EDUCATION ECONOMICS, 2017
VOL. 25, NO. 3, 304–322
http://dx.doi.org/10.1080/09645292.2016.1266301

A comparison of financial literacy between native and immigrant school students

Iulian Gramațki

Chair of Econometrics, Faculty of Economics (02), Goethe-Universität, Frankfurt am Main, Germany

ABSTRACT
This paper investigates the gap in Financial Literacy (FL) between native and immigrant 15-year-old school students using data from the 2012 PISA Financial Literacy Assessment. The size of the gap is about 0.15 standard deviations, going up to 0.3 for first-generation immigrants. This is partly because immigrants have poorer economic background, parents that work in lower-skilled jobs, do not speak the test language at home and are placed in later student cohorts. Controlling for this via OLS or matching reduces the unexplained gap, but it still remains significant and displays considerable country-level heterogeneity. It ceases to be significant when the Math score is partialled out.

ARTICLE HISTORY
Received 26 July 2016
Accepted 24 November 2016

KEYWORDS
Financial literacy; education; PISA; immigration

JEL CLASSIFICATION
A20; A21; I20; I21; I24

1. Introduction

The world we live in is becoming increasingly financially sophisticated. In order to perform well in a finance-heavy society, good knowledge of basic financial issues is crucial (Lusardi and Mitchell 2014). Such knowledge, which includes handling money and transactions, managing finances, identifying financial information, analyzing risk and return, is known under the umbrella term 'financial literacy.' (OECD 2013). Improving financial literacy (FL) is key for ensuring financial stability and, more broadly, a well-functioning society, both in developed and developing countries.

However, while standard literacy is close to 100% in a vast majority of countries in the world, recent empirical studies have shown that when it comes to FL, even developed countries are performing rather poorly. For instance, Lusardi (2008) shows that only 33% of a US sample received a full score in a short test of basic FL. Even among the young, who grew up in a more evolved financial sector, results are disappointing. Only 27% of the respondents in a sample from the National Longitudinal Survey of Youth (NLSY, USA) grasped the ideas of interest rates, inflation, and risk diversification (Lusardi, Mitchell, and Curto 2010). Furthermore, both studies show that FL is disproportionately distributed over demographic variables, with illiteracy being more pronounced in respondents that are Black, Hispanic, female, and/or poorly educated.

This paper uses the PISA Financial Literacy Assessment, which covers a subset of 18 countries that participate in PISA. It is the first comprehensive international survey of FL among teenagers. The rich micro-level dataset can be used to investigate interesting relationships between school students' FL and their various background characteristics. In particular, it is of interest whether there is a significant gap between students with an immigrant background and natives. Since FL does not constitute a separate school subject and is still not systematically taught in school in many countries (with some exceptions), adolescents often have to rely on their families to give them this knowledge. Their FL will therefore be highly dependent on their parents' expertise. In turn, immigrant parents generally originate from countries with lower FL levels.

CONTACT Iulian Gramațki iulian.gramatki@hof.uni-frankfurt.de

There is a growing body of literature which addresses the problem of native-immigrant disparities among school students. In this context, the PISA results are one of the main datasets used. For example, Kunz (2016) investigates the score gap between natives and immigrants, which he finds to be much larger in Germany than in German-speaking Switzerland, even after controlling for background characteristics. The author suggests that Swiss authorities have developed an education system which is more accommodative for immigrant schoolchildren. Entorf (2015) investigates bivariate cross-country correlation between the PISA Reading and Mathematics score of immigrants and their background characteristics. Unsurprisingly, the average country-wide PISA score of immigrant students is positively correlated with their index of economic and social status and with the number of books at home.

However, to date there have been no studies which combine migration issues and FL in a school setting. Native-immigrant disparities have been mostly analyzed in terms of Math or Language performance (Entorf 2015; Kunz 2016). Although papers on FL exist which narrow down to the subpopulation of young people (Lusardi, Mitchell, and Curto 2010), none so far have considered school students as the target category and none have focused on migration background as a determining factor of FL. This is the main contribution that the present paper aims to make.

It uses the PISA 2012 Financial Literacy Assessment to investigate whether immigrant students exhibit lower FL than native students and why this is the case. It then controls for relevant background characteristics to find out whether the gap remains, using the ordinary least squares (OLS) method and also propensity score matching (PSM) estimation as a sensitivity check to account for nonlinearities. It finds out that native students perform on average 15 PISA points better than all immigrant students, 30 PISA points better than first-generation immigrants and 5 PISA points better than second-generation immigrants (the mean of the PISA score being close to 500 and the standard deviation close to 100). Potential channeling factors are poorer economic background, parents that work in lower-skilled jobs, not speaking the test language at home and placement in later student cohorts. Controlling for all these factors reduces the unexplained gap, but it still remains significant at 12 points overall and 18 for first-generation immigrants. Using a propensity score matching estimator reduces the gap even further, and imposing a caliper restriction makes it insignificant. The FL gap also ceases to be significant when fixed effects are included, and there is significant country-level heterogeneity not just in levels, but in treatment effects as well. The gap in Math is slightly higher than the gap in FL, whereas the gap in reading is lower. Partialling out a student's numeracy by controlling for the Math score in the OLS regression makes the remaining FL gap insignificant, while the same outcome is not achieved when the Reading score is controlled for. Thus, the gap is not just an isolated effect on FL, but rather correlated with lower educational achievement in general and poorer numeracy in particular.

The remainder of the paper is structured as follows. Section 2 presents the PISA Financial Literacy dataset and Section 3 presents some introductory descriptive statistics. Section 4 briefly outlines the estimation method and the assumptions involved. Section 5 presents the findings, with three subsections devoted to the baseline OLS estimation, robustness checks, and PSM estimation, respectively, and Section 6 concludes.

2. Data

This study is based upon the data from the 2012 PISA Financial Literacy Assessment, run by the Organization for Economic Cooperation and Development (OECD). It is an integral but separate part of the 2012 wave of PISA. It has been conducted in 18 countries, of which 13 were OECD members. Thus, from some of the participating schools in these eighteen countries, besides the students taking the standard PISA test, eight additional students were chosen to receive the PISA Financial Literacy test (OECD 2014). The target groups of the test are school students aged between 15.3 and 16.2, who have completed at least six years of education. The PISA 2012 Financial Literacy data encompasses 29,041 students.

The test assessed students' performance in the areas of FL, Reading and Mathematics, with a separate score for each category. It is a paper-based test lasting two hours, split into four 30-minutes blocks. Two of these blocks assessed FL, one assessed Reading and one assessed Mathematics. The FL questions included the following content areas: money and transactions, planning and managing finances, risk and reward, and financial landscape. They require the student to identify financial information, analyze information in a financial context, evaluate financial issues, and apply financial knowledge and understanding in contexts typical for a 15-year-old, such as education, home and family, individual, and societal (further details in OECD 2013). Besides the test proper, the assessment also consisted of survey questions on the student's background, ICT familiarity, educational background and career interests, and experience with money matters. School principals and parents were also handed out background questionnaires.

Concerning sample selection, The PISA 2012 core test and the PISA 2012 FL test are different both in terms of the countries sampled and in terms of the students sampled in each participating school. The former difference is much more likely to affect the results than the latter. The naïve native-immigrant achievement gaps in Math and Reading are not replicated when we compare the whole core sample to the whole FL sample. Also, the total sample mean for the Math and Reading scores is about 20 points lower in the PISA core sample. However, the main reason for this is that the core PISA includes many more developing countries. 13 out of 18 countries in PISA FL are OECD members, whereas only 34 out of 68 in the core PISA are in the OECD. Also note that non-OECD countries mostly have low immigration (see Colombia and Shanghai-China in Table 3, for example). Therefore, it is much more reasonable to restrict ourselves to the subsample of OECD countries only. In this case, the native-immigrant gap, as well as the total sample means in Math and Reading are quite closely replicated.

The main dependent variable for our estimation is the student's projected score in FL. Aside from that, the scores in Mathematics and Reading will also be used for comparison. It is important to point out that the PISA student score is originally discrete, due to the finite number of test questions. In order to make it continuous, a distribution is constructed for each value of the point estimate, and then five random numbers, denoted plausible values, are drawn from it. The scores have been standardized such that the two-tier country average is 500 and the respective standard deviation is 100. The simple average of the FL score for the whole sample is 491.25.

In theory, analysis should be conducted on each of the five values independently, and then averaged in order to obtain the most accurate results. However, OECD (2009) illustrates that, with a sample of 6400 students, using one plausible value instead of five affects the mean estimate by less than 0.1% and the standard error estimate by less than 1%. The main findings of this paper are based on analysis using a single plausible value. Results using a different plausible value have been generated as a sensitivity check. They are not significantly different and are not reported for the sake of brevity.

