Working Paper

ADAPT

Associazione per gli Studi Internazionali e Comparati sul Diritto del lavoro e sulle Relazioni industriali

In collaborazione con il Centro Studi Internazionali e Comparati Marco Biagi

The school performance of immigrant students. Cross-country evidence from PISA 2006

by Marina Murat, Giulia Pirani***

Introduction

It is generally agreed that education implies individual social promotion, higher productivity and economic growth. Countries, however, differ in their education histories, institutions and also, as shown by the empirical evidence, results.

During the last decade, some international appraisals, especially the Programme for International Student Assessment (PISA), have provided data that have helped to shed light on the performance of students at school in several countries and on the factors more frequently related to it. Among these factors, family background, gender and nationality have shown to be especially important. While several studies have investigated on the influence of the two first two, only a few have focused on the impact of nationality, despite the raw data show that, except for few countries, immigrant students tend to perform below natives (Entorf and Minoiu, 2004; Entorf and Lauk 2006; Snepf 2007; OECD, 2006).

This paper analyses the distribution across countries of the performance of immigrant students at school using data from PISA 2006. To this purpose, we run a first set of estimates, one for each country, using the condition of being native or immigrant as the only regressor. This gives an initial picture of the immigrants-natives performance gaps across countries. Subsequently, we run more complete regressions that include a long list of variables related to students' characteristics and family backgrounds. This allows us to check for the results of our variables of interest once the other factors have been accounted for and, also, to search for possible common patterns regarding the immigrants' performance across countries. We find a geographical pattern, with central Western Europe as the world area with the more pronounced negative gaps.

We then add structural characteristics of the school systems into the analysis. We distinguish between countries where the education programs differ between schools, or school "tracks", and countries with "comprehensive" schools. We look at the institutional characteristics of each country and, for those with the tracking system, we split schools in at most three types, representing, respectively, schools preparing students mainly for university studies, intermediate schools, and schools leading just to the labor market. We then introduce into the regressions a variable indicating the school "type" each student attends.

We find that the cross-country pattern of gaps between immigrant and native students is not

Registrazione n. 1609, 11 novembre 2001 – Tribunale di Modena

independent from the structural characteristics of the countries' schooling systems. Education is more deeply based on the tracking schools system in countries of central Western Europe, where the performance gaps of immigrants with respect to natives are higher, while differences tend to be lower, or non-significant, in other regions of the world and in countries with comprehensive schools. In terms of educational policies, this implies that educational systems more based on comprehensive schools may increase the performance, and, hence, future social mobility, of immigrant students. This, in turn, may have positive effects on the society and the economy.
The paper is structured as follows. Section 1 explains the estimation methods used in the paper, Section 2 presents the data and some descriptive statistics, Section 3 discusses some main results and Section 4 concludes.

1. Estimation Method

We estimate three different specifications of an educational production function (EPF) in order to assess the determinants of differences in natives and immigrants' performances at school. In each specification the dependent variable is represented by the average score of each student obtained as the arithmetic mean between the student's score in each field ([1]).
In the first place, we aim at investigating the pattern of gaps in performance between immigrants and natives between countries. To this purpose, we run a first set of univariate estimates, one for each country, with the student's condition of being native or immigrant as the only regressor. The linear relation can be described as:

$$Y_i = \beta_0 + \beta_I I_i + \varepsilon_i \qquad i = 1,....,n \qquad (1)$$

where Y_i is the response variable representing the average score obtained by student i, β_0 is the intercept of the regression line, I_i is the explanatory variable for student i, β_I is the coefficient of independent variable and ε_i represents the error term, $\varepsilon_i \sim N(0 ; \sigma^2)$.
Subsequently, we run more complete regressions, which include a long list of variables that relate to students' characteristics and family backgrounds (see Appendix). The multivariate model can be described as:

$$Y_i = \beta_0 + \beta_I I_i + \beta_L L_i + \beta_X X_i + \varepsilon_i \qquad i = 1,....,n \quad (2)$$

where L_i is a variable witch indicates the language spoken at home by the student i, β_L is the coefficient of the language variable, X_i is the vector of regressors added into the model for student i and β_X is the vector of coefficients on regressors.
We select the variables to be used in order to optimize the trade-off between non-complexity and the explaining-ability of the model by using the Bayesian Information criterion (BIC):

$$BIC = -2 \ln(L) + k \ln(n) \qquad (4)$$

where L is the maximized value of the likelihood function for the estimated model and k and n are, respectively, the number of regressors and observations. As the maximum likelihood estimators correspond to OLS under the assumption of normally distributed experimental errors, BIC can be also written as:

$$BIC = RSS / \sigma^2 + k \ln(n) \qquad (5)$$

where RSS is the residual sum of squares from the estimated model and σ^2 is the error variance. The BIC is an increasing function of RSS and k. Thus, lower BIC implies either fewer explanatory variables, better fit, or both: given any two estimated models, the model with the lower value of BIC is the one to be preferred. As mode of stepwise search we apply the forward selection ([2]). Hence, finally, we run an OLS regression.

2. Data and descriptive statistics

The Programme for International Student Assessment (PISA) is an every-three-year internationally standardised assessment promoted by OCSE since 2000. Its main purpose is to collect data on the 15-year-old students' ([3]) competencies in reading, mathematics and science, that can be used to compare, both within and between countries. In each PISA survey, one subject has been chosen as a focus while the other domains have been assessed more briefly. In this paper we use the third wave of PISA, including 30 OECD jurisdictions and 27 non-OECD jurisdictions,

which refers to data collected in spring 2006 and whose main focus is on science ([4]).
Countries included in this study fulfil the following conditions: immigrant students are at least 3% of the students' population and their total number is above 100 ([5]). Twenty-eight countries satisfy these conditions. Table 1 (see Appendix) depicts the list of these countries and shows the share of first and second generations immigrant students. As there are no data in PISA 2006 on the reading competencies of students in the USA, and we use the average scores resulting from reading, mathematics and science, we had to exclude this country ([6]).
The Table shows that the countries with the highest shares of immigrant students are scattered in different areas of the world: Asia, Oceania, North America, Europe.

