 0.336***	0.186***	0.186***
	(0.005)	(0.019)	(0.006)	(0.024)
11-25 books	0.297***	0.297***	0.282***	0.282***
	(0.012)	(0.031)	(0.012)	(0.026)
26-100 books	0.628***	0.628***	0.623***	0.623***
	(0.010)	(0.036)	(0.012)	(0.044)
101-200 books	0.882***	0.882***	0.880***	0.880***
	(0.011)	(0.041)	(0.012)	(0.055)
201-500 books	1.099***	1.099***	1.115***	1.115***
	(0.011)	(0.046)	(0.011)	(0.055)
> 500 books	1.120***	1.120***	1.145***	1.145***
	(0.013)	(0.054)	(0.013)	(0.060)
Highest parental education:	0.210***	0.210***	0.184***	0.184***
Upper Secondary	(0.009)	(0.039)	(0.009)	(0.037)
Highest parental education:	0.307***	0.307***	0.291***	0.291***
Tertiary	(0.009)	(0.048)	(0.009)	(0.049)
2nd generation immigrant	- 0.220***	- 0.220***	-0.249***	-0.249***
	(0.017)	(0.069)	(0.015)	(0.055)
1st generation immigrant	- 0.270***	- 0.270***	- 0.297***	- 0.297***
	(0.017)	(0.046)	(0.017)	(0.039)
Different language	- 0.288**	- 0.288**	-0.253**	-0.253
	(0.117)	(0.017)	0.123)	(0.162)
Late immigrant	- 0.214***	- 0.214***	- 0.179***	- 0.179***
	(0.025)	(0.046)	(0.029)	(0.041)
Constant	-0.703***	-0.703***	-0.914***	-0.914***
	(0.013)	(0.069)	(0.012)	(0.074)
R^2	0.201		0.182	
Individual Observations	115105			
Number of countries	20			

*Dependent variables: Dependent variables are the original test scores which are standardized within each country to have and international mean 0 and standard deviation of 1. Odd-numbered columns report Robust Standard errors while even-numbered columns report standard errors clustered at the country level in parentheses. Number of countries with a Specialized system: 6. Significance levels: ***1%, **5%, *1%*

The final approach taken, as described in section 2, is to run Quantile regressions with these micro-data. Rather than estimating the conditional mean of the dependent variable, the objective is to identify if there are different effects on specific quantiles, by estimating the conditional quantile of the dependent variable. It could be that a Specialized system does not have an effect on average but does have an effect for lower or higher performing students. Analogously to the results presented in Table 3, the even-numbered columns of Table 4 present estimates with standard errors clustered at the country level, for the variable of interest. Results are only presents for Mathematics achievement, as results were never statistically significant for Reading achievement. The regression results suggest that the strongest effects associated with a Specialized system are in the lower quantiles (quantiles 0.1 and 0.25) and the smallest effects are seen in the median. Once again, however, these results are not statistically significant when standard errors are clustered at the country-level.

Table 4
Quantile Regressions for Mathematics achievement

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
0.1		0.25		0.5		0.75		0.9	
0.187*	0.187	0.189*	0.189	0.138	0.138	0.148*	0.148	0.148	0.148
(0.106)	(0.219)	(0.100)	0.164)	(0.094)	(0.178)	(0.083)	(0.146)	(0.106)	(0.154)

All entries show the coefficient on the interaction between Upper Secondary x Specialized system, the variable of interest. Odd-numbered columns report Robust Standard errors while even-numbered columns report standard errors clustered at the country level in parentheses. Number of observations: 115105. Number of countries: 20. Number of countries with a Specialized system: 6. Significance levels: ***1%, **5%, *1%

5. Conclusion

The analysis carried out in this paper provides preliminary results about the effects of a specialized versus a general upper secondary school curriculum on students' performance and inequality which, to the best of our knowledge, is the first analysis ever done regarding this topic.

Because countries only follow one system or the other, a cross-country analysis is required to estimate the possible effects of these institutional differences. We choose an international differences-in-differences analysis to account for country heterogeneity and unobserved factors influencing student outcomes. The impact of having a specialized system in upper secondary school, as opposed to a general one, is identified by comparing the change in outcomes between a point just before students have started this stage of schooling – measured using PISA data - and a later point after they have completed it – using data from the recently done PIAAC study – across countries with a specialized and general systems. In the country-level regressions following a differences-in-differences methodology, these changes are analyzed in terms of mean performance levels and different measures of inequality for the subjects of mathematics and reading and for the two combinations of PISA 2009-PIAAC 2012 and PISA 2012–PIAAC 2012. Although mean performance in PIAAC is statistically significantly related to that of PISA in both test-pairs and subjects, the regression analysis suggests that the choice of one system or the other does not account for differences across countries in either the mean performance or the inequality of students' test scores.