The main independent variable is the student's immigration status. The PISA definition of an immigrant student will be used in this study. According to it, a student is defined as an immigrant if both his parents were born outside the country of residence. Whenever this applies, the student is classified as a first-generation immigrant if he/she was born outside the country of residence him/herself; the student is classified as a second-generation immigrant if he/she was born in the country of residence (OECD 2014). Immigration status is recoded into three dummy variables: *immigrant* will take the value 1 for all immigrants and 0 for natives; *immiFG* will take the value 1 for first-generation immigrants only and 0 for everybody else; *immiSG* will take the value 1 for second-generation immigrants only and 0 for everybody else. Other control variables will be described in the following section. Of the whole sample, 88.21% are natives, 4.12% are first-generation and 5% are second-generation immigrants.

A note is due here with respect to selection issues. There are 777 students who have not reported immigration status (2.68% of the sample). These students have clearly different characteristics from the total sample. Their PISA scores are 70–80 points below the average. Thus depending on whether these students are more or less likely to be immigrants, the total achievement gap might be over- or understated. Unfortunately, this information is unknown, and can only be conjectured. One way to estimate the share of immigrants among the observations with missing data is imputation, which will not be used in this paper since it may provide false significance and requires complex treatment of standard errors. However, looking at variables such as language spoken at home and material endowment, we can see that they are closer to immigrant levels than to native levels. One might therefore make an educated guess that there is a high share of immigrants among those who have not reported immigration status. As a consequence, the true achievement gap is higher, and the results in this paper provide a reasonable lower bound of the gap.

It is also important to note that the dataset features sampling weights in order to mitigate the selection bias arising from the two-tier sampling procedure (schools are sampled first, and then students are sampled within schools) (OECD 2009). These weights have been included in the analysis. Bootstrapping will not be used, and analytical standard errors based on Abadie and Imbens (2016) will be used instead for the PSM estimation. Note that Abadie and Imbens (2008) show that bootstrapping procedures to estimate standard errors are not consistent for matching estimators, even in the simplest case of one continuous covariate.

3. Descriptive statistics

Table 1 presents a summary list of the relevant variables used in this study, including their number of observations, means, standard deviations, minima, maxima, and descriptions. The sample has been narrowed down to those units included in the baseline estimation, thus dropping the observations with missing values for one or more covariates. Some of the explanatory variables are raw answers, such as gender, age, grade, family structure, language spoken at home, highest level among the mother's and father's education according to the ISCED classification, school size, class size, and student-to-teacher ratio. At the same time, other variables are constructed indices (*ESCS*, *HEDRES*, *HOMEPOS*, *WEALTH*, *SCMATEDU*, *SCMATBUI*). They are derived from several distinct raw answers to the background questionnaire and standardized to have a mean of 0 and a standard deviation of 1. For further details, see OECD (2014a, 2014b). *HISEI* is the highest among the mother's and father's occupation index, as developed by Ganzeboom (2010).

As a preliminary analysis, we will first see how these variables differ for natives and immigrants. The means of each variable for the whole sample, as well as for each immigration status category are presented in Table 2, along with *t*-statistics testing the hypothesis that the mean of the respective value for second-generation (first-generation) immigrants is not different from the mean for natives.

We can see that first-generation immigrants score significantly worse than natives on all three PISA FL subjects (FL, Mathematics, and Reading). Their score is about 30 PISA points (0.3 standard deviations) worse, and the difference is highly statistically significant. Second-generation immigrants, however, do not exhibit such an achievement gap. The gap in Math and FL is only about 5 PISA points (significant at the 10% level), whereas the results in Reading are statistically indistinguishable.

Native and immigrant students are also very different in terms of background characteristics, some of which may serve as channels of the performance gap. The biggest and most important difference, without doubt, is in the language spoken at home. While 87% of native students speak the test language at home, the share is 63% for second-generation and 42% for first-generation immigrants. Furthermore, all indices of material endowment are strongly and significantly lower for first-generation immigrants than for natives. The differences persist for second-generation immigrants as well, although to a lesser extent. Immigrants tend to go to bigger schools, but their average class size tends to be somewhat smaller. Surprisingly perhaps, schools attended by immigrants tend to be better endowed – the averages of the school resource quality indices are higher for immigrants, and the difference is statistically significant. Also, immigrants' parents tend to be employed more in low-skilled jobs (the parents' occupation index is lower). There are somewhat more male students among immigrants (53.3% for first generation and 52.4% for second generation) than among

Table 1. Description and summary statistics of the PISA variables relevant for our study.

Variable	Observations[a]	Mean	Standard deviation	Minima	Maxima	Description
PV1MATH	21,477	503.049	103.266	42.360	978.771	Plausible value 1 in mathematics
PV1READ	21,477	502.843	97.116	0.789	860.171	Plausible value 1 in reading
PV1FLIT	21,477	502.536	96.349	38.138	921.029	Plausible value 1 in FL
GENDER	21,477	1.491	0.500	1	2	Gender (1 = female, 2 = male)
AGE	21,477	15.784	0.290	15.25	16.33	Age of student
GRADE	21,477	−0.092	0.565	−3	2	Grade relative to modal grade in the country
CULTPOS	21,367	0.141	0.982	−1.51	1.27	Index of Cultural Possessions
ESCS	21,477	−0.038	0.936	−4.29	2.7	Index of economic, social and cultural status
FAMSTRUC	21,477	1.877	0.372	1	2	1 = single parent, 2 = two parents
HEDRES	21,446	0.088	0.940	−3.93	1.12	Home educational resources
hisced	21,477	4.511	1.380	0	6	Highest educational level of parents
hisei	21,477	50.155	21.240	11.01	88.96	Highest parental occupational status
HOMEPOS	21,464	0.024	0.941	−4.89	3.95	Home Possessions
WEALTH	21,463	−0.059	0.959	−5.63	3.12	Wealth
LANGhome	21,477	0.840	0.367	0	1	Test language spoken at home (1 = yes)
STRATIO	21,477	12.786	6.355	0.554	117.756	Student–Teacher ratio
SCMATEDU	21,477	0.057	1.026	−3.592	1.976	Quality of school educational resources
SCMATBUI	21,194	−0.057	1.033	−2.755	1.305	Quality of physical infrastructure at school
SCHLTYPE	21,477	2.809	0.494	1	3	1 = private, 2 = government dependent, 3 = public
SCHSEL	21,477	2.243	0.808	1	3	School selectivity (3 categories)
SCHSIZE	21,477	799.923	619.775	5	7600	Total school enrollment
CLSIZE	21,477	26.997	8.643	13	53	Class size
immigrant	21,477	0.086	0.280	0	1	1 if student is an immigrant
immiFG	21,477	0.039	0.195	0	1	1 if student is a first-generation immigrant
immiSG	21,477	0.046	0.210	0	1	1 if student is a second-generation immigrant

Notes: Includes the PISA variable name (for reference), number of observations, mean, standard deviation, minimum, and maximum.

[a]Summary statistics are reported here only for the observations included in the baseline regression, thus eliminating from the total 29,041 observations those that have missing values for at least one of the variables in the baseline model specification. Variables included in alternative model specifications may have a yet small number of observations reported.

natives (49.7%), with the differences significant at the 5% level. First-generation immigrants tend to be placed into a later student cohort than their age would imply (most likely due to repeating grades, especially at the moment of migration). The variables which are not significantly different are age (unsurprisingly, as PISA students represent the same age cohort), family structure and school selection criteria, with parents' education being significantly higher at the 10% level for first-generation immigrants only.

Next, we turn to an overview of the countries participating in the PISA Financial Literacy Assessment. The 18 countries along with their sample sizes, country-wide average scores and averages of some important control variables are presented in Table 3.

Although this has been discussed in great detail in OECD (2014), we will point out again that there are large disparities among the country-wide averages, which can go beyond two standard deviations (such is the case with 602 for China [Shanghai] and 395 for Colombia). Immigration levels also vary significantly. In some countries, the first- and second-generation immigrants together exceed 15% of the sample (Australia, New Zealand, the USA, and Israel). Other countries have remarkably low immigration levels (Colombia, Czech Republic, Poland, China [Shanghai] and Slovakia). The parents' occupation index (*hisei*) also varies across the sampled countries, as does school size, ranging from an average of around 300 in Poland and Slovenia to over 1000 in Colombia, New Zealand, Shanghai, and the USA. Finally it is worth noting that test language spoken at home is an imperfect predictor of immigration – a lower share of respondents who speak the test language at home can also be due to multilingualism in the country, as in the case of Belgium [Flanders], China [Shanghai] or Italy, for instance.