3. Results

3.1. The immigrant status

To check for the performance of immigrant students in each country, we run a first set of regressions with the average scores as the dependent variable and the immigrant or the native status of students as the only regressor. Non-native students are split into first ([7]) and second ([8]) - generation immigrants. The first columns of Table 2 (see Appendix), labelled *immigr*, depict the coefficients of the immigrant variables in each country. The intercept includes the performance of natives and the coefficients of the regressors indicate deviations from the natives' average score, or intercept. Here and in the remaining columns of the Table we include only the significant coefficients of the variables of interest ([9]).
The general picture emerging from these first regressions is a wide disparity of performances of immigrant students across countries. However, as can be easily checked by comparing these results with the data of Table 1, the cross-country distribution of immigrant students results does not seem to relate to their absolute number or to their relative presence in countries.
Rather, the cross-country immigrants-natives gaps do seem to display a geographical pattern. To check this, we have ranked countries and grouped them into four categories, according to the to the values of coefficients. The results of this classification are depicted in Figure 1 (see Appendix). Countries with the highest absolute values of negative coefficients are labelled as type A. As it can be observed, these countries are all located in central Western Europe and are contiguous.
The coefficients of countries of type B are still negative but their absolute values are lower. Some of these countries, as Sweden, could easily be shifted to Group A, but this country's coefficients of second generation immigrants are significantly lower in absolute value than those of first generation ones, an this makes it different from the generality of countries of type A. All countries of type B are also European and, with the exception of Estonia, are located in Western Europe. It may also be observed that, geographically, they are scattered around the area covered by countries of type A.
Countries of type C are instead scattered worldwide. Two of them, Great Britain and Greece, are in Western Europe, three, Latvia, Russia and Slovenia, are in Eastern Europe, and two, New Zealand and Hong Kong, are respectively in Oceania and East Asia.
Immigrant students in countries of type D, the last in the ordering, perform at the same level or better than natives. These countries are even more dispersed worldwide: while Ireland and Montenegro are in Europe, Canada, Australia, Israel, Macao, Qatar, are located in different continents.
Table 2 depicts separate coefficients for first and second generation immigrants. It is often presumed that, everything else equal, the performance of second generation immigrants should be more similar to that of natives than that of first-generation ones, both in countries where gaps are positive and were they are negative. This is because the relevant characteristics of immigrant students born into the country are expected to be nearer to those of natives than those of students born abroad. For several of the countries under investigation, however, the "everything else equal" condition may not apply. Since the fall of the Berlin Wall, the composition of the immigrant population has significantly changed, especially in Western Europe, were the presence of people originating from Eastern Europe has steadily grown. Also, during the last two decades, several

countries have modified their immigration policies in the direction of increasingly favouring the inflows of skilled immigrants. These factors, together with exogenous modifications in the composition of migrant flows from other areas of the world, may have contributed to modify the characteristics of immigrants' cohorts and, presumably, the performance of immigrant students at school. For example, in Table 2, three countries of group A, Austria, Germany and Netherlands, do not meet the above expectation of second generation immigrants performing better than first generation ones, but also in the other countries of Table 2 there could be important differences between first and second generation cohorts of immigrants.

The splitting of immigrant students into first and second generation is useful, however, because it allows a more disaggregated view of the patterns of performance across countries. For example, as seen above, Sweden differs from countries of Group A in the performance of second generation immigrants. More generally, the first columns of Table 2 show that while all countries of Group A and B are characterized by significantly negative performances of both first and second generation immigrants, this is not so in countries of Groups C and D, where, with the exception of Great Britain, Russia and Slovenia, the performance of at least one generation is characterized by a non negative, less significant or non-significant coefficient.

To control whether this geographical distribution of students' performance is robust to the inclusion of a more complete list of variables and, more generally, to check for the possible existence of other common patterns in the immigrant students performance, we add to the estimation equation a wider set of regressors, which are taken from the PISA dataset and are related to the students' main characteristics and backgrounds.

3.2. Full regressions. Students' characteristics, family backgrounds, language at home

This new set of regressions are now run on a large number of independent variables, indicated by L_i and X_i and in equation (2) above. The complete list of the latter can be seen in the Appendix; they are about forty, including gender, another (national) language spoken at home, number of books at home, other possessions at home, level of education of each parent, occupation of each parent, number of hours spent at school and number of hours spent at home studying reading, science and mathematics, interest in studying science (the main subject of PISA 2006) and in issues concerning the environment. Particularly important for this study is a variable indicating that a non-national language is spoken at home: more than 50% of immigrant students speak a foreign language at home in Austria, Italy, Norway, Sweden, Luxembourg, Switzerland; the percentage is below this rate but above 40% in Slovenia, Denmark, Germany, Canada and Netherland, and above 30% in Israel, New Zealand, Great Britain, Australia, Belgium, France and Ireland.

As said above, we use the BIC method to select variables and, as a consequence, can have different regressors in different countries. The second column of each country in Table 2, named *full regr*, depict only the significant coefficients of our variable of interest, immigration, and of the non-national language spoken at home variable, *Lang.home other*.

The general picture that emerges from the full regressions of Table 2 mostly confirms the geographical distribution of countries' coefficients seen above. At the same time, the coefficients of the control variables (not shown into the Table) do not seem to evidence the existence of common patterns across countries that could be related to the performance of immigrant students. Partly, their values confirm the results of the previous literature: the number of books at home, the parents' level of education and type of occupation, as well as other variables as, for example, the hours of lessons taken at school regarding science, mathematics and reading, are often significantly correlated with the performance of students ([10]).