Concerning the micro-level regressions following a differences-in-differences methodology, a positive effect for Mathematics achievement is identified for students following a specialized system in academic Upper-Secondary school. However, these effects are no longer statistically significant when standard errors are clustered at the country-level, which allows for arbitrary correlations of the error term for students within one country. This seems to be the correct approach to be taken here and the one followed by previous research that analysed the impact of tacking in a differences-in-differences setting. Moreover, when estimating whether there are different effects for different Mathematics performance quantiles, through Quantile regressions, the positive and largest effects for the lower quantiles lose their statistical significance when clustering standard errors at the country level.

As this approach of clustering standard errors at the country-level greatly changes the relevance of the results, we believe that it is too early to conclude whether a specialized upper-secondary academic schools has any effects on student achievement and that further research in this topic is needed.

In the future, we propose to extend this analysis by implementing this same methodology to the TIMSS 1995 survey, which tested students in their 4th, 8th and 12th grade. Besides providing a robustness check to these results, it brings the possibility of adding students' school characteristics and clustering standard errors at the school-level.

References:

Benabou, Roland, 1996. "Equity and Efficiency in Human Capital Investment: The Local Connection," Review of Economic Studies, Wiley Blackwell, vol. 63(2), pages 237-64, April.

Betts, Julian R., (2011), "The Economics of Tracking in Education", in Hanushek, Eric A., Stephen Machin and Ludger Woessmann (Eds.), Handbook of the Economics of Education, Volume 3, Amsterdam: North Holland, pp. 341-381

Bishop, J. H. (2010). "Which secondary education systems work best? United States or Northern Europe" Retrieved from Cornell University Working papers: www.digitalcommons.ilr.cornell.edu/workingpapers/

Brunello, Giorgi & Checchi, Danielle (2007). "Does school tracking affect equality of opportunity? New international evidence," Economic Policy, CEPR;CES;MSH, vol. 22, pages 781-861, October.

Cid, Marília, et al., 2014. Portuguese Ministry of Education "Os cursos científico-humanísticos e o alargamento da escolaridade obrigatória – medidas educativas de inclusão"

Dobbelsteen, Simone, Jesse Levin, Hessel Oosterbeek (2002). "The Causal Effect of Class Size on Scholastic Achievement: Distinguishing the Pure Class Size Effect from the Effect of Changes in Class Composition". Oxford Bulletin of Economics and Statistics 64 (1): 17-38.

Hanushek, Eric A. (2003). "The Failure of Input-Based Schooling Policies". Economic Journal 113 (485): F64-F98.

Hanushek, Eric A. & Ludger Wössmann, (2006). "Does Educational Tracking Affect Performance and Inequality? Differences- in-Differences Evidence Across Countries," Economic Journal, Royal Economic Society, vol. 116(510), pages C63-C76, 03.

Heath, Anthony, ed. (1984). Comprehensive and Selective Schooling. Special Issue of the Oxford Review of Education 10 (1): 6-123

Jakubowski, Maciej. "Institutional tracking and achievement growth: Exploring difference-in-differences approach to PIRLS, TIMSS, and PISA data." *Quality and inequality of education*. Springer Netherlands, 2010. 41-81.

Kirjavainen, Tanja (2007). "Efficiency of Finnish Upper Secondary Schools: An Application of Stochastic

Frontier Analysis with Panel Data," Discussion Papers 428, Government Institute for Economic Research Finland (VATT)

Maaz, K., Trautwein, U., Lüdtke, O. and Baumert, J. (2008), "Educational Transitions and Differential Learning Environments: How Explicit Between-School Tracking Contributes to Social Inequality in Educational Outcomes. Child Development Perspectives", 2: 99–106

Meier, Volker & Schütz, Gabriela (2007). "The Economics of Tracking and Non-Tracking," Ifo Working Paper Series Ifo Working Papers No. 50, Ifo Institute for Economic Research at the University of Munich.

OECD (2009). "PISA data analysis manual: SPSS second edition" OECD (2012). "PISA 2012 results: What makes schools successful? Resources, policies and politics"

OECD (2012). "PISA 2012 results: What makes schools successful? Resources, policies and politics"

OECD (2013). "The Survey of Adult Skills: Reader's Companion"

Pekkarinen, Tuomas & Uusitalo, Roope & Kerr, Sari, (2009). "School tracking and intergenerational income mobility: Evidence from the Finnish comprehensive school reform," Journal of Public Economics, Elsevier, vol. 93(7-8), pages 965-973, August.

Pokropek, Artur & Maciej Jakubowski, (2013). "PIAACTOOLS: Stata module to provide PIAAC tools," Statistical Software Components S457728, Boston College Department of Economics, revised 31 Dec 2013.

Ruhose, Jens, and Guido Schwerdt. "Does early educational tracking increase migrant-native achievement gaps? Differences-in-differences evidence across countries." *Economics of Education Review* 52 (2016): 134-154.

UNESCO (1997). "International Standard Classification of Education – ISCED 97"

Waldinger, Fabian. "Does ability tracking exacerbate the role of family background for students' test scores." *Unpublished manuscript, London School of Economics, London, England* (2007).

Woessmann, Ludger (2009). "International Evidence on School Tracking: A Review," CESifo DICE Report, Ifo Institute for Economic Research at the University of Munich, vol. 7(1), pages 26-34, 04.