Table 2. Summary of differences between natives and immigrants.

Variable	Total	Native	Second generation	First generation	*t*-Statistics of second generation	*t*-Statistics of first generation
Number of observations	29,041	25,617	1451	1196		
Share	100.00%	88.21%	5.00%	4.12%		
Math score	491.760	495.157	489.940	465.379	−1.820*	−9.454***
Reading score	492.324	495.401	496.095	466.641	0.254	−9.537***
FL score	491.253	494.850	489.878	464.924	−1.839*	−10.084***
Gender	1.503	1.497	1.524	1.533	1.963**	2.445**
Age	15.784	15.784	15.781	15.793	−0.342	1.082
Grade (cohort)	−0.121	−0.111	−0.079	−0.338	2.028**	−12.899***
Cultural possessions	0.110	0.143	−0.120	−0.216	−9.808***	−12.282***
Economic status	−0.082	−0.067	−0.151	−0.268	−3.224***	−7.058***
Family structure	1.872	1.873	1.858	1.871	−1.347	−0.136
Home educational resources	0.034	0.046	−0.013	−0.080	−2.244**	−4.384***
Parents' education	4.483	4.482	4.492	4.552	0.266	1.675*
Parents' occupation	49.856	50.146	48.974	45.351	−1.989**	−7.411***
Home possessions	−0.033	−0.014	−0.115	−0.307	−3.870***	−10.157***
Wealth index	−0.101	−0.099	−0.060	−0.226	1.464	−4.313***
Test language at home	0.829	0.868	0.629	0.417	−25.561***	−43.986***
Student–teacher ratio	12.954	12.981	12.900	12.375	−0.430	−2.944***
School educational materials	0.016	−0.003	0.184	0.230	6.456***	7.303***
School buildings	−0.065	−0.073	−0.002	−0.008	2.426**	2.023**
School type	2.825	2.826	2.824	2.790	−0.105	−2.480**
School selectivity	2.225	2.227	2.219	2.205	−0.356	−0.872
School size	797.860	785.424	930.119	872.009	8.487***	4.619***
Class size	27.185	27.308	26.823	25.535	−1.983**	−6.593***

Notes: Columns 2–5 present the means of relevant variables by immigration status. The last two columns present the results of the *t*-tests for the difference in means between natives and the respective immigrant category.
*$p < .1$.
**$p < .05$
***$p < .01$.

Table 3. Overview by country.

Country	Number of Observations	FL score	Second generation (%)	First generation (%)	ESCS	hisei	LANGhome (%)	SCHSIZE
Australia	3293	515.193	9.77	8.28	0.178	56.720	86.58	922.811
Belgium (Flanders)	1093	544.462	6.26	4.11	0.198	52.756	66.24	714.067
Colombia	2100	394.892	0.34	0.15	−1.071	41.223	96.90	1441.912
Czech Republic	1207	529.465	1.67	1.25	0.077	49.811	94.37	548.927
Spain	1108	486.314	2.11	9.64	−0.135	48.211	81.95	757.541
Estonia	1088	531.280	9.52	0.65	0.141	51.311	92.83	590.423
France	1068	488.158	9.85	3.82	−0.021	52.074	89.04	835.217
Croatia	1145	479.631	7.78	3.89	−0.368	44.439	96.68	603.614
Israel	1006	480.758	12.10	5.07	0.206	58.596	84.29	821.627
Italy	7068	472.232	2.22	4.79	−0.011	47.344	74.18	678.664
Latvia	970	501.504	3.59	0.42	−0.131	50.566	88.66	543.260
New Zealand	957	523.061	10.49	15.31	0.102	55.902	83.18	1145.765
Poland	1054	514.555	0.00	0.10	−0.143	46.782	97.34	308.697
China (Shanghai)	1197	601.901	0.42	0.84	−0.418	50.282	36.51	1409.168
Russia	1187	486.854	7.44	2.74	−0.052	53.573	91.66	681.481
Slovakia	1055	474.614	0.48	0.58	−0.136	43.540	89.95	479.299
Slovenia	1312	466.962	5.42	3.56	0.009	49.549	90.55	345.488
United States	1133	491.289	15.90	5.71	0.161	54.360	82.97	1374.723

Note: Reports the sample size, average FL score, share of immigrants, and the means of some relevant covariates for each country included in the PISA FL.

There are several channels through which migration background can lead to worse educational achievement. One such potential candidate is the grade (i.e. form, cohort, school year) that the student is enrolled in. Advancing one grade leads to an increase in about 50 PISA points on the FL assessment, and first-generation immigrants tend to be in lower cohorts than their age would imply. Other valid candidates are all indices of the student's economic and cultural endowment, as well as the parents' occupation index. Speaking the test language at home is also a valid channel, as it leads to an improvement of circa 15 PISA points and is overwhelmingly higher for natives. Not valid candidates are age, since there are no significant differences between natives and immigrants, gender, as males perform slightly better but there are also more males among immigrant students, and also family structure and parents' education for similar reasons. Curiously, school-related variables (school material endowment, school size, class size) do a poor job in explaining the achievement gap.

In order to paint a more accurate picture of the native-immigrant FL gap and its drivers, we turn to more complex estimation techniques, the methodology of which is outlined in the next section, and the results are presented in Section 5.

4. Estimation method

The findings of this paper are a result of OLS and PSM estimation. OLS is one of the most basic econometric techniques and is extensively covered in the literature. It requires the zero conditional mean assumption $E(u_i|X_i) = 0$, full rank, and random sampling, in order to be both consistent and unbiased. Finally, under the extra assumption of homoscedasticity $E(u_i u_i'|X_i) = \sigma^2 I_N$, the OLS estimator is efficient.

The extent to which these assumptions hold for our dataset is debatable. The zero conditional mean can be violated in many cases, such as the presence of omitted variables, simultaneous (bidirectional) causal links between the regressand and the regressors, or measurement error, among others. In our particular case, the test score is an outcome variable, and is therefore unlikely to exhibit reverse causality onto demographic variables. Measurement error is an issue in every survey, but since its true functional form is unknown, there is generally no straightforward method to account for it. Finally, there is no reason to suspect that homoscedasticity holds, as poorer students are more likely to exhibit higher score variance, for example. To account for that and also for potential correlation of the error terms of students from the same school due to school-level effects, errors will be heteroscedasticity-robust and clustered at the school level.

Besides the considerations above, immigration status is not an exogenous variable. The decision to migrate is usually taken by the student's parents based on several factors: economic status in the country of origin, potential calamities in the country of origin (e.g. fleeing a war area), labor market opportunities in the destination country, cultural similarities between the home and destination countries, cost of migration, and others. While some of these, such as calamities, are unlikely to affect the migrant student's test score, others might have a bigger impact. Those that do have an impact are related to the family's socio-economic status, the parents' education and employment area, as well as language barriers between the home and destination countries. We make use of the extensive questions asked in the PISA background surveys to control for as many of the relevant factors causing endogeneity as possible. This approach is also used, among others, by Fuchs and Woessmann (2004) in their analysis of PISA data. The inclusion of additional covariates captures that part of the hidden bias which is correlated with at least one control variable. Although it does not free us from the problem of omitted variables, it substantially diminishes the bias arising from violation of the conditional mean independence and unconfoundedness assumptions by reducing the correlation between the treatment variable and the residual error term.

Matching estimators in general, and PSM in particular, are a more recent addition to the pool of econometric tools. Matching is used to analyze the impact of a binary variable D on an outcome variable Y. The former is often called a treatment variable, although it can be any binary variable, such as gender or immigration status, and does not have to imply deliberately induced treatment. The technical derivation, proof of consistency and asymptotic properties are presented in detail in Abadie et al. (2004) and Abadie and Imbens (2006, 2008, 2016). The remainder of this section will present a brief overview of these.

Suppose that we defined for every observation the potential outcomes for the treated and untreated states, Y_{1i} and Y_{0i}. One of them is always realized and the other one is always counterfactual. If we assume that these two potential outcomes are independent of the assignment of the treatment variable, that is, $Y_{1i}, Y_{0i} \perp D_i$, then the effect of the variable D can be extracted as the simple difference of means between the treated and untreated group. The above assumption is equivalent to random assignment of the treatment variable D, which is very seldom the case. However, suppose that we know the confounding variables X which affect both the outcome Y and the treatment status D. It then holds that $Y_{1i}, Y_{0i} \perp D_i | X_i$, also known as the ignorability assumption. We can then select, or match, units from the treated and untreated groups which are 'very close' in terms of X, and for these units, assignment into the treated or untreated group will be 'as good as random.' In order to be able to find a match for every unit, we also need what is known as the overlap assumption or the common support assumption, namely that $0 < P(D = 1|X = x) < 1$ for every value of x. Then, the difference in the outcome variable between two matched units will constitute the Local Average Treatment Effect (LATE): $LATE_i = Y_i - Y_{m(i)}$ if unit i is treated and $LATE_i = Y_{m(i)} - Y_i$ if i is untreated, where $m(i)$ is the unit matched to unit i. The sample average of these differences will constitute the Average Treatment Effect (ATE):

$$\widehat{ATE} = \sum_{i=1}^{N} \frac{1}{N} LATE_i,$$

which is used to make inference about the impact of the binary variable D on the outcome variable Y.