As could be easily predicted, the immigrant coefficients of Table 2 tend to converge, from positive and negative values, to the natives' averages. However, countries of type A still have higher negative coefficients than those of type B, which in turn are more negative or less positive than those of Groups C and D ([11]). Two partial exceptions appear to be Germany, in Group A, and Italy, in Group B, where the immigrant coefficients are now non-significant, suggesting that in these countries the immigrant students characteristics and family background explain most of their low performance at school. In the other countries of

Groups A and B, coefficients have lower absolute values, but remain negative and significant.
In Germany and Italy, speaking a foreign language at home is significantly related to a lower record at school. In both countries, as seen above, a relevant proportion of immigrant students speak a non-national language at home. In general, a foreign language spoken at home is an important channel through which family backgrounds affect the school performance of students. With different values and degrees of significance, it is also negatively related to performance in Belgium and Spain, as well as in countries where the overall immigrant performance appears to be nearer to that of natives, as Hong Kong, Russia, Greece, Great Britain, or even better, as Australia, Canada and Montenegro. It also matters, but with a positive sign, in Qatar, where immigrants perform much better than natives.
A question that arises after controlling for this set of variables and after finding that they do not capture entirely the factors lying behind the students performance and, also, that they do not significantly alter the cross-country ranking of coefficients, is to what extent specific institutions in the receiving countries may be related to the immigrants' performance at school. Institutions that are expected to directly matter, in this case, are those regarding schooling and education. In particular, we are interested in a structural characteristic of educational systems, which is that of being based on either a common national program for secondary schools or on different programs and schools. To our knowledge, with the exceptions of Ammermueller (2007), Entorf and Lauk (2006) and Snepf (2007), very few studies have analysed this issue in relation to the PISA results.

3.3. Structural features of education systems

Most of the literature based on PISA that includes school variables take the latter from the school PISA dataset, which specifies schools characteristics, as, for example, being publicly or privately run, the teachers-students rates, the kind of final examination and other. Differently from these studies, we focus on structural features of the countries' education systems and hence on school educational programs. To have this information, we first use the UNESCO classification of countries' education programs (UNESCO, 2006) regarding year 2006 to distinguish between countries with education systems that, for fifteen year olds, are based on a common educational program and "comprehensive" schools, and those that are based on differentiated programs and school "tracks". For the latter, we use the UNESCO classification to split schools into three main categories: type 1 refers to schools preparing students for tertiary studies after graduation, type 2 may lead to further studies but, mainly, prepares them for direct access to the labour market, and type 3 leads just to the labour market. Several countries have also "special schools" for children with special needs; when appropriate, we include these schools in type 3. Once each school in each country is labelled following this classification (details are in Table A2 of the Appendix), we establish the type of school each student attends by using the variable *PROGN* of the PISA students' codebook, which provides this information.
Table 3 (see Appendix) lists the countries of our sample having differentiated school types. It can be easily seen that, with the exception of Denmark, education in all countries of Group A of Table 2 is based on the differentiated or the "tracking" school system. In Group B, the tracking system is present in France, Italy and Portugal. Differently, comprehensive schools characterize education in Norway, Sweden, Spain and Estonia of Group B, and in Great Britain, New Zealand, Australia, Canada and Qatar of Groups C and D. The Table depicts the values of an index of "specialization" of immigrants relatively to natives for each type of school in each country. The index number results from a fraction were the numerator is the share of immigrant students of the immigrant students' population in a certain school type, and the denominator is the share of native students of the native students' population in the same school type. Index values higher than unity indicate a higher relative presence of immigrant students in a certain type of school and lower than unity a higher relative presence of native students. Switzerland is not included in Table 3 because many foreign students move to the country independently from their families to complete their high school studies, so they are not proper "immigrants"; mainly they attend schools of type 1. The country is, however, included in Table 2 because the coefficients of the variables concerning school types 2 and 3 are

significant ([12]).
Table 3 shows that educational systems with the three different school types are especially present in our countries of group A, and only in some of groups B, C and D. The presence of first-generation immigrant students in schools of type 1, leading to university studies, is significantly below unity in all countries of groups A and B. The situation improves for second generation immigrant students in countries of group B, but remains below unity. The column concerning School type 3 offers a clear picture: in the first place, schools of type 3 are more present in countries of Group A, followed by countries of Group B; secondly, in these countries the relative presence of immigrant students in these schools is particularly high. This shows that the tracking system is especially located in central Western Europe, were, also, the relative presence of immigrant students is higher in schools of types 2 and 3. All index values are more heterogeneous in countries of groups C and D.
Hence, we run a third set of regressions were the structural characteristics of the school systems are added into the analysis. More specifically, we introduce into the regressions a variable indicating the school "type" each student attends. The linear regression can be written as:

$$Y_i = \beta_0 + \beta_I I_i + \beta_L L_i + \beta_X X_i + \beta_S S_i + \varepsilon_i \; i = 1,....,n \quad (3)$$

where S_i is a dummy representing the school "type" attended by student i and β_S is its coefficient.
The coefficients of S_i show the correlations between school types and student's performance, once family background and the other student characteristics have been controlled for. Together with the significant coefficients of the other variables of interest, they are shown in the third column, labelled **school,* regarding each country, of Table 2. It may be observed that *school.type* 2 and 3 have negative and significant coefficients in several countries where they are present. Their absolute values are higher and more significant, however, in the countries of Group A and in the three countries of Group B that have the tracking system. In group C they are high only in Greece and Russia, but lower than in Group A, and, in Group D, in Macao (were, however, immigrants perform better than natives). Hence, school types appear to matter especially in countries of Group A and, where present, in those of Group B.
For most countries, the coefficients of the immigrant variables in the columns **school* are significantly different than those of the columns *full regr*. It may be observed, however, that also in this case negative coefficients have higher absolute values in relation to the countries of central Western Europe: Belgium, France, Netherlands, Luxembourg, Switzerland, Germany, Italy and Portugal. In most cases the values of intercepts in the **school* columns, which comprehend the schools of type 1, are higher than in the *full regr* column. This is because now they include schools of type 1, were the average performance of students tends to higher. This may affect the immigrant variable coefficients: For example, in Germany and Italy, these coefficients become again significant.
As indicated by the coefficients of the *school types* variable, the existence of a tracking school system may add an important disadvantage to immigrant students, who may already be affected by a less favourable family background and the language spoken at home, and possibly also to other components of the students' population, that is not present in countries with comprehensive schools systems. Table 2 shows that several countries of central Western Europe tend to have not only more negative coefficients of the immigrant variable but that they also have the negative impact of the schools of types 2 and 3.