The concept of propensity score is introduced to address the dimensionality problem. The regressor matrix X can contain a multitude of variables, and it can become hard to interpret what it actually means for two observations to be 'very close' in terms of X. A one-dimensional metric must be defined in order to be able to evaluate and compare proximity. The PS is an example of such a metric. It is denoted $\pi(X)$ and defined as the probability of being treated given a certain value of X: $\pi(X) = P(D = 1|X)$. It is normally estimated by a binary outcome model such as Logit or Probit. The underlying system of equations is a latent variable model:

$$D_i = I[D_i^* > 0]$$

$$D_i^* = \boldsymbol{x}_i \boldsymbol{\beta} + u_i, \quad u_i \sim N(0, 1)$$

The coefficient estimators $\hat{\beta}$ are obtained via maximum likelihood estimation, after which the predicted propensity score $\widehat{\pi(X_i)} = \Phi(X_i \hat{\beta})$ (where Φ is the cumulative distribution function of a standard normal distribution) is generated for every observation and used in the next step as the matching variable. Rosenbaum and Rubin (1983) prove that the original ignorability assumption implies $Y_{1i}, Y_{0i} \perp D_i | \pi(X_i)$, and therefore consistent estimation can be done using the PS. The common support assumption then becomes $0 < P(D = 1|\pi(X)) < 1$, that is, that both treated and untreated units exist for every realized value of the propensity score.

It is important to note that, unlike OLS, the matching estimator is non-parametric and therefore robust to nonlinear model specifications. Given that it is based on averaging individually estimated local treatment effects, it will be valid given any functional relationship between Y

and X, so long as the ignorability assumption holds. However, like OLS, it is not immune to omitted variable bias. The ignorability assumption is in many ways equivalent to a conditional exogeneity assumption for D. Therefore, just as in the case of OLS, we rely on our large pool of control variables included in X to try to ensure random assignment of the treatment variable given these controls. However, we cannot make it absolutely certain that no variables have been left out that may lead to violations of the ignorability assumption (and conditional mean independence for OLS).

The common support assumption is tested by providing density graphs of the propensity scores for the treated and control group, which can be observed in Figure 1. We can see that the fit is imperfect. Matching second-generation immigrants to natives produces a better fit than matching first-generation immigrants to natives in terms of the closeness of the two lines on the graph. However, note that it is not the similarity of the density values at every point that matters, but rather the existence of positive density over the same domain. No propensity score values below 0.2 are encountered (for all immigrants; first graph), and the minimum PS values for the treated group are about 0.07 higher than for the untreated group.

Also, balance between the treated and untreated groups in terms of covariate means is tested. Appendix Table A1 presents the means of covariates for the treated and untreated groups (considering all natives as treated and all immigrants as untreated, that is, using *immigrant* as the relevant treatment variable), as well as their percentage deviation and the t-statistics for the significance of the difference in means. These are presented before matching, after matching without calipers and with calipers, for different caliper sizes. A caliper restriction means that matching will be performed only when the two units have a difference in propensity score lower than the specified tolerance level. It is normally used to achieve better balance and to enforce the common support assumption, but its effectiveness is questionable in our particular case.

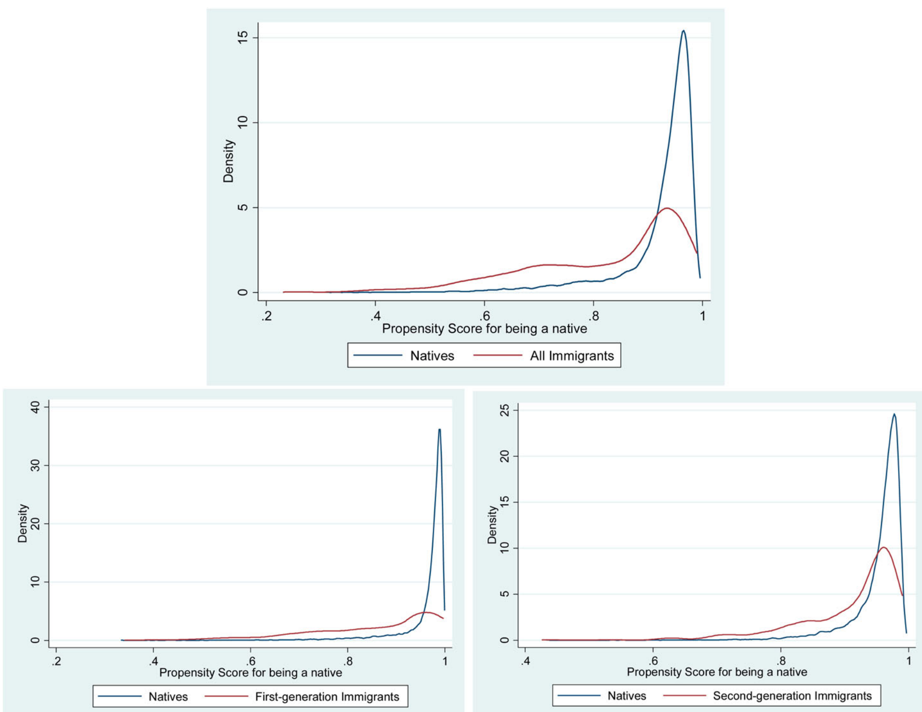

Figure 1. Common support check. Density graphs for the propensity score of treated and untreated groups.

5. Findings

5.1. OLS estimation

Table 4 presents the results of the comprehensive OLS regressions, including the variables of interest (immigration status) and the whole pool of control variables mentioned before. Alongside the FL score (our main outcome variable), the Math and Reading scores have been regressed for comparison. There are two main model specifications, one using a single dummy variable for all immigrants, and the other using separate dummies for first- and second-generation immigrants.

The first result one can see is that the native-immigrant gap in FL still persists, even after controlling for numerous student and school background characteristics. The gap amounts to about 12 points if we consider all immigrants, 18 points for the first generation, and 7 points for the second generation. The first two are significant at the 1% level, whereas the latter is significant at the 5% level.

Table 4. Estimated coefficients of multivariate OLS regressions.

	FL score	Math score	Reading score	FL score	Math score	Reading score
Immigrant	−12.048	16.112	−6.509			
	(4.25)***	(5.26)***	(2.43)**			
First-generation immigrant				−18.139	−22.123	−14.330
				(4.94)***	(5.73)***	(3.87)***
Second-generation immigrant				−7.105	−11.233	−0.161
				(1.96)**	(2.93)***	(0.05)
Gender	8.227	20.358	−26.520	8.221	20.352	−26.527
	(6.24)***	(14.46)***	(20.28)***	(6.24)***	(14.46)***	(20.29)***
Age	−0.457	−1.564	−5.904	−0.330	−1.439	−5.741
	(0.20)	(0.63)	(2.58)***	(0.14)	(0.58)	(2.51)**
Grade (school cohort)	31.954	25.831	26.850	31.758	25.637	26.598
	(21.40)***	(15.68)***	(17.24)***	(21.26)***	(15.58)***	(17.11)***
Economic status	26.005	30.654	31.223	25.949	30.599	31.151
	(11.41)***	(11.42)***	(13.37)***	(11.37)***	(11.39)***	(13.33)***
Family structure	−7.898	−4.249	−5.878	−7.858	−4.210	−5.827
	(4.64)***	(2.32)**	(3.49)***	(4.62)***	(2.30)**	(3.46)***
Parents' education	−3.420	−3.940	−5.236	−3.354	−3.875	−5.151
	(3.76)***	(3.90)***	(5.74)***	(3.69)***	(3.83)***	(5.64)***
Parents' occupation	0.317	0.257	0.293	0.316	0.256	0.292
	(4.77)***	(3.38)***	(4.41)***	(4.76)***	(3.38)***	(4.40)***
Test language at home	−6.573	−12.275	1.144	−6.917	−12.614	0.702
	(2.51)**	(4.04)***	(0.44)	(2.63)***	(4.14)***	(0.27)
Student–teacher ratio	−1.801	−3.063	−2.080	−1.798	−3.060	−2.077
	(5.01)***	(6.20)***	(6.06)***	(5.00)***	(6.18)***	(6.03)***
School material endowment	9.014	8.895	8.104	9.031	8.911	8.125
	(7.88)***	(6.62)***	(7.23)***	(7.90)***	(6.63)***	(7.25)***
School type (public/private)	−15.871	−8.079	*−9.479*	−15.885	−8.092	−9.497
	(7.29)***	(3.06)***	*(4.47)****	(7.30)***	(3.07)***	(4.49)***
School selection criteria	2.207	3.586	*2.740*	2.230	3.609	2.770
	(1.81)*	(2.60)***	*(2.25)***	(1.83)*	(2.62)***	(2.28)**
School size	0.016	0.018	0.017	0.016	0.018	0.017
	(6.73)***	(6.53)***	(7.03)***	(6.71)***	(6.51)***	(7.01)***
Class size	0.879	1.059	0.776	0.875	1.054	0.771
	(5.62)***	(5.49)***	(5.15)***	(5.60)***	(5.48)***	(5.12)***
constant	548.050	535.622	671.146	546.101	533.699	668.643
	(14.15)***	(12.60)***	(17.57)***	(14.12)***	(12.56)***	(17.52)***
R^2	0.20	0.20	0.21	0.20	0.20	0.21
N	21,477	21,477	21,477	21,477	21,477	21,477

Notes: Dependent variables presented in the column header. Independent variables listed in the row header. *t*-Statistics in parentheses.
*$p < .1$.
**$p < .05$.
***$p < .01$.

The second thing to note is that the gap in FL is smaller than the gap in Math, but larger than the gap in Reading. Also, the gap in Math is significant at the 1% level even for second-generation immigrants, whereas the gap in Reading for the second generation is nonexistent. This might suggest that one of the hurdles of immigration, namely the language barrier, which is most likely to affect Reading, is no longer an issue for second-generation immigrant students, who were already born in the host country. At the same time, other hurdles, such as material underendowment and financial hardship, are much more persistent and still likely to affect the second generation.

The coefficients of various control variables also deserve attention. For example, males perform (*ceteris paribus*) 8 PISA points better in FL and 20 points better in Math, but they perform 26 points worse in Reading, which reinforces common gender-based ability stereotypes. The effect of being in a higher grade is still large and highly significant. Increasing one's economic and socio-cultural status (ESCS) by one standard deviation yields an increase of 26 points in FL, along with even slightly higher improvements in Reading and Math. Speaking the test language at home now has a negative effect on FL and Math (but not Reading) once immigration status has been accounted for; this points out to better numerical performance (*ceteris paribus*) of bilingual students. FL is significantly and positively affected by the parents' occupation, but (surprisingly) negatively affected by their education (significant at the 1% level). One can suppose that parents who have higher-skilled jobs are more likely to be financially literate themselves and pass that knowledge on to their children. School material endowment and a lower student-to-teacher also positively affect the student's test performance.

5.2. Robustness checks

We have seen previously (in Section 3 and Table 3) that there is significant variation both in the FL score and the relative number of immigrants across the sampled countries. It might therefore be the case that the original treatment effect estimates are biased by country-wide differences in ability or in other unobserved variables. To remedy this, we include country-level and region-level fixed effects in alternative specifications. After such fixed effects are included, the resulting gap ceases to be statistically significant, as can be seen in the first three columns of Table 5.

While at a first glance one might interpret the baseline results as spurious and owing to mere country-level differences in levels, fixed effects might also capture migrant selection issues or the different relative success achieved in different countries or regions in terms of migrant integration. It is thus important to account for the fact that not only the levels of the outcome variable may be different across countries, but also the native-immigrant gaps, as well as the effects of our covariates. To further explore this possibility, disaggregated regressions by country have been performed. The results of these are displayed in Table 6 (left part). The regression model used is the same as in the baseline model, except that the observations are restricted to be from the respective country For the sake of brevity, only the coefficient of the migration dummy variable is shown.

The gap is small in magnitude and not statistically significant for New Zealand and the USA, moderate and somewhat significant for Israel, and large and highly significant for Australia. Since immigrants in these countries are often admitted on skill-based criteria, this leads to the migrant pool being more skilled both than the average native in the host country and their average compatriot in the home country. With the exception of Australia, the second generation exhibits milder gaps, as they are born in the host country and thus no longer subject to the skill-based selection mechanism.

The situation is mostly opposite in recent European destinations. Dominated by labor migration, newcomers to these countries are no longer more skilled than the average, but also have to bear the migration costs and cope with adaptation difficulties. One of the highest negative gaps can be seen in France, and amount to over 30 points in the FL test. The Flemish Community of Belgium displays a large and significant (at the 10% level) gap of 37 points for first-generation immigrants, but this vanishes for the second generation. Spain shows a smaller negative gap which is statistically

Table 5. OLS robustness checks.

	Baseline	Country FE	Region FE	Partial out Math	Partial out Reading	Partial out both
Math score				0.714		0.221
				(126.53)***		(9.65)***
Reading score					0.723	0.105
					(104.05)***	(4.86)***
Math–reading interaction						0.001
						(12.15)***
Immigrant	−12.048	−3.197	−1.029	−0.542	−7.342	−2.276
	(4.25)***	(1.22)	(0.40)	(0.33)	(4.12)***	(1.53)
Gender	8.227	9.711	13.750	−6.312	27.403	7.843
	(6.24)***	(7.98)***	(12.25)***	(7.50)***	(30.24)***	(9.49)***
Age	−0.457	−2.539	−2.245	0.661	3.812	1.186
	(0.20)	(1.25)	(1.16)	(0.45)	(2.41)**	(0.87)
Grade (school cohort)	31.954	38.526	38.080	13.507	12.539	11.749
	(21.40)***	(29.78)***	(29.06)***	(14.37)***	(12.22)***	(13.45)***
Economic status	26.005	32.027	25.471	4.112	3.427	2.394
	(11.41)***	(16.17)***	(13.48)***	(3.09)***	(2.37)**	(1.93)*
Family structure	−7.898	−7.928	−6.895	−4.863	−3.647	−4.486
	(4.64)***	(5.06)***	(4.58)***	(4.29)***	(2.98)***	(4.25)***
Parents' education	−3.420	−6.145	−5.272	−0.606	0.366	−0.335
	(3.76)***	(7.41)***	(6.66)***	(1.10)	(0.62)	(0.66)
Parents' occupation	0.317	0.064	−0.024	0.134	0.105	0.036
	(4.77)***	(1.11)	(0.43)	(3.38)***	(2.46)**	(0.98)
Test language at home	−6.573	12.688	15.358	2.193	−7.400	0.078
	(2.51)**	(5.76)***	(7.91)***	(1.67)*	(4.96)***	(0.07)
Student–teacher ratio	−1.801	0.247	0.079	0.387	−0.297	0.346
	(5.01)***	(1.23)	(0.38)	(3.61)***	(1.94)*	(3.49)***
School material endowment	9.014	4.869	3.399	2.662	3.155	2.089
	(7.88)***	(5.35)***	(4.20)***	(4.66)***	(4.85)***	(4.03)***
School type (public/private)	−15.871	−12.116	−13.885	−10.101	−9.017	−8.205
	(7.29)***	(6.23)***	(6.87)***	(8.93)***	(6.80)***	(7.91)***
School selection criteria	2.207	2.414	2.661	−0.354	0.225	−0.777
	(1.81)*	(2.12)**	(2.50)**	(0.50)	(0.29)	(1.20)
School size	0.016	0.006	0.009	0.003	0.004	0.001
	(6.73)***	(3.64)***	(5.82)***	(2.61)***	(3.06)***	(1.08)
Class size	0.879	0.369	0.172	0.123	0.318	0.033
	(5.62)***	(3.00)***	(1.53)	(1.73)*	(3.56)***	(0.50)
constant	548.050			165.522	62.741	204.878
	(14.15)***			(6.84)***	(2.35)**	(7.89)***
R^2	0.20	0.32	0.41	0.67	0.61	0.72
N	21,477	21,477	21,477	21,477	21,477	21,477

Notes: Model specification listed in column header. Dependent variable is FL score in all specifications. *t*-Statistics in parentheses.
*$p < .1$.
**$p < .05$.
***$p < .01$.

insignificant, whereas in Italy the gap is close to 0. Slovenia shows a result which fits into this category, with a gap of −25 points in FL, significant at the 10% level. Estonia, a regional migration destination (for the former Soviet Union mostly) also exhibits a large and significant positive gap. It is striking that some pairs of neighboring countries with similar cultures, such as Croatia and Slovenia or Latvia and Estonia, have rather different gaps. The results for some countries were not reported because the share of immigrants among the students who took the PISA test was less than 2%.