Conclusions

This study sheds a preliminary light on the relation between schooling institutions and the performance of immigrants at school. It shows that gaps in the immigrants-natives performance tend to be higher in countries located in central Western Europe, that these countries' educational systems are strongly based on the tracking system and that it significantly affects results.
These findings do not imply a causal link between schools and performance, they only register significant correlations. Also, other characteristics of schools might influence results. For example, in countries based on "comprehensive" schools, segregation may still exist and take place through

geography (poorer education in schools of poor neighbourhoods) or public vs. private schools. However, it might also happen that different types of school segregation simply add up rather than compensate for each other. Future developments of this research will focus on this and related questions.

Marina Murat, Giulia Pirani

* Department of Economics, University of Modena and Reggio Emilia.
** The present paper is the result of joint discussion between authors. However to Giulia Pirani are attributed the sections: 1, 2 and 3.1 and to Marina Murat sections: 3.2, 3.3, introduction and conclusions.

Notes

([1]) High correlation between the different domains considered allows as to do such an operation.
([2]) Up to the point where adding up a new regressor into the model makes the BIC increase.
([3]) As they are closed to the end of compulsory education.
([4]) In all cycles, the domains are covered in terms of students' ability to use their knowledge and skills to meet real life challenges and real world issues. All students took pencil-and-paper tests and answered to a mixture of multiple-choice items and questions requiring students to construct their own responses. Students were also asked to answer a background questionnaire in order to provide individual information on their social and economic background. For further details see the PISA web site: *www.pisa.oecd.org*.
([5]) A similar criterion of selection was adopted in OECD (2006), based on PISA 2003, which was focused on the performance in mathematics. Then, 17 countries were selected: Australia, Austria, Belgium, Canada, Denmark, France, Germany, Luxembourg, the Netherlands, New Zealand, Norway, Sweden, Switzerland, the United States, Hong Kong, Macao and the Russian Federation.
([6]) The assessment based on PISA 2003 (OECD, 2006), considers the share of all students that speak a non-national language at home, while we focus on just the immigrant population speaking a foreign language.
([7]) Students who are born outside the country of assessment and whose parents are also born in a different country.
([8]) Students who were born in the country of assessment but whose parents were born in a different country.
([9]) Full regressions are available from the authors upon request.
([10]) These and other cross-country results will be presented in future work by the authors.
([11]) The marked difference in the degrees of freedom between the first and the second set of regressions makes clear that the values of the two sets of coefficients cannot be compared. What can instead be compared are the orderings of countries in the two sets of regressions.
([12]) Data from the *Statistique Swisse* show that foreign students that have not completed elementary school in Switzerland show significantly lower rates of participation in vocational schools, and higher rates in general high schools or gymnasiums than foreign students that have attended elementary school in Switzerland (higher also than those of the general students' population): *www.bfs.admin.ch/bfs/portal/fr/index/themen/15/04/ind4.indicator.40101.401.html?open=412#412*.

References

Ammermueller, A. (2007) "PISA: What makes the difference? Explaining the gap in test scores between Finland and Germany", *Empirical Econ*omics 33:263-287
Entorf, H., Minoiu, N. (2004) "PISA results: what a difference immigration law makes", *IZA Discussion Paper*, No.1021.
Entorf, H., Lauk, M. (2006) "Peer effects, social multipliers and migrants at school: an international comparison", *IZA Discussion Paper*, No.2182.
OECD (2006) "Where immigrant student succeed- A comparative review of performance and engagement in PISA 2003", Paris, OECD Publishing.
Schnepf, S.V. (2007) "Immigrants' educational disadvantage: an examination across ten countries and three surveys", *Journal of Population Economics,* 20:527-545.
UNESCO (2006) "World data on education. Sixth edition, 2006/07", *www.ibe.unesco.org/Countries/WDE/2006/index.html*.

Main web sites

www.pisa.oecd.org/pages/0,2987,en_32252351_32235731_1_1_1_1_1,00.html
www.oecd.org/home/0,2987,en_2649_201185_1_1_1_1_1,00.html
www.wes.org/ewenr/research.asp
www.ibe.unesco.org/Countries/WDE/2006/index.html
www.bfs.admin.ch/bfs/portal/fr/index/themen/15/04/ind4.indicator.40101.401.html?open=412#412

Appendix

Table 1 – Share of immigrant students

	First generation	*Second generation*
HKG	18,96	24,74
QAT	17,89	21,86
LUX	16,24	19,38
MAC	15,5	56,61
NZL	15,09	7,45
ISR	11,71	11,19
CHE	10,28	11,48
AUS	8,16	11,53
AUT	7,59	5,01
DEU	6,56	7,71
BEL	6,51	5,92
CAN	5,32	6,67
GRC	5,26	1,19
RUS	5,18	4,2
MNE	4,93	1,44
SWE	4,77	6,26
IRL	4,37	1,08
ESP	4,18	0,52
DNK	3,52	4,23
ITA	3,47	0,66
NLD	3,38	7,71
FRA	3,3	9,53
NOR	3,16	3,14
PRT	3,07	2,12
GBR	2,19	2,67
SVN	1,97	8,34
EST	1,07	10,95
LVA	0,57	7,81

Table 2 – Performance of immigrant students: first and second generation, full regressions, and school types

Group A	AUT			BEL			CHE			DEU			LUX		
	im-migr	*full regr*	* *school*	*im-migr*	*full regr*	* *school*	*im-migr*	*full regr*	* *school*	*im-migr*	*full regr*	* *school*	*immigr*	*full regr*	* *school*
intercept	515,1	350,5	435,6	528,6	388,0	440,6	528,0	392,5	429,6	521,1	371,3	423,1	508,8	393,2	491,4
Immigr 1st gen	-65,1	-14,4	-35,2	-88,8	-28,4	-28	-86,5	-30,8	-32,3	-72,7			-64,8	-20,3	-24,4
Immigr 2nd gen	-75,1	-25,0	-23,8	-82,0	-34,1	-35,8	-62,0	-21,6	-23,5	-82,2		-13,7	-58,4	-19,2	-20,4
Lang. home other					-28,4	-34,1					-29,1				
School.type 2			-26,9			-46,8			-39,6			-37,2			-36,5
School.type 3			-52,9			-93,5			-31,8			-45,5			-59,1
degrees free-dom	4888	3469	4479	8740	5099	5354	12018	7894	7875	4600	2798	3263	4487	2905	2951
Adjusted *R2*	0,065	0,537	0,574	0,090	0,518	0,573	0,126	0,532	0,544	0,084	0,510	0,569	0,109	0,53	0,59

Group A	NLD			DNK	
	im-migr	*full regr*	* *school*	*im-migr*	*full regr*
intercept	535,3	414,6	546,7	506,8	354,9
Immigr 1st gen	-53,6	-34,2	-25,9	-80,8	-30,7
Immigr 2nd gen	-65,2	-28,6	-37,4	-73,9	-32,1
Lang. home other					
School.type 2			-67,0		
School.type 3			-144,4		
degrees free-dom	4784	3109	3514	4490	2689
Adjusted *R2*	0,053	0,505	0,701	0,065	0,420

Group B	FRA			ITA			PRT			ESP		EST		NOR		SWE	
	im-migr	*full regr*	* *school*	*im-migr*	*full regr*	* *school*	*im-migr*	*full regr*	* *school*	*im-migr*	*full regr*	*lm-migr*	*full regr*	*im-migr*	*full regr*	*immigr*	*full regr*
intercept	501,9	386,5	441,6	483,3	313,1	371,3	478,8	409,7	429,7	498,9	365,3	523,6	358,9	492,9	329,2	513,4	387,5
Immigr 1st gen	-56,7	-26,8	-19,6	-62,0		*-10,8*	-53,9	-15,6		-63,6	-17,8	-50,1	-21,1	-59,4	*-17,0*	-71,7	-27,9
Immigr 2nd gen	-42,9	-13,1	-20,1	-23,2			-31,0	-19,7	-19	-34,7		-35,0	-19,4	-50,0	-20,4	-39,2	-17,0
Lang. home other					-14,9	-11,0					*-8,1*						
School.type 2			-70,0			-43,8			-42,2								
School.type 3			-135,3														
degrees free-dom	4572	3188	3252	21257	11611	11974	5050	3413	3427	19364	11825	4753	3780	4582	3450	4359	3471
Adjusted *R2*	0,030	0,558	0,662	0,017	0,405	0,453	0,015	0,567	0,601	0,028	0,480	0,025	0,505	0,023	0,438	0,043	0,467

Group C	GRC			LVA			RUS			SVN			HKG			GBR		NZL	
	immigr	*full regr*	* *school*	*immigr*	*full regr*	* *school*	*immigr*	*full regr*	* *school*	*immigr*	*full regr*	* *school*	*immigr*	*full regr*	* *school*	*immigr*	*full regr*	*immigr*	*full regr*
intercept	470,1	384,7	413	492,6	353,9	381,1	469,9	325,4	367,2	486,6	369,5	374,8	550,1	420,0	433,4	505,2	348,0	530,8	377,2
Immigr 1st gen	-36,6						-12,6			-37,1			-27,7	-8,4	*6,7*	-28,9	-21,4	-8,1	-12,2
Immigr 2nd gen				-11,0	-12,0	-12,4	-20,9	*-10,5*	*-9,9*	-35,5	*-6,7*	*-6,7*				-13,6		*-17,0*	
Lang. home other		-22,2						-35,6	-40,5					-41,2	-35,8		*-13,3*		
School.type 2			-65,4			-27,5			-25,4			-9,7			-37,2				
School.type 3									-43,4										
degrees of freedom	4792	4021	4132	4593	3098	3097	5711	3862	3843	6483	4491	4490	4581	4372	4371	12748	8356	4708	3201
Adjusted *R2*	0,010	0,478	0,543	0,002	0,447	0,452	0,004	0,385	0,407	0,018	0,472	0,475	0,021	0,442	0,474	0,003	0,478	0,003	0,497

Group D	IRL			ISR			MNE			MAC			QAT		AUS		CAN	
	immigr	*full regr*	**school*	*immigr*	*full regr*	* *school*	*immigr*	*full regr*	* *school*	*immigr*	*full regr*	* *school*	*immigr*	*full regr*	*Immigr*	*full regr*	*immigr*	*full regr*
intercept	512,3	440,6	397,09	453,6	428,6	432,2	397,8	359,7	383,8	501,6	410,9	454,2	307,4	288,6	516,4	346,5	520,2	331,1
Immigr 1st gen						11,9	21,5	*13,0*	12,9			11,2	85,7	50,5			*-6,1*	
Immigr 2nd gen										13,2	*5,8*	*4,1*	37,0	16,2	11,2	4,7		*6,1*
Lang. home other								-31,1	-36,9					20,3		-8,4		-10,0
School.type 2			-25,65			-25,5			-35,5			-49,6						
School.type 3																		
degrees of freedom	4439	3474	3304	4198	3248	3306	4299	2216	2214	4669	2713	4380	5715	4837	13841	10027	21740	12218
Adjusted *R2*	0,001	0,347	0,418	2,975	0,423	0,420	0,004	0,454	0,498	0,008	0,458	0,472	0,154	0,411	0,002	0,476	0,000	0,408

BIC selection, OLS estimates. Coefficients: significant at 1% level; in Italics: significant at 5 and 10 % levels.