In an attempt to further elaborate on this issue, we present in Table 7 some details on migrant composition in the countries under investigation. We can see that countries with zero or positive gaps (Australia, New Zealand, Israel) have negligible disparities in background characteristics between natives and immigrants, having been more successful in integrating them. On the other hand, for the countries with high negative effects (Spain, France, Belgium, Slovenia), migrant integration seems more problematic: immigrant schoolchildren tend to be considerably poorer than their native counterparts, and their parents are on average less skilled. The only oddity appears for the USA, where migrants are worse endowed, but the performance gap is positive nonetheless.

Table 6. OLS and matching results by country (listed in row header). Model type and treatment variable specified in column header.

	OLS (all immigrants)	OLS (first generation)	OLS (second generation)	Matching (all immigrants)	Matching (first generation)	Matching (second generation)
Total	−12.048	−18.139	−7.105	−6.715	−1.009	−7.390
	(4.25)***	(4.94)***	(1.96)**	(1.99)**	(0.19)	(1.90)*
Australia	33.631	23.652	48.561	21.520	2.070	40.732
	(6.38)***	(3.35)***	(7.36)***	(2.45)**	(0.21)	(3.88)***
Belgium (Flanders)	−13.676	−37.330	3.950	−41.710	−55.642	−27.952
	(1.13)	(1.76)*	(0.34)	(1.85)*	(1.67)*	(0.85)
Czech Republic	20.566					
	(1.21)					
Spain	−8.443	−3.293	−25.293	17.582	−68.058	76.169
	(0.93)	(0.37)	(0.98)	(1.35)	(4.47)***	(0.84)
Estonia	−35.569		−36.322	−23.467		−45.891
	(4.20)***		(4.17)***	(1.84)*		(5.11)***
France	−31.721	−37.019	−28.131	−5.403		−24.562
	(3.28)***	(1.62)	(2.70)***	(0.53)		(3.13)***
Croatia	−2.455	12.829	−9.124	−4.217	9.917	−16.115
	(0.30)	(0.90)	(1.00)	(0.29)	(0.44)	(1.66)*
Israel	22.373	39.895	18.687	18.981	14.806	10.387
	(2.29)**	(2.25)**	(1.78)*	(2.29)**	(2.23)**	(1.21)
Italy	2.511	−0.428	6.441	−4.762	−4.149	3.497
	(0.51)	(0.07)	(0.90)	(0.48)	(0.15)	(0.28)
Latvia	−7.407		1.372			
	(0.63)		(0.14)			
New Zealand	−3.630	8.801	2.128	−1.156	17.874	−41.387
	(0.28)	(0.72)	(0.10)	(0.06)	(1.34)	(2.02)**
Russia	−22.049	1.986	−30.216	−2.159	17.006	−22.301
	(2.28)**	(0.12)	(2.73)***	(0.16)	(0.73)	(1.38)
Slovenia	−24.488	−26.369	−26.977	−33.551	−66.666	−17.655
	(1.81)*	(1.27)	(1.80)*	(3.65)***	(1.01)	(1.26)
USA	4.801	20.225	3.707	6.035	−38.919	21.518
	(0.42)	(1.41)	(0.28)	(0.17)	(0.72)	(1.33)

Notes: Estimated treatment effect (coefficient of the immigrant dummy variable) reported. Note that some countries have a remarkably low number of immigrant students. The coefficients for these countries are highly unreliable and should not be used for inference, in spite of their occasional statistical significance. That is why they have been left blank. *t*-Statistics in parentheses.
*$p < .1$.
**$p < .05$.
***$p < .01$.

Therefore, there may be two reasons why the baseline gap disappears in the fixed effect model: countries with a high share of immigrants are both (a) attracting more skilled immigrants owing to their immigration policy, and (b) more successful in integrating them.

Similarly to the country-level heterogeneity, one might expect considerable heterogeneity by gender or student ability. To provide some insight on gender disparities, separate estimations have also been performed for the male and female subsamples. These are reported in Table 8. Note that the coefficients represent the gap between immigrants and natives for the total sample and for the male and female subsamples, and not the gap between males and females. One can note that there are no significant differences between boys and girls in the native-immigrant gaps.

An approach to investigating disparities due to ability is to estimate the treatment effect not only at the mean, but also at other, more extreme points in the distribution. One way to do this would be to perform quantile regressions according to Koenker and Bassett (1978). More recently, a refined approach, known as generalized quantile regression, has been developed by Powell (2016). This approach can also be used to generate counterfactual distributions in a straightforward manner, and under some conditions outperforms the conditional distribution regression developed by Chernozhukov, Fernandez-Val, and Melly (2013). The results of generalized quantile regression are displayed in Table 9. The 25th, 50th and 75th percentiles display similar gaps to the mean. First-

Table 7. Migrant composition by country. Means of economic status (ESCS), parents' occupation (hisei) and student-to-teacher ratio (STRATIO) by immigration status.

	ESCS (natives)	ESCS (second generation)	ESCS (first generation)	hisei (natives)	hisei (second generation)	hisei (first generation)	STRATIO (natives)	STRATIO (second generation)	STRATIO (first generation)
Australia	0.186	0.199	0.194	56.85	56.89	57.62	13.20	12.89	12.91
Belgium (Flanders)	0.271	−0.500	−0.242	54.07	40.31	43.50	9.07	8.25	10.84
Colombia[a]	−1.065	−1.034	−2.023	41.29	34.34	39.42	27.55	40.26	15.46
Czech Republic[b]	0.081	−0.075	0.066	49.99	43.8	47.23	13.59	12.50	13.34
Spain	−0.058	−0.681	−0.665	49.91	40.68	35.45	12.53	10.57	11.67
Estonia[b]	0.159	−0.025	−0.067	51.79	47.54	42.65	11.83	11.69	10.46
France	0.054	−0.378	−0.847	53.03	46.47	39.77	11.74	12.14	13.23
Croatia	−0.328	−0.700	−0.658	45.32	37.29	36.52	12.59	12.67	12.93
Israel	0.229	0.221	−0.243	58.88	58.33	52.85	11.24	10.76	10.52
Italy	0.021	−0.320	−0.476	48.14	40.28	35.07	9.98	9.67	9.54
Latvia[a]	−0.147	0.143	0.555	49.99	58.79	63.33	10.40	11.23	12.05
New Zealand	0.119	−0.123	0.180	56.17	51.23	58.31	15.02	15.68	15.59
Poland[a]	−0.134		−1.18	46.91		24.53	9.29		10.28
China (Shanghai)[a]	−0.410	−0.902	−1.097	50.26	51.43	50.11	12.16	9.91	12.93
Russia	−0.056	−0.030	−0.036	53.49	54.19	56.34	14.94	15.15	16.02
Slovakia[a]	−0.133	0.162	−0.148	43.56	51.32	38.03	13.39	16.65	12.47
Slovenia	0.055	−0.384	−0.598	50.38	44.98	35.31	9.56	10.11	9.43
USA	0.350	−0.317	−0.629	57.51	44.85	42.00	16.85	18.73	20.22

[a,b] Countries with low/very low number of immigrants, respectively (see Table 3 for details).

Table 8. OLS and matching results by gender.

	OLS (all immigrants)	OLS (first generation)	OLS (second generation)	Matching (all immigrants)	Matching (first generation)	Matching (second generation)
Total	−12.048	−18.139	−7.105	−6.715	−1.009	−7.390
	(4.25)***	(4.94)***	(1.96)**	(1.99)**	(0.19)	(1.90)*
Male	−13.021	−18.830	−6.018	−4.741	−6.659	−11.349
	(3.38)***	(3.75)***	(1.18)	(0.90)	(0.83)	(1.65)*
Female	−11.009	−16.021	−5.527	−5.365	3.427	−6.688
	(3.05)***	(3.30)***	(1.26)	(1.07)	(0.60)	(0.99)

Notes: Model type and treatment variable specified in column header. Estimated coefficient of the immigration dummy reported. *t*-Statistics in parenthesis.
*$p < .1$.
**$p < .05$.
***$p < .01$.

Table 9. Quantile regression results.