Figure 1 – Performance gap of immigrant students: groups of countries

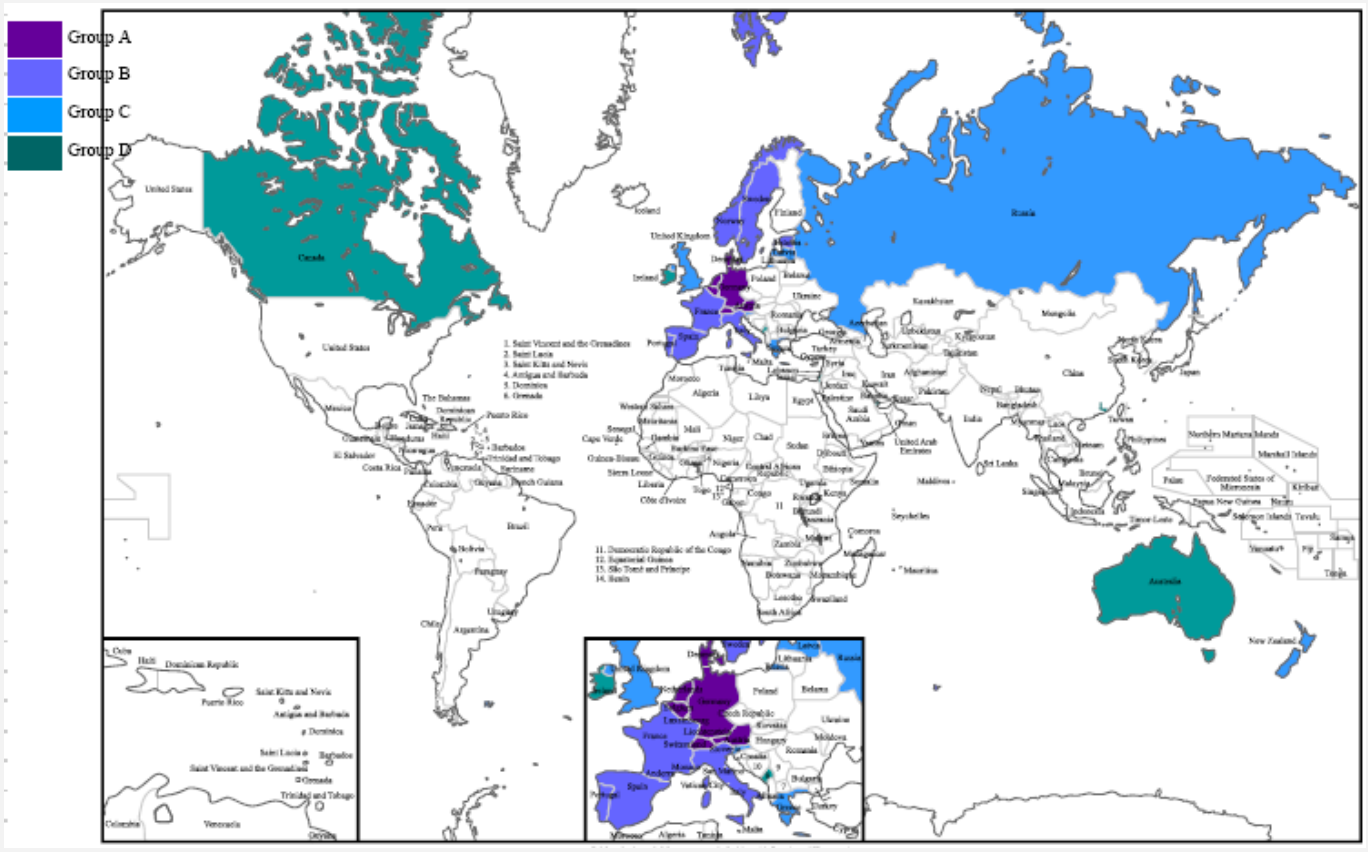

Table 3 – School types index: (% immigrant students) /(% native students)

	School 1		*School 2*		*School 3*	
	1st gen	*2nd gen*	*1st gen*	*2nd gen*	*1st gen*	*2nd gen*
AUT	0,84	1	1,08	0,91	1,01	1,12
BEL	0,81	0,97	1,09	0,95	3,65	2,74
DEU	0,73	0,7	0,48	0,14	1,24	1,28
LUX	0,74	0,68	0,7	0,85	1,17	1,19
NLD	0,71	0,58	0,86	0,99	2,03	1,75
average A	*0,77*	*0,79*	*0,84*	*0,77*	*1,82*	*1,62*
FRA	0,64	0,92	1,39	1,07	1,45	1,41
ITA	0,48	0,87	1,36	1,09		
PRT	0,54	0,8	1,34	1,15		
average B	*1,3*	*0,86*	*1,36*	*1,1*	*1,45*	*1,41*
GRC	0,61	0,93	2,92	1,36		
HKG	0,29	1	3,09	1,01		
LVA	2,28	1,24	0,96	0,99		
RUS	0,89	0,86	1,19	1,11	0,57	1,47
SVN	1,05	0,88	0,91	1,19		
average C	*1,02*	*0,98*	*1,81*	*1,13*	*0,57*	*1,47*
ISR	0,69	0,92	1,8	1,19		
MAC	0,37	1,09	1,41	0,94		
MNE	1,06	0,94	0,93	1,07		
IRL	1,48	0,68	0,93	1,04		
average D	*0,9*	*0,91*	*1,27*	*1,05*		
School 1: tertiary studies; School 2, mixed; School 3: labour market						