	All immigrants	First generation	Second generation
1st percentile	−24.482	−28.520	−24.004
	(3.09)***	(2.62)***	(2.37)**
5th percentile	−9.001	−5.908	−18.287
	(2.18)**	(0.71)	(2.14)**
10th percentile	−6.274	−9.588	−1.926
	(2.00)**	(1.80)*	(0.26)
25th percentile	−13.535	−18.551	−8.387
	(3.80)***	(3.52)***	(2.21)**
50th percentile	−10.246	−13.738	−6.828
	(3.88)***	(2.89)***	(1.97)**
75th percentile	−12.169	−21.200	−6.271
	(4.58)***	(4.34)***	(2.06)**
90th percentile	−8.348	−15.589	1.142
	(2.08)**	(4.94)***	(0.22)
95th percentile	−8.951	−21.388	−1.233
	(1.55)	(3.63)***	(0.20)
99th percentile	1.517	−7.538	16.366
	(0.13)	(0.65)	(1.32)

Notes: Treatment variable specified in column header and quantile listed in row header. Estimated coefficient of the immigration dummy reported. *t*-Statistics in parenthesis.
*$p < .1$.
**$p < .05$.
***$p < .01$.

generation immigrants lag behind natives more strongly than their second-generation counterparts everywhere except at the 5th percentile. The gaps for the first generation are negative and significant at least at the 10% level everywhere except at the 5th and 99th percentiles, whereas second-generation immigrants display significant gaps only at the 1st, 5th, 25th, 50th and 75th percentiles. At the 99th percentile, second-generation immigrants actually perform better than natives, although the gap is not statistically significant.

Finally, in order to see how much of the gap is due to better numeracy or overall learning skills, the student's Math and Reading scores have been partialled out in alternative model specifications (Table 5). We can see that a student's Math score contributes to around 70% of his/her FL score. While some background variables remain significant, such as the grade, parents' occupation or economic status, the native-immigrant gap gets very close (and statistically identical) to 0. This suggests that numeracy is one of the main channels through which the miscellaneous effects of migration background translate into poorer FL. Note that a similar result is not achieved with Reading, as the gap still remains statistically significant if only the Reading score is partialled out. In the case when both scores and their interaction term are included, the effect of the Math score on FL and its *t*-statistic are about twice as high as those of the Reading score.

5.3. Matching estimation

As was mentioned previously, non-parametric ATE estimation via PSM allows for less restrictive assumptions. In particular, it does not assume a linear model specification, but is rather asymptotically valid for any functional relationship between *Y* and *X*, given that hidden bias is not present. In this regard, PSM estimation serves as a sensitivity check in order to point out how much of the residual gap from the OLS estimation might be due to a nonlinear relationship among test score, immigration status and the covariates.

PSM estimation results are presented in Table 6 (right part). The first row presents the gap for the whole sample, whereas the following rows present by-country gaps (both using within-country matching). The untreated group always consists of natives; the treated group is alternatively all immigrants, first-generation immigrants and second-generation immigrants.

The results of matching estimation tell us that the unexplained gap in FL between natives and immigrants is even lower – around 7 PISA points, though still significant at the 5% level. Contrary to OLS though, the originally much higher gap for first-generation immigrants is now insignificant, whereas the gap for the second generation stays virtually unchanged. This may be a joint consequence of (a) convex effects of some covariates, in particular the student's grade, economic status, or parents' occupation, and (b) the inferior observables of first-generation immigrants relative to natives and second-generation immigrants.

Country-level heterogeneity after matching is even more volatile, but still largely in line with previous OLS findings. Australia and Israel still exhibit a positive FL gap, although for first-generation immigrants in Australia it is now insignificant. Slovenia, Estonia, France, and Flanders still show negative gaps, though for France it is significant for second-generation immigrants only. Spain displays a very peculiar and unreliable result, with a gap of −68 (significant at the 1% level) for first-generation and +76 (but still not significant) for second-generation immigrants. Similar irregularities, but on a lesser scale, can be observed for New Zealand. Naturally, there is a large random error factor due to the small number of first-generation immigrant students in some of our sample countries.

It is worth noting that while non-parametric PSM provides the additional flexibility over OLS of not assuming a functional form for the relationship between the test score, immigration status and covariates, this does not come at the cost of extra bias resulting from endogeneity. It is true that, if the unconfoundedness assumption is violated, the estimation will be inconsistent. However, the unconfoundedness assumption for matching is equivalent to the conditional mean independence assumption for OLS, and thus PSM will be consistent whenever OLS is consistent. In the cases when endogeneity is present, there is no reason to suggest that matching estimation will be more biased than OLS (Rubin 1973). Therefore, PSM results are more robust than their OLS counterparts and the matching specification is preferred.

Matching without a caliper restriction achieves reasonable but not perfect balance in covariates between natives and immigrants (see Appendix Table A1). The overall mean percentage bias drops from 13.9% to 2.5%, but 7 regressors are still significantly different at the 1% level, with class size having the most noticeable difference. The remaining imbalance can only be mitigated by imposing a rather narrow caliper of 10^{-5}, which leaves less than 20% of the sample on the common support. Also it is worth noting that the treatment effect dampens progressively and becomes statistically insignificant as the caliper becomes narrower. While the baseline result fails to withstand this robustness check, the insignificant treatment effect when using a narrow caliper can be a consequence of the small remaining sample. A very narrow caliper filters out a lot of observations, thus inflating the resulting standard error. It is also possible, however, that the initial matching estimation result is due to lack of balance and driven by some outliers, for which the matched counterpart is still quite far in terms of propensity score. Caliper use, in this particular study, does not help to achieve better balance, while reducing the sample size. Thus, matching without calipers is preferred.

6. Conclusion

The PISA 2012 FL test scores reveal that there is a gap of about 15 PISA points on the FL test between native and immigrant students in the sampled 18 countries. This gap grows to 30 points if we consider first-generation immigrants only and shrinks to 5 points for the second generation. This gap is largely, but not completely explained by the student's background characteristics. Immigrants tend to have a poorer economic and social status, have less education materials available at home and less wealth overall. Furthermore, their parents are more likely to be employed in low-skilled jobs, and first-generation immigrant children are more likely to be enrolled in a grade below the expected one for their age. All these factors contribute to immigrant students' underperformance in FL as well as other subjects. Curiously, school-related characteristics cannot be unambiguously highlighted as a contributing factor.

When these factors are controlled for, OLS still reveals a gap of about 12 points (18 for first generation, 7 for second generation). This is due to other particularities of the native and immigrant student population, different from student, parent or school background characteristics. It might be, especially for the first generation, owing to the difficulty and stress of coming into a different country and adapting to a different culture and different rules, including variations in functioning of the financial system. However, it might also be due to nonlinear effects of covariates or hidden bias. Matching estimation was employed to account for nonlinear effects of the control variables on the test score. As a result, the unexplained gap in FL shrinks to 6.7, though still remains significant at the 5% level. It ceases to be significant once a caliper restriction is enforced, however. The gap also shrinks after including country- or region-level fixed effects, and there is a lot of country-level heterogeneity in the native-immigrant gaps.

In particular, traditional immigration countries (the USA, Australia, New Zealand, Israel) show negative gaps in FL, which are in some cases statistically significant. On the other hand, European destinations, which are dominated by labor migration, exhibit mostly positive gaps, which are also sometimes statistically significant. As a result, we can see that the migrant selection process and the different success of migrant integration are important determinant of the disparities in FL. There is no similar gender-level heterogeneity, but there is some ability-based heterogeneity, with the gap progressively disappearing as we move towards the top of the score distribution.

The gap in FL has been compared to the ones in Reading and Math. Over the whole sample, the Reading gap is smaller and statistically insignificant for second-generation immigrants, whereas the Math gap is slightly larger than the FL gap. Both the students' Math and Reading scores capture a good deal of the variation in FL, however the Math score does a better job.

This paper's novel contribution is its investigation of the FL gap between native and immigrant students with the help of a large cross-country micro-level dataset. It confirms that the native-immigrant gap in other school subjects, as found by Entorf (2015) and Kunz (2016), carries over to FL. However, given that the scores in PISA are normalized, this analysis cannot serve as an evaluation of the overall level of FL of the target population, similar to Lusardi, Mitchell, and Curto (2010). The high correlation of FL with student and parent background characteristics is unlikely to be solely due to the fact that in many countries FL is not systematically taught in schools, but rather learned in the family. The gap also manifests itself in Math (and Reading to a lesser extent), which suggests that a poorer background leads not just to an isolated effect on FL, but to lower educational ability in general and lower numeracy, which then also translates into poorer FL. This argument is supported by the fact that partialling out the Math effect makes the coefficient of immigration status very small and statistically insignificant. However, it might still be the case that school courses targeted to develop FL can attenuate the discrepancy, as in the case of Hospido, Villanueva, and Zamarro (2015). Studying efficient policies to address this gap leaves broad opportunities for future research.

Disclosure statement

No potential conflict of interest was reported by the author.