Table A1 – List of variables

Code PISA 2006	Variable	Meaning
st04q01	*gender*	Gender of student (1= female, 2= male)
st12q01	*language.home*	Language spoken at home (1= test language, 2 = other national language, 3= foreign language)
st15q01	*books*	How many books at home (1 = 0-10, 2 = 11-25, 3 = 26-100, 4= 101-200, 5 = 201-500, 6 = more than 500)
st13q04	*pc*	Computer at home (1= yes, 2 = no)
st14q03	*pcs*	How many computers at home (1 = none, 2 = one, 3 = two, 4 = more than three)
ic02q01	*usepc*	How long used computers (1 = less than 1 year, 2 = 1-3 years, 3 = 3-5 years, 4 = more than 5 years)
misced, fisced	*misced, fisced*	Educational level of mother/father (1 = none, 2 = ISCED 1, 3 = ISCED 2, 4 = ISCED 3B/C, 5 = ISCED 3A/4, 6 = I-SCED 5B, 7 = ISCED 5A/6)
bmmj, bfmj	*occupM, occumF*	Occupational status of mother/father (range 16- 90)
msecateg, fseca-teg	*categM, categF*	Socio-economics employment category of mother/father (1 = white collar high skilled, 2 = white collar low skilled, 3 = blue collar high skilled, 4 = blue collar low skilled)
hedres	*hedres*	Index of educational resources at home derived from students' reports on the availability of the following items in their home: *i)* a desk to study at; *ii)* a quiet place to study; *iii)* a computer they can use for school work; *iv)* educational software; *v)* their own calculator; *vi)* books to help with their school work; and *vii)* a dictionary.
homepos	*homepos*	Index of home possessions obtained by asking students whether they had at their home: a desk to study at, a room of their own, a quiet place to study, a computer they can use for school, an educational software, a link to the Internet, their own calculator, classic literature, books of poetry, works of art (*e.g.* paintings), books to help with their school work, a dictionary, a dishwasher, a DVD player or VCR, the number of cellular phones, televisions, computers, cars and books at home, and three other country-specific items.
cultposs	*cultposs*	Index of cultural possessions at home derived from students' reports on the availability of the following items in their home: classic literature (examples were given), books of poetry and works of art (examples were given).
wealth	*wealth*	Index of family wealth.
escs	*escs*	Index derived from the following variables: the *highest international socioeconomic index of occupational status* (HISCEI) of the father or mother; the *index of highest educational level of parents* (HISCED) converted into years of schooling and the *index of home possessions.*
St31q01, st31q0-4, stq07	*regularlessons. scie, regularlessons. math, regularlessons. read*	Number of regular lessons (weekly) – science, mathematics, reading (1 = none, 2= up to 2 ours, 3 = 2-4 ours, 4 = 4-6 ours, 5 = more than 6 ours)
st31q03, st31q06, st31q09	*selfstudy.scie, selfstudy.math, selfstudy.read*	Out of school study (weekly) – science, mathematics, reading (1 = none, 2= up to 2 ours, 3 = 2-4 ours, 4 = 4-6 ours, 5 = more than 6 ours)
envware	*envware*	Index of students' awareness of environmental issues was derived from students' beliefs regarding their own level of information on the following environmental issues: *i)* the increase of greenhouse gases in the atmosphere; *ii)* the use of genetically modified organisms (<GMO>); *iii)* acid rain; *iv)* nuclear waste; and *v)* the consequences of clearing forests for other land use.
envopt	*envopt*	Index of students' optimism regarding environmental issues was derived from students' optimism concerning the development over the next 20 years of the problems associated with the following environmental issues: i) air pollu-tion; ii) energy shortages; iii) extinction of plants and animals; iv) clearing of forests for other land use; v) water shorta-ges; and vi) nuclear waste.
envperc	*envperc*	Index of students' level of concern for environmental issues was derived from students' level of concern about the following environmental issues: i) air pollution; ii) energy shortages; iii) extinction of plants and animals; iv) clearing of forests for other land use; v) water shortages; and vi) nuclear waste.
respdev	*respdev*	Index of students' responsibility for sustainable development was derived from students' level of agreement with the following statements: i) it is important to carry out regular checks on the emissions from cars as a condition of their use; ii) it disturbs me when energy is wasted through the unnecessary use of electrical appliances; iii) I am in favour of having laws that regulate factory emissions even if this would increase the price of products; iv) to reduce waste, the use of plastic packaging should be kept to a minimum; v) industries should be required to prove that they safely dispo-se of dangerous waste materials; vi) I am in favour of having laws that protect the habitats of endangered species; and vii) electricity should be produced from renewable sources as much as possible, even if this increases the cost.