ORCID

Iulian Gramaţki http://orcid.org/0000-0002-5257-2364

References

Abadie, A., D. Drukker, J. Herr, and G. Imbens. 2004. “Implementing Matching Estimators for Average Treatment Effects in Stata.” *The Stata Journal* 4: 290–311. http://econpapers.repec.org/RePEc:tsj:stataj:v:4:y:2004:i:3:p:290-311

Abadie, A., and G. Imbens. 2006. “Large Sample Properties of Matching Estimators for Average Treatment Effects.” *Econometrica* 74 (1): 235–267. doi:10.1111/j.1468-0262.2006.00655.x

Abadie, A., and G. Imbens. 2008. “On the Failure of the Bootstrap for Matching Estimators.” *Econometrica* 76 (6): 1537–1557. doi:10.3982/ECTA6474.

Abadie, A., and G. Imbens. 2016. “Matching on the Estimated Propensity Score.” *Econometrica* 84 (2): 781–807. doi:10.3982/ECTA11293

Chernozhukov, V., I. Fernandez-Val, and B. Melly. 2013. “Inference on Counterfactual Distributions.” *Econometrica* 81 (6): 2205–2268. doi:10.3982/ECTA10582

Entorf, H. 2015. “Migrants and Educational Achievement Gaps: Avoiding Segregation and lCompensating for Parental Disadvantage can Reduce Migrants’ Educational Achievement Gaps.” IZA World of Labor. doi:10.15185/izawol.146

Fuchs, T., and L. Woessmann. 2004. “Computers and Student Learning: Bivariate and Multivariate Evidence on the Availability and Use of Computers at Home and at Schools.” *Brussels Economic Review* 47: 359–389. http://www.cesifo-group.de/portal/page/portal/DocBase_Content/WP/WP-CESifo_Working_Papers/wp-cesifo-2004/wp-cesifo-2004-11/cesifo1_wp1321.pdf.

Ganzeboom. 2010. “Occupational Status Measures for the New International Standard Classification of Occupations ISCO-08.” Working Paper. http://www.harryganzeboom.nl/isol/isol2010c2-ganzeboom.pdf.

Hospido, L., E. Villanueva, and G. Zamarro. 2015. “Finance for All: The Impact of Financial Literacy Training in Compulsory Secondary Education in Spain.” IZA DP No. 8902. http://dx.doi.org/10.1787/9789264208094-en

Koenker, R., and G. Bassett. 1978. “Regression Quantiles.” *Econometrica* 46 (1): 33–50. doi:10.2307/1913643

Kunz, J. S. 2016. “Analyzing Educational Achievement Differences Between Second-Generation Immigrants: Comparing Germany and German-Speaking Switzerland.” *German Economic Review* 17 (1): 61–91 doi:10.1111/geer.12062

Lusardi, A. 2008. “Financial Literacy: An Essential Tool for Informed Consumer Choice?” CFS Working Paper, No. 2008/19, http://nbn-resolving.de/urn:nbn:de:hebis:30-56927

Lusardi, A., and O. S. Mitchell. 2014. “The Economic Importance of Financial Literacy: Theory and Evidence.” *Journal of Economic Literature* 52 (1): 5–44. doi:10.1257/jel.52.1.5

Lusardi, A., O. S. Mitchell, and V. Curto. 2010. “Financial Literacy Among the Young.” *Journal of Consumer Affairs* 44 (2): 358–380. doi:10.1111/j.1745-6606.2010.01173.x

OECD. 2009. *PISA Data Analysis Manual. SPSS*. 2nd ed. PISA, OECD Publishing. doi:10.1787/9789264056275-en

OECD. 2013. *PISA 2012 Financial Literacy Framework*. PISA, OECD Publishing. doi:10.1787/9789264190511-en

OECD. 2014a. *PISA 2012 Results: Students and Money: Financial Literacy Skills for the 21st Century (Volume VI)*. PISA, OECD Publishing. doi:10.1787/9789264208094-4-en

OECD. 2014b. *PISA 2012 Technical Report*. PISA, OECD Publishing. https://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf

Powell, D. (2016). “Quantile Treatment Effects in the Presence of Covariates.” Working Paper, https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbnxkYXZpZG1hdHRoZXdwb3dlbGx8Z3g6NTQyNzg5MTM2MWRiMjgzMg

Rosenbaum, P., and D. B. Rubin. 1983. “The Central Role of the Propensity Score in Observational Studies for Causal Effects.” *Biometrika* 70 (1): 41–55. doi:10.1093/biomet/70.1.41

Rubin, D. B. 1973. “Matching to Remove Bias in Observational Studies.” *Biometrics* 29: 159–184. doi:10.2307/2529684

Appendix

Table A1. Matching estimation with different caliper sizes (indicated in column header).

	Unmatched		No caliper		1e-4 Caliper		5e-5 Caliper		1e5-Caliper	
Variable	%Bias	*p*-Value	%Bias	*p*-Value	%Bias	*p*-Value	%Bias	*p*-Value	% Bias	*p*-Value
Gender	−6.2	0.003***	−5.2	0***	−2.7	0.022**	−1.9	0.178	0.2	0.935
Age	−0.9	0.646	−1.1	0.287	0.9	0.464	2.2	0.118	3.8	0.165
Grade (cohort)	13.3	0***	−1.6	0.091*	−0.6	0.617	−1.4	0.306	1.7	0.495
Cultural possessions	31	0***	−0.8	0.448	−3.7	0.002***	−5.9	0***	−3.6	0.176
Economic status	14.4	0***	0.3	0.755	−1.8	0.127	−3.1	0.027**	0.6	0.827
Family structure	2.3	0.289	−2.1	0.032**	−2.5	0.031**	−3.1	0.026**	−1.2	0.658
Home educational resources	9	0***	3.2	0.002***	1.6	0.163	1.5	0.269	2	0.45
Parents' education	−2.6	0.199	−0.1	0.928	0.4	0.759	−0.8	0.542	1.7	0.532
Parents' occupation	12.9	0***	−1.1	0.26	−3.8	0.001***	−4.4	0.002***	0.5	0.864
Home possessions	19.8	0***	0.1	0.942	−3.1	0.011**	−4.3	0.002***	−2.6	0.339
Wealth index	3.7	0.076*	−1.4	0.168	−3.8	0.001***	−4.6	0.001***	−3.5	0.189
Test language at home	78.6	0***	−0.4	0.62	0	0.966	0.6	0.467	0.3	0.811
Student–teacher ratio	5.1	0.024**	3.8	0***	4.9	0***	4.5	0***	6.7	0.007***
School educational materials	−19.9	0***	−2.4	0.017**	−2.3	0.044**	−2.8	0.038**	−2.4	0.37
School buildings	−6.6	0.002***	0.2	0.811	0.9	0.445	−0.2	0.912	−0.4	0.885
School type	3.5	0.089*	3	0.002***	2.5	0.03**	2.8	0.044**	4.3	0.11
School selectivity	1.7	0.408	−3.7	0***	−3.4	0.003***	−2.9	0.036**	−2.9	0.286
School size	−19	0***	−5.1	0***	−6.1	0***	−6.4	0***	−5.4	0.017**
Class size	13.3	0***	−11.8	0***	−10.5	0***	−10.3	0***	−8.9	0.001***
Pseudo R^2	0.125		0.006		0.008		0.009		0.008	
Likelihood ratio χ^2	1539.58		336.09		321.51		239.76		58.33	
p-Value	0***		0***		0***		0***		0***	
Mean bias (%)	13.9		2.5		2.9		3.4		2.8	
Median bias (%)	9		1.6		2.5		2.9		2.4	
Unbalanced variables	12		7		7		7		2	
Observations on support										
Total	21,088		21,088		15,926		11,561		3626	
Treated	19,291		19,291		14,367		10,156		2695	
Untreated	1797		1797		1559		1405		931	
Observations off support										
Total	0		0		5162		9527		17,462	
Treated	0		0		4924		9135		16,596	
Untreated	0		0		238		392		866	
Treatment effects	12.048		6.715		3.289		2.896		2.366	
t-Statistic	4.25***		1.99**		0.59		0.48		0.62	

Notes: Balance check for differences of means of covariates between treatment groups. The top section presents the percentage variance for the respective variable (listed in the row header) between natives and immigrants. The middle section provides summary statistics for joint mean difference. The third section shows the number of observations on and off the support by treatment status, and the bottom section displays estimated treatment effects.

*$p < .1$.

**$p < .05$.

***$p < .01$.

Education Economics

ISSN: 0964-5292 (Print) 1469-5782 (Online) Journal homepage: https://www.tandfonline.com/loi/cede20

A comparison of financial literacy between native and immigrant school students

Iulian Gramațki