genscie	*genscie*	Index of general value of science was derived from students' level of agreement with the following statements: i) advances in <broad science and technology> usually improve people's living conditions; ii) <broad science> is important for helping us to understand the natural world; iii) advances in <broad science and technology> usually help improve the economy; iv) <broad science> is valuable to society; and v) advances in <broad science and technology> usually bring social benefits.
perscie	*perscie*	Index of personal value of science was derived from students' level of agreement with the following statements: i) some concepts in <broad science> help me see how I relate to other people; ii) I will use <broad science> in many ways when I am an adult; iii) <broad science> is very relevant to me; iv) I find that <broad science> helps me to understand the things around me; v) when I leave school there will be many opportunities for me to use <broad science>; and vi) some concepts in <broad science> help me see how I relate to other people.
scieact	*scieact*	Index of students' science-related activities was derived from the frequency with which students did the following things: i) watch TV programs about <broad science>; ii) borrow or buy books on <broad science> topics; iii) visit web sites about <broad science> topics; iv) listen to radio programs about advances in <broad science>; v) read <broad science> magazines or science articles in newspapers; and vi) attend a <science club>.
joyscie	*joyscie*	Index *of enjoyment of science* was derived from students' level of agreement with the following statements: *i)* I generally have fun when I am learning <broad science> topics; *ii)* I like reading about <broad science>; *iii)* I am happy doing <broad science> problems; *iv)* I enjoy acquiring new knowledge in <broad science>; and *v)* I am interested in learning about <broad science>.
instscie	*instscie*	Index of instrumental motivation to learn science was derived from students' level of agreement with the following statements: i) making an effort in my <school science> subject(s) is worth it because this will help me in the work I want to do later on; ii) what I learn in my <school science> subject(s) is important for me because I need this for what I want to study later on; iii) I study <school science> because I know it is useful for me; iv) studying my <school science> subject(s) is worthwhile for me because what I learn will improve my career prospects; and v) I will learn many things in my <school science> subject(s) that will help me get a job.
intscie	*intscie*	Index of general interest in science was derived from students' level of interest in learning the following topics: i) topics in physics; ii) topics in chemistry; iii) the biology of plants; iv) human biology; v) topics in astronomy; vi) topics in geology; vii) ways scientists design experiments; and viii) what is required for scientific explanations. A four-point scale with the response categories "high interest", "medium interest", "low interest" and "no interest" was used. All items were inverted for IRT scaling and positive values on this new index for PISA 2006 indicate higher levels of interest in science.
sciefut	*sciefut*	Index of future-oriented motivation to learn science was derived from students' level of agreement with the following statements: i) I would like to work in a career involving <broad science>; ii) I would like to study <broad science> after <secondary school>; iii) I would like to spend my life doing advanced <broad science>; and iv) I would like to work on <broad science> projects as an adult.

Table A2 – List of school types by country

AUT	BEL	CHE	DEU
0400002 = mixed	0560101 = mixed	7560001 = mixed	2760001 = tertiary studies
0400003 = mixed	0560103 = mixed	7560002 = labour market	2760002 = labour market
0400004 = labour market	0560104 = tertiary studies	7560003 = tertiary studies	2760003 = labour market
0400005 = labour market	0560105 = tertiary studies	7560004 = labour market	2760004 = tertiary studies
0400006 = mixed	0560106 = mixed	7560005 = labour market	2760005 = tertiary studies
0400007 = tertiary studies	0560107 = tertiary studies	7560006 = mixed	2760006 = tertiary studies
0400008 = mixed	0560108 = mixed	7560007 = labour market	2760008 = labour market
0400009 = tertiary studies	0560109 = labour market		2760009 = mixed
0400010 = labour market	0560110 = labour market		2760010 = mixed
0400011 = labour market	0560111 = labour market		2760012 = labour market
0400012 = labour market	0569612 = tertiary studies		2760013 = labour market
0400013 = labour market	0569613 = labour market		2760014 = labour market
0400014 = mixed	0569614 = mixed		2760015 = labour market
0400015 = mixed	0569615 = labour market		2760016 = mixed
	0569616 = tertiary studies		2760017 = tertiary studies
	0569617 = mixed		2760018 = labour market
	0569618 = mixed		2760019 = labour market
	0569619 = mixed		2760020 = labour market
	0569620 = labour market		
	0569622 = labour market		
	0569623 = labour market		
	0569624 = labour market		
FRA	GRC	HKG	IRL
2500001 = mixed	3000001 = mixed	3440001 = mixed	3720001 = mixed
2500002 = labour market	3000002 = tertiary studies	3440002 = tertiary studies	3720002 = mixed
2500003 = tertiary studies	3000003 = mixed	3440003 = mixed	3720003 = mixed
2500004 = mixed	3000004 = tertiary studies	3440004 = tertiary studies	3720004 = tertiary studies
	3000097 = NA		3720005 = mixed
ISR	ITA	LUX	LVA
3760001 = mixed	3800001 = tertiary studies	4420001 = labour market	4280001 = mixed
3760002 = mixed	3800002 = mixed	4420002 = labour market	4280002 = mixed
3760003 = tertiary studies	3800003 = mixed	4420003 = labour market	4280004 = tertiary studies
3760004 = tertiary studies	3800004 = mixed	4420004 = labour market	4280006 = mixed
3760005 = tertiary studies	3800005 = mixed	4420005 = mixed	
3760006 = mixed		4420006 = tertiary studies	
3760007 = mixed		4420007 = tertiary studies	
3760008 = mixed		4420008 = mixed	
3760009 = tertiary studies		4420009 = tertiary studies	
3760010 = mixed			
3760011 = tertiary studies			

MAC	MNE	NLD	PRT
4460001 = mixed	4990001 = mixed	5280001 = labour market	6200001 = mixed
4460002 = tertiary studies	4990002 = tertiary studies	5280002 = labour market	6200002 = mixed
4460003 = mixed	4990003 = mixed	5280003 = labour market	6200003 = tertiary studies
4460004 = tertiary studies	4990004 = mixed	5280004 = labour market	6200004 = mixed
	4990005 = tertiary studies	5280005 = labour market	6200005 = mixed
	4990006 = tertiary studies	5280006 = mixed	6200006 = mixed
	4990008 = tertiary studies	5280007 = labour market	6200007 = mixed
	4990009 = tertiary studies	5280008 = mixed	6200008 = mixed
	4990010 = mixed	5280009 = mixed	
	4990011 = mixed	5280010 = mixed	
		5280011 = tertiary studies	
		5280012 = tertiary studies	
		5280097 = NA	
RUS	SVN		
6430001 = mixed	7050001 = mixed		
6430002 = tertiary studies	7050002 = mixed		
6430003 = labour market	7050003 = tertiary studies		
6430004 = mixed	7050004 = mixed		
	7050005 = tertiary studies		
	7050006 = tertiary studies		
Source: PISA 2006 codebook and UNESCO (2006)			